Matrix Analysis and Applications

Chapter 9: Kronecker Product

Instructor: Kai Lu

(http://seit.sysu.edu.cn/teacher/1801)

School of Electronics and Information Technology
Sun Yat-sen University

December 20, 2020

Table of Contents

Table of Contents

Motivating Problem: Matrix Equations

Consider the following problem: given $\boldsymbol{A}$ and $\boldsymbol{B}$, find an $\boldsymbol{X}$ such that

$$\boldsymbol{A}\boldsymbol{X} = \boldsymbol{B}. \tag{1}$$

This problem is easy, e.g., if $\boldsymbol{A}$ has full column rank and Eq. (1) has a solution, then the solution is $\boldsymbol{X} = \boldsymbol{A}^{\dagger}\boldsymbol{B}$.

Motivating Problem: Matrix Equations

Consider the following problem: given $\boldsymbol{A}$ and $\boldsymbol{B}$, find an $\boldsymbol{X}$ such that

$$\boldsymbol{A}\boldsymbol{X} = \boldsymbol{B}. \tag{1}$$

This problem is easy, e.g., if $\boldsymbol{A}$ has full column rank and Eq. (1) has a solution, then the solution is $\boldsymbol{X} = \boldsymbol{A}^{\dagger}\boldsymbol{B}$.

Question 1

What about matrix equations like

1) $\boldsymbol{A}\boldsymbol{X} + \boldsymbol{X}\boldsymbol{B} = \boldsymbol{C}$,
2) $\boldsymbol{A}_1\boldsymbol{X}\boldsymbol{B}_1 + \boldsymbol{A}_2\boldsymbol{X}\boldsymbol{B}_2 = \boldsymbol{C}$,
3) $\boldsymbol{A}\boldsymbol{X} + \boldsymbol{Y}\boldsymbol{B} = \boldsymbol{C}$, *where both* $\boldsymbol{X}$ *and* $\boldsymbol{Y}$ *are unknown?*

Motivating Problem: Matrix Equations

Consider the following problem: given $\boldsymbol{A}$ and $\boldsymbol{B}$, find an $\boldsymbol{X}$ such that

$$\boldsymbol{A}\boldsymbol{X} = \boldsymbol{B}. \tag{1}$$

This problem is easy, e.g., if $\boldsymbol{A}$ has full column rank and Eq. (1) has a solution, then the solution is $\boldsymbol{X} = \boldsymbol{A}^{\dagger}\boldsymbol{B}$.

Question 1

What about matrix equations like

1) $\boldsymbol{A}\boldsymbol{X} + \boldsymbol{X}\boldsymbol{B} = \boldsymbol{C}$,
2) $\boldsymbol{A}_1\boldsymbol{X}\boldsymbol{B}_1 + \boldsymbol{A}_2\boldsymbol{X}\boldsymbol{B}_2 = \boldsymbol{C}$,
3) $\boldsymbol{A}\boldsymbol{X} + \boldsymbol{Y}\boldsymbol{B} = \boldsymbol{C}$, *where both* $\boldsymbol{X}$ *and* $\boldsymbol{Y}$ *are unknown?*

The above matrix equations can be tackled via matrix tools arising from the **Kronecker product**, which has a rich algebra that supports a wide range of fast, practical algorithms. More important, it provides a bridge between **matrix computations** and **tensor computations**.

Kronecker Product

The Kronecker product of matrices $\boldsymbol{A} \in \mathbb{C}^{m \times n}$ and $\boldsymbol{B} \in \mathbb{C}^{p \times q}$ is defined as

$$\boldsymbol{A} \otimes \boldsymbol{B} \triangleq \begin{bmatrix} a_{11}\boldsymbol{B} & a_{12}\boldsymbol{B} & \cdots & a_{1n}\boldsymbol{B} \\ a_{21}\boldsymbol{B} & a_{22}\boldsymbol{B} & \cdots & a_{2n}\boldsymbol{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}\boldsymbol{B} & a_{m2}\boldsymbol{B} & \cdots & a_{mn}\boldsymbol{B} \end{bmatrix} \in \mathbb{C}^{mp \times nq}, \tag{2}$$

which is also known as the **tensor product** or the **direct product**.

Kronecker Product

The Kronecker product of matrices $\boldsymbol{A} \in \mathbb{C}^{m\times n}$ and $\boldsymbol{B} \in \mathbb{C}^{p\times q}$ is defined as

$$\boldsymbol{A} \otimes \boldsymbol{B} \triangleq \begin{bmatrix} a_{11}\boldsymbol{B} & a_{12}\boldsymbol{B} & \cdots & a_{1n}\boldsymbol{B} \\ a_{21}\boldsymbol{B} & a_{22}\boldsymbol{B} & \cdots & a_{2n}\boldsymbol{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}\boldsymbol{B} & a_{m2}\boldsymbol{B} & \cdots & a_{mn}\boldsymbol{B} \end{bmatrix} \in \mathbb{C}^{mp\times nq}, \tag{2}$$

which is also known as the **tensor product** or the **direct product**.

Example 1 (Kronecker product vs. Outer product)

Let $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{C}^{m\times 1}$. By definition, we have $\boldsymbol{a} \otimes \boldsymbol{b} = \begin{bmatrix} a_1\boldsymbol{b} \\ a_2\boldsymbol{b} \\ \vdots \\ a_m\boldsymbol{b} \end{bmatrix} \in \mathbb{C}^{m^2\times 1}$. On the other hand, we know that $\boldsymbol{b}\boldsymbol{a}^T = [a_1\boldsymbol{b}, a_2\boldsymbol{b}, \cdots, a_m\boldsymbol{b}] \in \mathbb{C}^{m\times m}$. Hence, it is clear that $\boldsymbol{a} \otimes \boldsymbol{b}$ is a column-by-column concatenation of the **outer product** $\boldsymbol{b}\boldsymbol{a}^T$ (vs. the **inner product** $\boldsymbol{a}^T\boldsymbol{b}$).

Kronecker (1823-1891, Liegnitz, Prussia [now Poland])

God made the integers, all else is the work of man.

-------- Leopold Kronecker

Carl Jacobi (1804-1851) (1929, 1939, 1943)

Ernst Kummer
(1810-1893)

- 1831, Ph.D.
- 1842, University of Breslau (University of Wroclaw)
- 1855, Professor, University of Berlin
- 1868-1869, President, University of Berlin

Leopold Kronecker
(1823-1891)

- 1845, Ph.D. (Supervisor: Dirichlet)
- 1855, Berlin (became close friend with Weierstrass, a new teacher at Berlin U.)
- 1861, member of the Berlin Academy
- 1862, University of Berlin
- 1866, offer from University of Göttingen (Carl Gauss, Dirichlet, Riemann)
- 1883, Kummer retired, ordinary professor

Gustav Lejeune Dirichlet
(1805-1859)

- 1825, comprising part of a proof of Fermat's last theorem for the case n=5 (Fourier and Poisson)
- 1827, honorary doctorate, University of Bonn
- 1928, Berlin (Friedrich Bessel)
- Number Theory, Fourier series
- Dirichlet defines a function by the property that "to any *x* there corresponds a single finite *y*"

Karl Weierstrass
(1815-1897)

- father of modern analysis
- formalized the definition of the continuity of a function
- proved the intermediate value theorem
- Bolzano–Weierstrass theorem (i.e., each bounded sequence in R^n has a convergent subsequence)

Hermann Schwarz
(1843-1921)

