Chapter 2
Semi-tensor Product of Matrices

Abstract Starting from bilinear functions we show an alternative way of calculating bilinear mapping than using matrix form, and then extend this new "product" to multi-linear case. This new product leads to the definition of general left semi-tensor product (STP) of matrices, which is a generalization of conventional matrix product. Then certain basic properties of STP have been revealed. Roughly speaking, all the major properties of conventional matrix product remain true for this generalized product. In the light of swap matrix, certain pseudo-commutative properties of STP are obtained, which shows one of the advantages of STP over the conventional matrix product. Finally, the bi-linearity property of the STP of two vectors is investigated.

2.1 Multilinear Function

Definition 2.1. Let V_i, $i = 0, 1, \cdots, k$ be vector spaces. A mapping $f : V_1 \times V_2 \times \cdots \times V_k \to V_0$ is called a k-linear mapping, if

$$f(X_1, \cdots, \alpha X_i + \beta Y_i, \cdots, X_k) = af(X_1, \cdots, X_i, \cdots, X_k) + bf(X_1, \cdots, Y_i, \cdots, X_k),$$
$$X_j \in V_j,\ j = 1, \cdots, k;\ Y_i \in V_i,\ \alpha, \beta \in \mathbb{R}. \tag{2.1}$$

When $V_0 = \mathbb{R}$, it is called a k-linear function.

Let $f : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$ be a bilinear function. Assume

$$f(\delta_m^i, \delta_n^j) = \mu_{i,j}, \quad i = 1, \cdots, m;\ j = 1, \cdots, n.$$

Then we can arrange $D = \{\mu_{i,j} \,|\, i = 1, \cdots, m;\ j = 1, \cdots, n\}$ into a matrix as

$$M_f = \begin{bmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1n} \\ \mu_{21} & \mu_{22} & \cdots & \mu_{2n} \\ \vdots & & & \\ \mu_{m1} & \mu_{m2} & \cdots & \mu_{mn} \end{bmatrix},$$

which is called the structure matrix of f. Now for any $X = (x_1, \cdots, x_m)^T$ and $Y = (y_1, \cdots, y_n)^T$, (Precisely, $X = \sum_{i=1}^m x_i \delta_m^i$, etc.) we have

$$f(X,Y) = \sum_{i=1}^m \sum_{j=1}^n \mu_{i,j} x_i y_j. \tag{2.2}$$

Using structure matrix, we have a matrix expression of (2.2) as

$$f(X,Y) = X^T M_f Y. \tag{2.3}$$

It was mentioned in Chapter 1 that to deal with the multi-linear functions, say 3-linear functions, the cubic matrices have been proposed by [1, 4]. Certain product rules between cubic matrices and conventional matrices have also been developed. It has some successful applications [5]. But they are rather complicated. Moreover, this approach can hardly be extended to higher-dimensional data. An attempt of using matrix to deal with higher-dimensional data is so-called multi-edge matrix, proposed by Zhang [6]. Because of the complexity, it can also hardly be used.

Alternatively, we may arrange D into a row vector by $\in (i, j; m, n)$ as

$$V_f = (\mu_{11}\ \mu_{12}\ \cdots\ \mu_{1n}\ \cdots \mu_{m1}\ \cdots\ \mu_{mn}) = V_r(M_f).$$

Note that if we split V_f into m equal-size blocks as

$$V_f = [V_1\ V_2\ \cdots\ V_m]$$

then the data in V_1 has the first index $i = 1$, in V_2 has the first index $i = 2$ and so on. Then it is clear that

$$f(X,Y) = \left(\sum_{i=1}^m x_i V_i \right) Y. \tag{2.4}$$

Here we may consider $\sum_{i=1}^m x_i V_i$ as the "product" of V_f with X.

The advantage of the "product" is it can easily be extended to higher-order case. For instance, consider $f : \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^t \to \mathbb{R}$. Let

$$\mu_{i,j,k} := f(\delta_m^i, \delta_n^j, \delta_t^k), \quad i = 1, \cdots, m,\ j = 1, \cdots, n,\ k = 1, \cdots, t.$$