- 1864, Ph.D.
- complex analysis
- Cauchy–Schwarz inequality $|\langle \mathbf{u}, \mathbf{v}\rangle|^2 \leq \langle \mathbf{u}, \mathbf{u}\rangle \cdot \langle \mathbf{v}, \mathbf{v}\rangle$

"he who does not honor the Smaller, is not worthy of the Greater,"

Elementary Properties

1) $\boldsymbol{A} \otimes (\alpha \boldsymbol{B}) = \alpha \boldsymbol{A} \otimes \boldsymbol{B}$ (scaling).
2) $(\boldsymbol{A} + \boldsymbol{B}) \otimes \boldsymbol{C} = \boldsymbol{A} \otimes \boldsymbol{C} + \boldsymbol{B} \otimes \boldsymbol{C}, \;\; \boldsymbol{A} \otimes (\boldsymbol{B} + \boldsymbol{C}) = \boldsymbol{A} \otimes \boldsymbol{B} + \boldsymbol{A} \otimes \boldsymbol{C}$ (distributive).
3) $\boldsymbol{A} \otimes (\boldsymbol{B} \otimes \boldsymbol{C}) = (\boldsymbol{A} \otimes \boldsymbol{B}) \otimes \boldsymbol{C}$ (associativity).
4) $\mathbf{0}_{mn} = \mathbf{0}_m \otimes \mathbf{0}_n, \;\; \boldsymbol{I}_{mn} = \boldsymbol{I}_m \otimes \boldsymbol{I}_n$.
5) $(\boldsymbol{A} \otimes \boldsymbol{B})^T = \boldsymbol{A}^T \otimes \boldsymbol{B}^T, \;\; (\boldsymbol{A} \otimes \boldsymbol{B})^H = \boldsymbol{A}^H \otimes \boldsymbol{B}^H$.

[1] [R1] A. Graham, *Kronecker Products and Matrix Calculus with Applications*, Dover Publications, 2018.

Elementary Properties

1) $\boldsymbol{A} \otimes (\alpha \boldsymbol{B}) = \alpha \boldsymbol{A} \otimes \boldsymbol{B}$ (scaling).
2) $(\boldsymbol{A} + \boldsymbol{B}) \otimes \boldsymbol{C} = \boldsymbol{A} \otimes \boldsymbol{C} + \boldsymbol{B} \otimes \boldsymbol{C}, \ \ \boldsymbol{A} \otimes (\boldsymbol{B} + \boldsymbol{C}) = \boldsymbol{A} \otimes \boldsymbol{B} + \boldsymbol{A} \otimes \boldsymbol{C}$ (distributive).
3) $\boldsymbol{A} \otimes (\boldsymbol{B} \otimes \boldsymbol{C}) = (\boldsymbol{A} \otimes \boldsymbol{B}) \otimes \boldsymbol{C}$ (associativity).
4) $\mathbf{0}_{mn} = \mathbf{0}_m \otimes \mathbf{0}_n, \ \ \boldsymbol{I}_{mn} = \boldsymbol{I}_m \otimes \boldsymbol{I}_n.$
5) $(\boldsymbol{A} \otimes \boldsymbol{B})^T = \boldsymbol{A}^T \otimes \boldsymbol{B}^T, \ \ (\boldsymbol{A} \otimes \boldsymbol{B})^H = \boldsymbol{A}^H \otimes \boldsymbol{B}^H.$
6) There exist permutation matrices $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ such that

$$\boldsymbol{U}_1(\boldsymbol{A} \otimes \boldsymbol{B})\boldsymbol{U}_2 = \boldsymbol{B} \otimes \boldsymbol{A}. \tag{3}$$

Note: Kronecker product is not commutative, i.e., $\boldsymbol{A} \otimes \boldsymbol{B} \neq \boldsymbol{B} \otimes \boldsymbol{A}$ in general. Property 6) above is a weak version of commutativity. For more details on how to determine $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ used in (3), please refer to Section 2.5 of [R1].[1]

[1] [R1] A. Graham, *Kronecker Products and Matrix Calculus with Applications*, Dover Publications, 2018.

More Properties

Property 1 (Mixed Product Rule)

$$(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{C} \otimes \boldsymbol{D}) = (\boldsymbol{A}\boldsymbol{C}) \otimes (\boldsymbol{B}\boldsymbol{D}), \tag{4}$$

for $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}$ *of appropriate matrix dimensions.*

More Properties

Property 1 (Mixed Product Rule)

$$(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{C} \otimes \boldsymbol{D}) = (\boldsymbol{AC}) \otimes (\boldsymbol{BD}), \tag{4}$$

for $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}$ of appropriate matrix dimensions.

Some properties from Property 1:

1) If $\boldsymbol{A} \in \mathbb{C}^{m\times m}$ and $\boldsymbol{B} \in \mathbb{C}^{n\times n}$ are nonsingular, then

$$(\boldsymbol{A} \otimes \boldsymbol{B})^{-1} = \boldsymbol{A}^{-1} \otimes \boldsymbol{B}^{-1}.$$

Proof: $(\boldsymbol{A}^{-1} \otimes \boldsymbol{B}^{-1})(\boldsymbol{A} \otimes \boldsymbol{B}) = (\boldsymbol{A}^{-1}\boldsymbol{A}) \otimes (\boldsymbol{B}^{-1}\boldsymbol{B}) = \boldsymbol{I}_m \otimes \boldsymbol{I}_n = \boldsymbol{I}_{mn}$.

2) If $\boldsymbol{Q}_1$, $\boldsymbol{Q}_2$ are semi-unitary, then $\boldsymbol{Q}_1 \otimes \boldsymbol{Q}_2$ is semi-unitary.

Proof: $(\boldsymbol{Q}_1 \otimes \boldsymbol{Q}_2)^H(\boldsymbol{Q}_1 \otimes \boldsymbol{Q}_2) = (\boldsymbol{Q}_1^H \otimes \boldsymbol{Q}_2^H)(\boldsymbol{Q}_1 \otimes \boldsymbol{Q}_2) = (\boldsymbol{Q}_1^H\boldsymbol{Q}_1) \otimes (\boldsymbol{Q}_2^H\boldsymbol{Q}_2) = \boldsymbol{I}$.

More Properties (cont'd)

Kronecker Product inherits the **structure** of special matrices:

1) If $\boldsymbol{A}$ is sparse, then $\boldsymbol{A} \otimes \boldsymbol{B}$ has the same sparsity pattern at the block level.
2) If $\boldsymbol{A}$ and $\boldsymbol{B}$ are permutation matrices, then $\boldsymbol{A} \otimes \boldsymbol{B}$ is also a permutation matrix.
3) (orthogonal) $\otimes$ (orthogonal) $=$ (orthogonal)
4) (stochastic) $\otimes$ (stochastic) $=$ (stochastic)
5) (symmetric positive definite, SPD) $\otimes$ (SPD) $=$ (SPD)

   Proof: Let $\boldsymbol{A} = \boldsymbol{L}_1 \boldsymbol{L}_1^T$ and $\boldsymbol{B} = \boldsymbol{L}_2 \boldsymbol{L}_2^T$. We have

   $$\boldsymbol{A} \otimes \boldsymbol{B} = \boldsymbol{L}_1 \boldsymbol{L}_1^T \otimes \boldsymbol{L}_2 \boldsymbol{L}_2^T \xlongequal{(4)} (\boldsymbol{L}_1 \otimes \boldsymbol{L}_2)(\boldsymbol{L}_1^T \otimes \boldsymbol{L}_2^T) = (\boldsymbol{L}_1 \otimes \boldsymbol{L}_2)(\boldsymbol{L}_1 \otimes \boldsymbol{L}_2)^T.$$

   Similarly, if $\boldsymbol{A} = \boldsymbol{Q}_1 \boldsymbol{R}_1$ and $\boldsymbol{B} = \boldsymbol{Q}_2 \boldsymbol{R}_2$, then $\boldsymbol{A} \otimes \boldsymbol{B} = (\boldsymbol{Q}_1 \otimes \boldsymbol{Q}_2)(\boldsymbol{R}_1 \otimes \boldsymbol{R}_2)$.

Example 2 (Hadamard Matrix)

Consider an 2×2 orthogonal matrix

$$\boldsymbol{H}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

From $\boldsymbol{H}_2$, construct a 4×4 matrix

$$\boldsymbol{H}_4 = \boldsymbol{H}_2 \otimes \boldsymbol{H}_2 = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix},$$

and inductively, $\boldsymbol{H}_{2^n} = \boldsymbol{H}_{2^{n-1}} \otimes \boldsymbol{H}_2$.

Example 2 (Hadamard Matrix)

Consider an 2×2 orthogonal matrix

$$\boldsymbol{H}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

From $\boldsymbol{H}_2$, construct a 4×4 matrix

$$\boldsymbol{H}_4 = \boldsymbol{H}_2 \otimes \boldsymbol{H}_2 = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix},$$

and inductively, $\boldsymbol{H}_{2^n} = \boldsymbol{H}_{2^{n-1}} \otimes \boldsymbol{H}_2$.

1) Is $\boldsymbol{H}_4$ orthogonal?

   Yes, as $\boldsymbol{H}_4 \boldsymbol{H}_4^T = (\boldsymbol{H}_2 \otimes \boldsymbol{H}_2)(\boldsymbol{H}_2^T \otimes \boldsymbol{H}_2^T) = (\boldsymbol{H}_2 \boldsymbol{H}_2^T \otimes \boldsymbol{H}_2 \boldsymbol{H}_2^T) = \boldsymbol{I}$.

2) For the same reason, any $\boldsymbol{H}_n$ is orthogonal.

More Properties (cont'd)

There is a direct correspondence between the eigen-equations of $\boldsymbol{A} \otimes \boldsymbol{B}$ and $\boldsymbol{A}$, $\boldsymbol{B}$. More specifically, we have

Theorem 3 (Kronecker Product and Eigen-Pairs)

Let $\boldsymbol{A} \in \mathbb{C}^{m \times m}$, $\boldsymbol{B} \in \mathbb{C}^{n \times n}$. Let $\{\alpha_i, \boldsymbol{x}_i\}_{i=1}^m$ be the set of m eigen-pairs of $\boldsymbol{A}$, and $\{\beta_j, \boldsymbol{y}_j\}_{j=1}^n$ be the set of n eigen-pairs of $\boldsymbol{B}$. The set of mn eigen-pairs of $\boldsymbol{A} \otimes \boldsymbol{B}$ is given by

$$\{\alpha_i \beta_j, \boldsymbol{x}_i \otimes \boldsymbol{y}_j\}_{i=1,\cdots,m,\ j=1,\cdots,n}.$$

More Properties (cont'd)

Properties arising from Theorem 3 ($\boldsymbol{A} \in \mathbb{R}^{m\times m}$, $\boldsymbol{B} \in \mathbb{R}^{n\times n}$):

6)
$$\sigma(\boldsymbol{A} \otimes \boldsymbol{B}) = \{\alpha_i\beta_j : \alpha_i \in \sigma(\boldsymbol{A}), \beta_j \in \sigma(\boldsymbol{B})\}$$

7)
$$\mathrm{rank}(\boldsymbol{A} \otimes \boldsymbol{B}) = \mathrm{rank}(\boldsymbol{A}) \times \mathrm{rank}(\boldsymbol{B})$$

8)
$$\det(\boldsymbol{A} \otimes \boldsymbol{B}) = \det(\boldsymbol{A})^n \times \det(\boldsymbol{B})^m$$

9)
$$\mathrm{tr}(\boldsymbol{A} \otimes \boldsymbol{B}) = \mathrm{tr}(\boldsymbol{A}) \times \mathrm{tr}(\boldsymbol{B})$$

10)
$$\|\boldsymbol{A} \otimes \boldsymbol{B}\|_F = \|\boldsymbol{A}\|_F \times \|\boldsymbol{B}\|_F$$

11)
$$\|\boldsymbol{A} \otimes \boldsymbol{B}\|_2 = \|\boldsymbol{A}\|_2 \times \|\boldsymbol{B}\|_2$$

Hadamard Product

For two matrices $\boldsymbol{A}$, $\boldsymbol{B}$ of the same dimension $m \times n$, the **Hadamard**[2] **product** $\boldsymbol{A} \circ \boldsymbol{B}$ is a matrix of the same dimension as the operands, with elements given by

$$(\boldsymbol{A} \circ \boldsymbol{B})_{i,j} = (\boldsymbol{A})_{i,j}(\boldsymbol{B})_{i,j}. \tag{5}$$

For matrices of different dimensions ($m \times n$ and $p \times q$, where $m \neq p$ or $n \neq q$ or both) the Hadamard product is undefined.

[2] For more about French mathematician **Jacques Hadamard**, please refer to `https://en.wikipedia.org/wiki/Jacques_Hadamard`.

Hadamard Product

For two matrices $\boldsymbol{A}$, $\boldsymbol{B}$ of the same dimension $m \times n$, the **Hadamard**[2] **product** $\boldsymbol{A} \circ \boldsymbol{B}$ is a matrix of the same dimension as the operands, with elements given by

$$(\boldsymbol{A} \circ \boldsymbol{B})_{i,j} = (\boldsymbol{A})_{i,j}(\boldsymbol{B})_{i,j}. \tag{5}$$

For matrices of different dimensions ($m \times n$ and $p \times q$, where $m \neq p$ or $n \neq q$ or both) the Hadamard product is undefined.

Example 4

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \circ \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} & a_{12}b_{12} & a_{13}b_{13} \\ a_{21}b_{21} & a_{22}b_{22} & a_{23}b_{23} \\ a_{31}b_{31} & a_{32}b_{32} & a_{33}b_{33} \end{bmatrix} \tag{6}$$

[2] For more about French mathematician **Jacques Hadamard**, please refer to `https://en.wikipedia.org/wiki/Jacques_Hadamard`.

Table of Contents

Vectorization

Let $\boldsymbol{A} = [\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n]$. The vectorization of $\boldsymbol{A}$ is given by

$$\mathrm{vec}(\boldsymbol{A}) \triangleq \begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \vdots \\ \boldsymbol{a}_n \end{bmatrix}, \tag{7}$$

i.e., stacking the columns of a matrix to form a column vector.

Vectorization

Let $\boldsymbol{A} = [\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n]$. The vectorization of $\boldsymbol{A}$ is given by

$$\mathrm{vec}(\boldsymbol{A}) \triangleq \begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \vdots \\ \boldsymbol{a}_n \end{bmatrix}, \tag{7}$$

i.e., stacking the columns of a matrix to form a column vector.

Property 2

$$\mathrm{tr}(\boldsymbol{A}\boldsymbol{B}) = \left(\mathrm{vec}(\boldsymbol{A}^H)\right)^H \mathrm{vec}(\boldsymbol{B}) = \left(\mathrm{vec}(\boldsymbol{B}^H)\right)^H \mathrm{vec}(\boldsymbol{A}). \tag{8}$$

Proof.