Then we arrange $\{\mu_{i,j,k}\}$ into a row vector V_f under the order of $\mathrm{id}(i, j, k; m, n, t)$. Let $X \in \mathbb{R}^m$, $Y \in \mathbb{R}^n$, and $Z \in \mathbb{R}^t$. We still split V_f into m equal-size blocks as

$$V_f = [V_1\ V_2\ \cdots\ V_m],$$

then

$$V_{i_0} = [\mu_{i_0 11}\ \mu_{i_0 12}\ \cdots\ \mu_{i_0 1t}\ \mu_{i_0 n1}\ \cdots\ \mu_{i_0 nt}].$$

which is the data of the first index $i = i_0$. Using the new product, we have

$$V_f X = \sum_{i=1}^{m} V_i x_i.$$

You now can see that x_i has been multiplied to the proper segment of data. Continuing this procedure, we split $V_f X$ into n blocks as

$$V_f X = [U_1, U_2, \cdots, U_n].$$

Then you can see that we have, using new product,

$$V_f XY = \sum_{j=1}^{n} U_j y_j.$$

Now y_j has been multiplied to the proper segment of data too. Finally, we split $V_f XY$ as

$$V_f XY = [W_1, W_2, \cdots, W_t].$$

Using new product, we have

$$V_f XYZ = \sum_{k=1}^{t} W_k z_k.$$

Again, z_k is multiplied to the proper segment of data, and finally we have

$$V_f XYZ = f(X,Y,Z). \tag{2.5}$$

Based on this observation, we give the following rigorous definition for this "new product":

Definition 2.2. 1. Let $X \in \mathbb{R}^{mn}$ be a row and $Y \in \mathbb{R}^{m}$ be a column. The we split X into m equal-size blocks as $(X^1\ X^2\ \cdots\ X^m)$, such that $X^i \in \mathbb{R}^n$, $i = 1, \cdots, m$, and define the left STP of X and Y, denoted by $X \ltimes Y$, as

$$X \ltimes Y := \sum_{i=1}^{m} X^i y_i \in \mathbb{R}^n. \tag{2.6}$$

2. Let $X \in \mathbb{R}^{m}$ be a row and $Y \in \mathbb{R}^{mn}$ be a column. The we define the left STP of X and Y as

$$X \ltimes Y := \left[Y^T \ltimes X^T\right]^T \in \mathbb{R}^n. \tag{2.7}$$

Using matrix product to express a bilinear function as in (2.3) is very convenient. Unfortunately, it can hardly be used for multi-linear functions. The advantage of

(2.4) is that the data are arranged as vectors, and the product can be realized by a sequence of products between two vectors. Then it can easily be extended to 3-linear case as in (2.5), as well as to general multi-linear mapping.

Consider a function $f : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \cdots \times \mathbb{R}^{n_k} \to \mathbb{R}$. Denote

$$\mu_{i_1,i_2,\cdots,i_k} = f(\delta_{n_1}^{i_1}, \delta_{n_2}^{i_2}, \cdots, \delta_{n_k}^{i_k}), \quad i_j = 1,\cdots,n_j,\ j = 1,\cdots,k.$$

Denote by $X_j = (x_1^j,\cdots,x_{n_j}^j)^T \in \mathbb{R}^{n_j}$, $j = 1,2,\cdots,k$. Then

$$f(X_1,\cdots,X_k) = \sum_{i_1=1}^{n_1} \cdots \sum_{i_k=1}^{n_k} \mu_{i_1,i_2,\cdots,i_k} x_{i_1}^1 x_{i_2}^2 \cdots x_{i_k}^k.$$

Arrange

$$D = \left\{ \mu_{i_1,i_2,\cdots,i_k} = f(\delta_{n_1}^{i_1}, \cdots, \delta_{n_k}^{i_k}) \,\middle|\, i_j = 1,\cdots,n_i,\ i = 1,\cdots,k \right\}$$

into a row vector V_f in the order of $\mathrm{id}(i_1,i_2,\cdots,i_k;n_1,n_2,\cdots,n_k)$. Then it is easy to see that

$$f(X_1,\cdots,X_k) = ((\cdots(V_f \ltimes X_1) \ltimes X_2) \ltimes \cdots X_k)\cdots)). \tag{2.8}$$

Remark 2.1. 1. From (2.8) one sees that the left STP can search for each factor vector its corresponding indix automatically. Hence it is very convenient in dealing with multi-dimensional data.

2. It is clear from the discussion of multi-linear mapping that the "multiplier dimension" requirement for two factor matrices is particularly important. Through this book we are mainly focused on this particular case, rather than considering the product of two arbitrary matrices.