It is left for your after-class exercise. □

Vectorization

Property 3

$$\mathrm{vec}(\boldsymbol{AXB}) = (\boldsymbol{B}^T \otimes \boldsymbol{A})\mathrm{vec}(\boldsymbol{X}). \tag{9}$$

Vectorization

Property 3

$$\mathrm{vec}(\boldsymbol{AXB}) = (\boldsymbol{B}^T \otimes \boldsymbol{A})\mathrm{vec}(\boldsymbol{X}). \tag{9}$$

Two special cases of Property 3:

$$\mathrm{vec}(\boldsymbol{AX}) = (\boldsymbol{I} \otimes \boldsymbol{A})\mathrm{vec}(\boldsymbol{X}), \tag{10}$$

$$\mathrm{vec}(\boldsymbol{XA}) = (\boldsymbol{A}^T \otimes \boldsymbol{I})\mathrm{vec}(\boldsymbol{X}). \tag{11}$$

Application: Recalling Question 1.

Sketch of the Proof of Property 3

1) Represent $\boldsymbol{X}$ by

$$\boldsymbol{X} = \sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij} \boldsymbol{e}_i \boldsymbol{e}_j^T,$$

where $\boldsymbol{e}_i$'s are unit vectors of appropriate dimension. Then,

$$\mathrm{vec}(\boldsymbol{AXB}) = \mathrm{vec}\left(\sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij} \boldsymbol{A}\boldsymbol{e}_i \boldsymbol{e}_j^T \boldsymbol{B}\right) = \sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij} \mathrm{vec}\left(\boldsymbol{A}\boldsymbol{e}_i \boldsymbol{e}_j^T \boldsymbol{B}\right).$$

2) Denote $\boldsymbol{a}_i \in \mathbb{C}^{p\times 1}$ as the i^{th} column of $\boldsymbol{A}$, and $\boldsymbol{b}_j \in \mathbb{C}^{q\times 1}$ as the j^{th} row of $\boldsymbol{B}$, we have

$$\mathrm{vec}\left(\boldsymbol{A}\boldsymbol{e}_i \boldsymbol{e}_j^T \boldsymbol{B}\right) = \mathrm{vec}\left(\boldsymbol{a}_i \boldsymbol{b}_j^T\right) = \mathrm{vec}\left([\boldsymbol{a}_i b_{j,1}, \cdots, \boldsymbol{a}_i b_{j,q}]\right) = \begin{bmatrix} b_{j,1}\boldsymbol{a}_i \\ \vdots \\ b_{j,q}\boldsymbol{a}_i \end{bmatrix} = \boldsymbol{b}_j \otimes \boldsymbol{a}_i.$$

3) Consequently,

$$\mathrm{vec}(\boldsymbol{AXB}) = \sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij} \boldsymbol{b}_j \otimes \boldsymbol{a}_i = (\boldsymbol{B}^T \otimes \boldsymbol{A})\mathrm{vec}(\boldsymbol{X}).$$

Table of Contents

Kronecker Sum

1) Motivation: Consider the linear matrix equation:

$$\boldsymbol{A}\boldsymbol{X} + \boldsymbol{X}\boldsymbol{B} = \boldsymbol{C}, \tag{12}$$

where $\boldsymbol{A} \in \mathbb{C}^{m\times m}$, $\boldsymbol{B} \in \mathbb{C}^{n\times n}$, and $\boldsymbol{C}, \boldsymbol{X} \in \mathbb{C}^{m\times n}$.

Kronecker Sum

1) Motivation: Consider the linear matrix equation:

$$\boldsymbol{A}\boldsymbol{X} + \boldsymbol{X}\boldsymbol{B} = \boldsymbol{C}, \tag{12}$$

where $\boldsymbol{A} \in \mathbb{C}^{m\times m}$, $\boldsymbol{B} \in \mathbb{C}^{n\times n}$, and $\boldsymbol{C}, \boldsymbol{X} \in \mathbb{C}^{m\times n}$.

2) By vectorizing Eq. (12) with the help of Eqs. (10) and (11), we get

$$\begin{aligned} & (\boldsymbol{I}_n \otimes \boldsymbol{A})\mathrm{vec}(\boldsymbol{X}) + (\boldsymbol{B}^T \otimes \boldsymbol{I}_m)\mathrm{vec}(\boldsymbol{X}) = \mathrm{vec}(\boldsymbol{C}) & (13) \\ \Longrightarrow \quad & \left((\boldsymbol{I}_n \otimes \boldsymbol{A}) + (\boldsymbol{B}^T \otimes \boldsymbol{I}_m)\right)\mathrm{vec}(\boldsymbol{X}) = \mathrm{vec}(\boldsymbol{C}) & (14) \\ \Longrightarrow \quad & \left(\boldsymbol{A} \oplus \boldsymbol{B}^T\right)\mathrm{vec}(\boldsymbol{X}) = \mathrm{vec}(\boldsymbol{C}), & (15) \end{aligned}$$

which is a **linear** system that we have already known how to solve it!

Kronecker Sum

1) Motivation: Consider the linear matrix equation:

$$\boldsymbol{A}\boldsymbol{X} + \boldsymbol{X}\boldsymbol{B} = \boldsymbol{C}, \tag{12}$$

where $\boldsymbol{A} \in \mathbb{C}^{m\times m}$, $\boldsymbol{B} \in \mathbb{C}^{n\times n}$, and $\boldsymbol{C}, \boldsymbol{X} \in \mathbb{C}^{m\times n}$.

2) By vectorizing Eq. (12) with the help of Eqs. (10) and (11), we get

$$(\boldsymbol{I}_n \otimes \boldsymbol{A})\mathrm{vec}(\boldsymbol{X}) + (\boldsymbol{B}^T \otimes \boldsymbol{I}_m)\mathrm{vec}(\boldsymbol{X}) = \mathrm{vec}(\boldsymbol{C}) \tag{13}$$

$$\Longrightarrow \quad \left((\boldsymbol{I}_n \otimes \boldsymbol{A}) + (\boldsymbol{B}^T \otimes \boldsymbol{I}_m)\right)\mathrm{vec}(\boldsymbol{X}) = \mathrm{vec}(\boldsymbol{C}) \tag{14}$$

$$\Longrightarrow \quad \left(\boldsymbol{A} \oplus \boldsymbol{B}^T\right)\mathrm{vec}(\boldsymbol{X}) = \mathrm{vec}(\boldsymbol{C}), \tag{15}$$

which is a **linear** system that we have already known how to solve it!

3) The Kronecker sum of $\boldsymbol{A} \in \mathbb{C}^{m\times m}$ and $\boldsymbol{B} \in \mathbb{C}^{n\times n}$ is defined as

$$\boldsymbol{A} \oplus \boldsymbol{B} \triangleq (\boldsymbol{I}_n \otimes \boldsymbol{A}) + (\boldsymbol{B} \otimes \boldsymbol{I}_m). \tag{16}$$

As a result, if a unique solution to (12) is desired, then we wish to know conditions under which $\boldsymbol{A} \oplus \boldsymbol{B}$ is nonsingular.

Kronecker Sum (cont'd)

Theorem 5 (Kronecker Sum and Eigen-Pairs)

Let $\{\lambda_i, \boldsymbol{x}_i\}_{i=1}^n$ be the set of n eigen-pairs of $\boldsymbol{A}$, and $\{\mu_j, \boldsymbol{y}_j\}_{j=1}^m$ be the set of m eigen-pairs of $\boldsymbol{B}$. The set of mn eigen-pairs of $\boldsymbol{A} \oplus \boldsymbol{B}$ is given by

$$\{\lambda_i + \mu_j, \boldsymbol{y}_j \otimes \boldsymbol{x}_i\}_{i=1,\cdots,n,\ \ j=1,\cdots,m}.$$

Kronecker Sum (cont'd)

Theorem 5 (Kronecker Sum and Eigen-Pairs)

Let $\{\lambda_i, \boldsymbol{x}_i\}_{i=1}^n$ be the set of n eigen-pairs of $\boldsymbol{A}$, and $\{\mu_j, \boldsymbol{y}_j\}_{j=1}^m$ be the set of m eigen-pairs of $\boldsymbol{B}$. The set of mn eigen-pairs of $\boldsymbol{A} \oplus \boldsymbol{B}$ is given by

$$\{\lambda_i + \mu_j, \boldsymbol{y}_j \otimes \boldsymbol{x}_i\}_{i=1,\cdots,n,\ \ j=1,\cdots,m.}$$

Theorem 6

The matrix equation

$$\boldsymbol{AX} + \boldsymbol{XB} = \boldsymbol{C} \tag{17}$$

has a unique solution for every given $\boldsymbol{C}$ if and only if

$$\lambda_i \neq -\mu_j, \text{ for all } i, j$$

where $\{\lambda_i\}_{i=1}^n$ and $\{\mu_j\}_{j=1}^m$ are the set of eigenvalues of $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively.

Kronecker Sum (cont'd)

Theorem 5 (Kronecker Sum and Eigen-Pairs)

Let $\{\lambda_i, \boldsymbol{x}_i\}_{i=1}^n$ be the set of n eigen-pairs of $\boldsymbol{A}$, and $\{\mu_j, \boldsymbol{y}_j\}_{j=1}^m$ be the set of m eigen-pairs of $\boldsymbol{B}$. The set of mn eigen-pairs of $\boldsymbol{A} \oplus \boldsymbol{B}$ is given by

$$\{\lambda_i + \mu_j, \boldsymbol{y}_j \otimes \boldsymbol{x}_i\}_{i=1,\cdots,n,\ \ j=1,\cdots,m}.$$

Theorem 6

The matrix equation

$$\boldsymbol{AX} + \boldsymbol{XB} = \boldsymbol{C} \tag{17}$$

has a unique solution for every given $\boldsymbol{C}$ if and only if

$$\lambda_i \neq -\mu_j, \ \text{for all } i, j$$

where $\{\lambda_i\}_{i=1}^n$ and $\{\mu_j\}_{j=1}^m$ are the set of eigenvalues of $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively.

Notice that, if $\lambda_i = -\mu_j$ for some i, j, then from Theorem 5 there exists a zero eigenvalue for the matrix $\boldsymbol{A} \oplus \boldsymbol{B}$.

Kronecker Sum (cont'd)

Consider a special case of Eq. (17):

$$\boldsymbol{A}^H\boldsymbol{X} + \boldsymbol{X}\boldsymbol{A} = \boldsymbol{C}, \tag{18}$$

which is known as the **Lyapunov[3] equations**.

- From Theorem 6, the Lyapunov equations admit a unique solution if

$$\lambda_i \neq -\lambda_j^*, \text{ for all } i, j \tag{19}$$

- If $\boldsymbol{A}$ is PD such that λ_i are real and positive, then the Lyapunov equations always have a unique solution.

[3] Aleksandr Lyapunov (1857–1918): Russian mathematician, a student (junior to Markov) of Chebyshev. In the theory of probability, he generalized the works of Chebyshev and Markov, and proved the Central Limit Theorem under more general conditions than his predecessors. The method of characteristic functions he used for the proof later found widespread use in probability theory.

The Solution to the Linear Matrix Equation

Finally, the solution to the linear matrix equation shown in Eq. (17) is

$$\boldsymbol{X} = \mathrm{vec}^{-1}\left(\left(\boldsymbol{A} \oplus \boldsymbol{B}^T\right)^{-1} \mathrm{vec}(\boldsymbol{C})\right). \tag{20}$$

For more details on Kronecker product and sum, please refer to Chapter 7 of [R2].[4]

[4][R2] D. S. Bernstein, *Matrix Mathematics*, 2nd Ed., Princeton University Press, 2009.

Table of Contents

The Nearest Kronecker Product (NKP) Problem

Suppose $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is given with $m = m_1 \times m_2$ and $n = n_1 \times n_2$. For these integer factorizations the nearest Kronecker product (NKP) problem involves minimizing

$$\phi(\boldsymbol{B}, \boldsymbol{C}) \triangleq \|\boldsymbol{A} - \boldsymbol{B} \otimes \boldsymbol{C}\|_F, \tag{21}$$

where $\boldsymbol{B} \in \mathbb{R}^{m_1 \times n_1}$ and $\boldsymbol{C} \in \mathbb{R}^{m_2 \times n_2}$.

A Toy Example

$$\phi(\boldsymbol{B},\boldsymbol{C}) = \left\| \left[\begin{array}{cc|cc} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ \hline a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \\ \hline a_{51} & a_{52} & a_{53} & a_{54} \\ a_{61} & a_{62} & a_{63} & a_{64} \end{array}\right] - \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix} \otimes \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} \right\|_F \tag{22}$$

$$= \left\| \underbrace{\begin{bmatrix} a_{11} & a_{21} & a_{12} & a_{22} \\ a_{31} & a_{41} & a_{32} & a_{42} \\ a_{51} & a_{61} & a_{52} & a_{62} \\ a_{13} & a_{23} & a_{14} & a_{24} \\ a_{33} & a_{43} & a_{34} & a_{44} \\ a_{53} & a_{63} & a_{54} & a_{64} \end{bmatrix}}_{\triangleq\, \mathcal{R}(\boldsymbol{A})} - \begin{bmatrix} b_{11} \\ b_{21} \\ b_{31} \\ b_{12} \\ b_{22} \\ b_{32} \end{bmatrix} \begin{bmatrix} c_{11} & c_{21} & c_{12} & c_{22} \end{bmatrix} \right\|_F \tag{23}$$

A Toy Example (cont'd)

It is not hard to find that

$$\mathcal{R}(\boldsymbol{A}) = \begin{bmatrix} \mathrm{vec}\left(\boldsymbol{A}_{11}\right)^T \\ \mathrm{vec}\left(\boldsymbol{A}_{21}\right)^T \\ \mathrm{vec}\left(\boldsymbol{A}_{31}\right)^T \\ \mathrm{vec}\left(\boldsymbol{A}_{12}\right)^T \\ \mathrm{vec}\left(\boldsymbol{A}_{22}\right)^T \\ \mathrm{vec}\left(\boldsymbol{A}_{32}\right)^T \end{bmatrix}. \tag{24}$$

Then, it is straightforward that Eq. (22) can be equivalently rewritten as

$$\phi(\boldsymbol{B}, \boldsymbol{C}) = \left\| \mathcal{R}(\boldsymbol{A}) - \mathrm{vec}(\boldsymbol{B}) \otimes \mathrm{vec}(\boldsymbol{C})^T \right\|_F. \tag{25}$$

A Toy Example (cont'd)

It is not hard to find that

$$\mathcal{R}(\boldsymbol{A}) = \begin{bmatrix} \operatorname{vec}\left(\boldsymbol{A}_{11}\right)^T \\ \operatorname{vec}\left(\boldsymbol{A}_{21}\right)^T \\ \operatorname{vec}\left(\boldsymbol{A}_{31}\right)^T \\ \operatorname{vec}\left(\boldsymbol{A}_{12}\right)^T \\ \operatorname{vec}\left(\boldsymbol{A}_{22}\right)^T \\ \operatorname{vec}\left(\boldsymbol{A}_{32}\right)^T \end{bmatrix}. \tag{24}$$

Then, it is straightforward that Eq. (22) can be equivalently rewritten as

$$\phi(\boldsymbol{B}, \boldsymbol{C}) = \left\| \mathcal{R}(\boldsymbol{A}) - \operatorname{vec}(\boldsymbol{B}) \otimes \operatorname{vec}(\boldsymbol{C})^T \right\|_F. \tag{25}$$

Clearly, the act of minimizing $\phi(\boldsymbol{B}, \boldsymbol{C})$ is equivalent to **finding a nearest rank-1 matrix to** $\mathcal{R}(\boldsymbol{A})$ (since the second term on the right of Eq. (25) is of unit rank). This problem has a simple SVD solution!

A Toy Example (cont'd)

More specifically, if the SVD of $\mathcal{R}(\boldsymbol{A})$ is given by

$$\mathcal{R}(\boldsymbol{A}) = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T, \tag{26}$$

then, the optimizing $\boldsymbol{B}$ and $\boldsymbol{C}$ are defined by

$$\mathrm{vec}(\boldsymbol{B}_{\mathrm{opt}}) = \sqrt{\sigma_1}\,\boldsymbol{U}(:,1), \tag{27}$$

$$\mathrm{vec}(\boldsymbol{C}_{\mathrm{opt}}) = \sqrt{\sigma_1}\,\boldsymbol{V}(:,1), \tag{28}$$

where the scalings $\sqrt{\sigma_1}$ are arbitrary. Indeed, if $\boldsymbol{B}_{\mathrm{opt}}$ and $\boldsymbol{C}_{\mathrm{opt}}$ solve the NKP problem and $\alpha \neq 0$, then $\alpha\boldsymbol{B}_{\mathrm{opt}}$ and $\frac{1}{\alpha}\boldsymbol{C}_{\mathrm{opt}}$ are also optimal.

Kronecker Product SVD

In general, if

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{1,1} & \cdots & \boldsymbol{A}_{1,n_1} \\ \vdots & \ddots & \vdots \\ \boldsymbol{A}_{m_1,1} & \cdots & \boldsymbol{A}_{m_1,n_1} \end{bmatrix} \in \mathbb{R}^{m_1 m_2 \times n_1 n_2}, \tag{29}$$

where each $\boldsymbol{A}_{i,j} \in \mathbb{R}^{m_2 \times n_2}$. Then, $\mathcal{R}(\boldsymbol{A})$ is defined by

$$\mathcal{R}(\boldsymbol{A}) \triangleq \begin{bmatrix} \tilde{\boldsymbol{A}}_1 \\ \vdots \\ \tilde{\boldsymbol{A}}_{n_1} \end{bmatrix} \in \mathbb{R}^{m_1 n_1 \times m_2 n_2}, \tag{30}$$

with

$$\tilde{\boldsymbol{A}}_j = \begin{bmatrix} \mathrm{vec}\left(\boldsymbol{A}_{1,j}\right)^T \\ \vdots \\ \mathrm{vec}\left(\boldsymbol{A}_{m_1,j}\right)^T \end{bmatrix} \in \mathbb{R}^{m_1 \times m_2 n_2}. \tag{31}$$

Kronecker Product SVD (cont'd)

Theorem 7 (Kronecker Product SVD)

If $\boldsymbol{A} \in \mathbb{R}^{m_1 m_2 \times n_1 n_2}$ *is blocked according to Eq.* (29) *and*

$$\mathcal{R}(\boldsymbol{A}) = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T = \sum_{k=1}^{r} \sigma_k \boldsymbol{u}_k \boldsymbol{v}_k^T, \tag{32}$$

is the SVD of $\mathcal{R}(\boldsymbol{A})$ *with* $\boldsymbol{u}_k = \boldsymbol{U}(:,k)$, $\boldsymbol{v}_k = \boldsymbol{V}(:,k)$ *and* $\sigma_k = \boldsymbol{\Sigma}(k,k)$, *then,*

$$\boldsymbol{A} = \sum_{k=1}^{r} \sigma_k \boldsymbol{U}_k \otimes \boldsymbol{V}_k, \tag{33}$$

where $\boldsymbol{U}_k = \mathrm{reshape}(\boldsymbol{u}_k, m_1, n_1)$ *and* $\boldsymbol{V}_k = \mathrm{reshape}(\boldsymbol{v}_k, m_2, n_2)$.

Kronecker Product SVD (cont'd)

Remark 1 (Kronecker Product SVD)

The integer r in Theorem 7 is the Kronecker product rank of $\boldsymbol{A}$ given the blocking Eq. (29). Note that if $\tilde{r} \leq r$, then

$$\boldsymbol{A}_{\tilde{r}} = \sum_{k=1}^{\tilde{r}} \sigma_k \, \boldsymbol{U}_k \otimes \boldsymbol{V}_k \tag{34}$$

is the closest matrix to $\boldsymbol{A}$ (in the sense of Frobenius norm) that is the sum of $\tilde{r}$ Kronecker products. If $\boldsymbol{A}$ is large and sparse and $\tilde{r}$ is small, then the Lanzcos SVD iteration can effectively be used to compute the required singular values and vectors of $\mathcal{R}(\boldsymbol{A})$.

For more about Kronecker product computations, please see Section 12.3 of [R3].[5]

[5] [R3] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th Ed., The John Hopkins University Press, 2013.

App: hybrid analog/digital precoding in massive MIMO

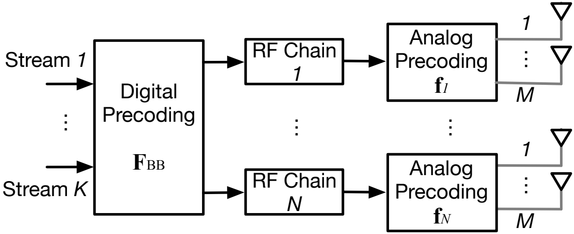

Figure 1: The sub-connected architecture of hybrid precoding systems.[6]

[6] C. Chen, Y. Wang, S. Aïssa, and M. Xia, "Low-complexity hybrid analog and digital precoding for mmWave MIMO systems," submitted to IEEE ICC'2020.

App: hybrid analog/digital precoding in massive MIMO

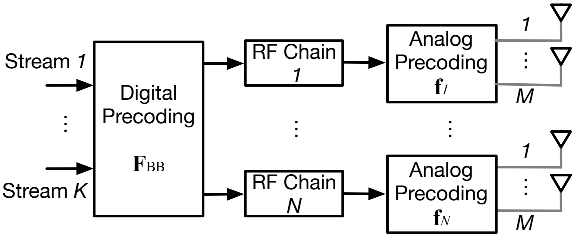

Figure 1: The sub-connected architecture of hybrid precoding systems.[6]

$$\mathbf{F}^{\mathrm{opt}} = \arg\max_{\mathbf{F}} \ \log_2\left(\left|\mathbf{I}_{N_r} + \frac{\rho}{K\sigma^2}\mathbf{H}\mathbf{F}\mathbf{F}^H\mathbf{H}^H\right|\right), \tag{35a}$$

$$\text{s.t. } \mathbf{F} = \mathbf{F}_{\mathrm{BB}} \otimes \mathbf{f}_{\mathrm{RF}}, \tag{35b}$$

$$\|\mathbf{F}\|_F^2 = K, \tag{35c}$$

$$\mathbf{f}_{\mathrm{RF}} \in \mathcal{F}. \tag{35d}$$

[6] C. Chen, Y. Wang, S. Aïssa, and M. Xia, "Low-complexity hybrid analog and digital precoding for mmWave MIMO systems," submitted to IEEE ICC'2020.

Table of Contents

Derivative of the Vectorization of a Matrix

Given the transformation

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X}\boldsymbol{B}, \tag{36}$$

where $\boldsymbol{Y}$, $\boldsymbol{A}$, $\boldsymbol{X}$ and $\boldsymbol{B} \in \mathbb{R}^{n \times n}$. By recalling (9), the derivative of the vectorization of the matrix $\boldsymbol{Y}$ is

$$\frac{\partial \mathrm{vec}(\boldsymbol{Y})}{\partial \mathrm{vec}(\boldsymbol{X})} = (\boldsymbol{B}^T \otimes \boldsymbol{A})^T = \boldsymbol{B} \otimes \boldsymbol{A}^T. \tag{37}$$

Derivative of the Vectorization of a Matrix

Given the transformation

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X}\boldsymbol{B}, \tag{36}$$

where $\boldsymbol{Y}$, $\boldsymbol{A}$, $\boldsymbol{X}$ and $\boldsymbol{B} \in \mathbb{R}^{n \times n}$. By recalling (9), the derivative of the vectorization of the matrix $\boldsymbol{Y}$ is

$$\frac{\partial \mathrm{vec}(\boldsymbol{Y})}{\partial \mathrm{vec}(\boldsymbol{X})} = (\boldsymbol{B}^T \otimes \boldsymbol{A})^T = \boldsymbol{B} \otimes \boldsymbol{A}^T. \tag{37}$$

The Jacobian of this transformation is

$$|\boldsymbol{J}| = \left| \frac{\partial \mathrm{vec}(\boldsymbol{Y})}{\partial \mathrm{vec}(\boldsymbol{X})} \right|^{-1} = |\boldsymbol{B}|^{-n} |\boldsymbol{A}|^{-n}, \tag{38}$$

where the operator $|\cdot|$ refers to the determinant of a matrix.

Derivative of the Vectorization of a Matrix

Example 8

Given the transformation

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X}\boldsymbol{B}, \tag{39}$$

where

$$\boldsymbol{A} = \begin{bmatrix} 2 & -4 \\ -1 & 3 \end{bmatrix} \quad \text{and} \quad \boldsymbol{B} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}, \tag{40}$$

find the Jacobian of this transformation.

Example (cont'd)

Solution:

1) **Direct method**:

$$
\begin{aligned}
\boldsymbol{X} &= \boldsymbol{A}^{-1}\boldsymbol{Y}\boldsymbol{B}^{-1} && (41)\\
&= \frac{1}{2}\begin{bmatrix} 3y_1 + 4y_2 - 3y_3 - 4y_4 & -3y_1 - 4y_2 + 6y_3 + 8y_4 \\ y_1 + 2y_2 - y_3 - 2y_4 & -y_1 - 2y_2 + 2y_3 + 4y_4 \end{bmatrix}, && (42)
\end{aligned}
$$

so that

$$
|\boldsymbol{J}| = \left|\frac{\partial \mathrm{vec}(\boldsymbol{X})}{\partial \mathrm{vec}(\boldsymbol{Y})}\right| = \left(\frac{1}{2}\right)^4 \begin{vmatrix} 3 & 1 & -3 & -1 \\ 4 & 2 & 4 & -2 \\ -3 & -1 & 6 & 2 \\ -4 & -2 & 8 & 4 \end{vmatrix} = \frac{1}{4}. \tag{43}
$$

Example (cont'd)

Solution:

1) **Direct method**:

$$\boldsymbol{X} = \boldsymbol{A}^{-1}\boldsymbol{Y}\boldsymbol{B}^{-1} \tag{41}$$

$$= \frac{1}{2}\begin{bmatrix} 3y_1 + 4y_2 - 3y_3 - 4y_4 & -3y_1 - 4y_2 + 6y_3 + 8y_4 \\ y_1 + 2y_2 - y_3 - 2y_4 & -y_1 - 2y_2 + 2y_3 + 4y_4 \end{bmatrix}, \tag{42}$$

so that

$$|\boldsymbol{J}| = \left|\frac{\partial \text{vec}(\boldsymbol{X})}{\partial \text{vec}(\boldsymbol{Y})}\right| = \left(\frac{1}{2}\right)^4 \begin{vmatrix} 3 & 1 & -3 & -1 \\ 4 & 2 & 4 & -2 \\ -3 & -1 & 6 & 2 \\ -4 & -2 & 8 & 4 \end{vmatrix} = \frac{1}{4}. \tag{43}$$

2) **Using** (38), it is easy to get $|\boldsymbol{A}| = 2$ and $|\boldsymbol{B}| = 1$, thus $|\boldsymbol{J}| = \frac{1}{4}$.

Derivative of a Matrix

1) **Derivative of a matrix w.r.t. a scalar** x_{rs}:

$$\frac{\partial \boldsymbol{Y}}{\partial x_{rs}} = \begin{bmatrix} \frac{\partial y_{11}}{\partial x_{rs}} & \frac{\partial y_{12}}{\partial x_{rs}} & \cdots & \frac{\partial y_{1q}}{\partial x_{rs}} \\ \frac{\partial y_{21}}{\partial x_{rs}} & \frac{\partial y_{22}}{\partial x_{rs}} & \cdots & \frac{\partial y_{2q}}{\partial x_{rs}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_{p1}}{\partial x_{rs}} & \frac{\partial y_{p2}}{\partial x_{rs}} & \cdots & \frac{\partial y_{pq}}{\partial x_{rs}} \end{bmatrix} = \sum_{i,j} \boldsymbol{E}_{ij} \otimes \frac{\partial y_{ij}}{\partial x_{rs}}, \tag{44}$$

where $\boldsymbol{E}_{ij} \triangleq \boldsymbol{e}_i \boldsymbol{e}_j^T \in \mathbb{R}^{p \times q}$ is an **elementary matrix**. In other words, $\boldsymbol{E}_{ij}$ is defined as the matrix that has a unity in the $(i,j)^{\text{th}}$ position and all other elements are zero.

Derivative of a Matrix

1) **Derivative of a matrix w.r.t. a scalar** x_{rs}:

$$\frac{\partial \boldsymbol{Y}}{\partial x_{rs}} = \begin{bmatrix} \frac{\partial y_{11}}{\partial x_{rs}} & \frac{\partial y_{12}}{\partial x_{rs}} & \cdots & \frac{\partial y_{1q}}{\partial x_{rs}} \\ \frac{\partial y_{21}}{\partial x_{rs}} & \frac{\partial y_{22}}{\partial x_{rs}} & \cdots & \frac{\partial y_{2q}}{\partial x_{rs}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_{p1}}{\partial x_{rs}} & \frac{\partial y_{p2}}{\partial x_{rs}} & \cdots & \frac{\partial y_{pq}}{\partial x_{rs}} \end{bmatrix} = \sum_{i,j} \boldsymbol{E}_{ij} \otimes \frac{\partial y_{ij}}{\partial x_{rs}}, \tag{44}$$

where $\boldsymbol{E}_{ij} \triangleq \boldsymbol{e}_i \boldsymbol{e}_j^T \in \mathbb{R}^{p\times q}$ is an **elementary matrix**. In other words, $\boldsymbol{E}_{ij}$ is defined as the matrix that has a unity in the $(i,j)^{\text{th}}$ position and all other elements are zero.

2) **Derivative of a matrix w.r.t. a matrix**: with $\boldsymbol{X} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{Y} \in \mathbb{R}^{p\times q}$, the derivative of $\boldsymbol{Y}$ with respect to $\boldsymbol{X}$ is

$$\frac{\partial \boldsymbol{Y}}{\partial \boldsymbol{X}} = \sum_{r,s} \boldsymbol{E}_{rs} \otimes \frac{\partial \boldsymbol{Y}}{\partial x_{rs}}, \tag{45}$$

where $\boldsymbol{E}_{rs} \in \mathbb{R}^{m\times n}$. It is seen that $\partial \boldsymbol{Y}/\partial \boldsymbol{X} \in \mathbb{R}^{mp\times nq}$.

Two Properties

Given $\boldsymbol{X} \in \mathbb{R}^{m\times n}$, $\boldsymbol{Y} \in \mathbb{R}^{u\times v}$ and $\boldsymbol{Z} \in \mathbb{R}^{p\times q}$.

1) **Product rules**:

- The derivative of a **product** of matrices (now $n = u$)

$$\frac{\partial \boldsymbol{X}\boldsymbol{Y}}{\partial \boldsymbol{Z}} = \frac{\partial \boldsymbol{X}}{\partial \boldsymbol{Z}}(\boldsymbol{I}_q \otimes \boldsymbol{Y}) + (\boldsymbol{I}_p \otimes \boldsymbol{X})\frac{\partial \boldsymbol{Y}}{\partial \boldsymbol{Z}}, \tag{46}$$

  where $\boldsymbol{I}_q \in \mathbb{R}^{q\times q}$ and $\boldsymbol{I}_p \in \mathbb{R}^{p\times p}$ are **unit matrices**.
- The derivative of a **Kronecker product** of matrices

$$\frac{\partial (\boldsymbol{X} \otimes \boldsymbol{Y})}{\partial \boldsymbol{Z}} = \frac{\partial \boldsymbol{X}}{\partial \boldsymbol{Z}} \otimes \boldsymbol{Y} + [\boldsymbol{I}_p \otimes \boldsymbol{U}_1]\left[\frac{\partial \boldsymbol{Y}}{\partial \boldsymbol{Z}} \otimes \boldsymbol{X}\right][\boldsymbol{I}_q \otimes \boldsymbol{U}_2], \tag{47}$$

  where $\boldsymbol{U}_1 \in \mathbb{R}^{mu\times mu}$ and $\boldsymbol{U}_2 \in \mathbb{R}^{nv\times nv}$ are **permutation matrices**.

Two Properties

Given $\boldsymbol{X} \in \mathbb{R}^{m\times n}$, $\boldsymbol{Y} \in \mathbb{R}^{u\times v}$ and $\boldsymbol{Z} \in \mathbb{R}^{p\times q}$.

1) **Product rules**:
   - The derivative of a **product** of matrices (now $n = u$)

$$\frac{\partial \boldsymbol{X}\boldsymbol{Y}}{\partial \boldsymbol{Z}} = \frac{\partial \boldsymbol{X}}{\partial \boldsymbol{Z}}(\boldsymbol{I}_q \otimes \boldsymbol{Y}) + (\boldsymbol{I}_p \otimes \boldsymbol{X})\frac{\partial \boldsymbol{Y}}{\partial \boldsymbol{Z}}, \tag{46}$$

   where $\boldsymbol{I}_q \in \mathbb{R}^{q\times q}$ and $\boldsymbol{I}_p \in \mathbb{R}^{p\times p}$ are **unit matrices**.
   - The derivative of a **Kronecker product** of matrices

$$\frac{\partial (\boldsymbol{X} \otimes \boldsymbol{Y})}{\partial \boldsymbol{Z}} = \frac{\partial \boldsymbol{X}}{\partial \boldsymbol{Z}} \otimes \boldsymbol{Y} + [\boldsymbol{I}_p \otimes \boldsymbol{U}_1]\left[\frac{\partial \boldsymbol{Y}}{\partial \boldsymbol{Z}} \otimes \boldsymbol{X}\right][\boldsymbol{I}_q \otimes \boldsymbol{U}_2], \tag{47}$$

   where $\boldsymbol{U}_1 \in \mathbb{R}^{mu\times mu}$ and $\boldsymbol{U}_2 \in \mathbb{R}^{nv\times nv}$ are **permutation matrices**.

2) **Chain Rule**: the matrix $\boldsymbol{Z}$ is a matrix function of a matrix $\boldsymbol{X}$, that is $\boldsymbol{Z} = \boldsymbol{Z}(\boldsymbol{Y}(\boldsymbol{X}))$, then

$$\frac{\partial \boldsymbol{Z}}{\partial \boldsymbol{X}} = \left[\frac{\partial [\mathrm{vec}(\boldsymbol{Y})]^T}{\partial \boldsymbol{X}} \otimes \boldsymbol{I}_p\right]\left[\boldsymbol{I}_n \otimes \frac{\partial \boldsymbol{Z}}{\partial \mathrm{vec}(\boldsymbol{Y})}\right]. \tag{48}$$

Example 9

Find an expression for $\frac{\partial \boldsymbol{X}^{-1}}{\partial \boldsymbol{X}}$.

Solution: using $\boldsymbol{X}\boldsymbol{X}^{-1} = \boldsymbol{I}$, we obtain

$$\frac{\partial \left(\boldsymbol{X}\boldsymbol{X}^{-1}\right)}{\partial \boldsymbol{X}} = \frac{\partial \boldsymbol{X}}{\partial \boldsymbol{X}} \left(\boldsymbol{I} \otimes \boldsymbol{X}^{-1}\right) + \left(\boldsymbol{I} \otimes \boldsymbol{X}\right) \frac{\partial \boldsymbol{X}^{-1}}{\partial \boldsymbol{X}} = \boldsymbol{0}, \tag{49}$$

hence

$$\begin{aligned} \frac{\partial \boldsymbol{X}^{-1}}{\partial \boldsymbol{X}} &= -(\boldsymbol{I} \otimes \boldsymbol{X})^{-1} \frac{\partial \boldsymbol{X}}{\partial \boldsymbol{X}} \left(\boldsymbol{I} \otimes \boldsymbol{X}^{-1}\right) & (50) \\ &= -(\boldsymbol{I} \otimes \boldsymbol{X}^{-1}) \bar{\boldsymbol{U}} \left(\boldsymbol{I} \otimes \boldsymbol{X}^{-1}\right), & (51) \end{aligned}$$

where $\bar{\boldsymbol{U}} = \sum_{r,s} \boldsymbol{E}_{rs} \otimes \boldsymbol{E}_{rs}$.

Remark 2 (More about Kronecker products and matrix calculus with applications)

For more details on Kronecker products and matrix calculus with applications, please refer to [R1].[7]

[7] [R1] A. Graham, *Kronecker Products and Matrix Calculus with Applications*, Dover Publications, 2018.

Thank you
for your attention!

Kai Lu
E-mail: lukai86@mail.sysu.edu.cn
Web: http://seit.sysu.edu.cn/teacher/1801