Example 2.1. 1. Let $X = [1\ 3\ 2\ 4]$ and $Y = [2\ -1]^T$. Then

$$X \ltimes Y = [1\ 3] \times 2 + [2\ 4] \times (-1) = [0\ 2].$$

2. Let $X = [1\ 2\ -1]$ and $Y = [2\ 1\ -1\ 0\ -2\ 1]^T$. Then

$$X \ltimes Y = 1 \times \begin{bmatrix} 2 \\ 1 \end{bmatrix} + 2 \times \begin{bmatrix} -1 \\ 0 \end{bmatrix} + (-1) \times \begin{bmatrix} -2 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}.$$

3. Recall Example 1.6. The volume tensor is defined as

$$v := \langle X, Y \times Z \rangle, \quad X,Y,Z \in \mathbb{R}^3.$$

Then

$$\mu_{111} = \left\langle \delta_3^1, \delta_3^1 \times \delta_3^1 \right\rangle = 0; \quad \mu_{112} = \left\langle \delta_3^1, \delta_3^1 \times \delta_3^2 \right\rangle = 0; \cdots.$$

Finally, we have the vector form of $\{\mu_{ijk} \,|\, i,j,k = 1,2,3\}$ as

$$V_v = [0\ 0\ 0\ 0\ 0\ 1\ 0\ -1\ 0\ 0\ 0\ -1\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 1\ 0\ -1\ 0\ 0\ 0\ 0\ 0].$$

Now assume $X = (1,1,-1)^T$, $Y = (2,1,-2)^T$, $Z = (-1,0,-2)^T$. Then we have

$$\begin{aligned} v(X,Y,Y) &= (((V_v \ltimes X) \ltimes Y) \ltimes Z) \\ &= (([0\ -1\ -1\ 1\ 0\ 1\ 1\ -1\ 0] \ltimes Y) \ltimes Z) \\ &= [-1\ 0\ -1] \ltimes Z \\ &= 3. \end{aligned}$$

2.2 Left Semi-tensor Product of Matrices

Let $A \in \mathscr{M}_{m\times n}$ and $B \in \mathscr{M}_{p\times q}$.

(i) If $n = p$, A and B are said to be of "equal dimension".

(ii) If $n = tp$ or $nt = p$ (where $t \in \mathbb{Z}_+$, then A and B are said to be of "multiplier dimension". If $n = tp$, we denote it by $A \succ_t B$, and if $nt = p$ we denote it by $A \prec_t B$.

(iii) Otherwise, we say A and B are of arbitrary dimension.

We use $\mathrm{Row}(A)$ ($\mathrm{Col}(A)$) for the set of rows (columns) of A, and $\mathrm{Row}_i(A)$ ($\mathrm{Col}_i(A)$) the ith row (column) of A.

Definition 2.3. Let $A \in \mathscr{M}_{m\times n}$ and $B \in \mathscr{M}_{p\times q}$, and A and B are of multiplier dimension. Then the left STP of A and B is defined as

$$A \ltimes B = \begin{bmatrix} \mathrm{Row}_1(A) \ltimes \mathrm{Col}_1(B) & \mathrm{Row}_1(A) \ltimes \mathrm{Col}_2(B) & \cdots & \mathrm{Row}_1(A) \ltimes \mathrm{Col}_q(B) \\ \mathrm{Row}_2(A) \ltimes \mathrm{Col}_1(B) & \mathrm{Row}_2(A) \ltimes \mathrm{Col}_2(B) & \cdots & \mathrm{Row}_2(A) \ltimes \mathrm{Col}_q(B) \\ \vdots & & & \\ \mathrm{Row}_m(A) \ltimes \mathrm{Col}_1(B) & \mathrm{Row}_m(A) \ltimes \mathrm{Col}_2(B) & \cdots & \mathrm{Row}_m(A) \ltimes \mathrm{Col}_q(B) \end{bmatrix} \tag{2.9}$$

Example 2.2. Let

$$X = \begin{bmatrix} 1 & 2 & -1 & 2 \\ 0 & 1 & 2 & 3 \\ 3 & 3 & 1 & 1 \end{bmatrix}, \quad Y = \begin{bmatrix} 1 & 2 \\ -1 & 3 \end{bmatrix}.$$

Then

$$X \ltimes Y = \begin{bmatrix} (1\ 2) - (-1\ 2) & 2(1\ 2) + 3(-1\ 2) \\ (0\ 1) - (2\ 3) & 2(0\ 1) + 3(2\ 3) \\ (3\ 3) - (1\ 1) & 2(3\ 3) + 3(1\ 1) \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 0 & -1 & 10 \\ -2 & -2 & 6 & 11 \\ 2 & 2 & 9 & 9 \end{bmatrix}.$$

Remark 2.2. Let $A \in \mathscr{M}_{m\times n}$ and $B \in \mathscr{M}_{p\times q}$. It follows from the definition that when $n = p$ we have

$$A \ltimes B = AB.$$

That is, when the conventional matrix product is defined the left STP of A and B coincides with their conventional product. This fact shows that the left STP is a generalization of conventional matrix product. Because of this fact, the symbol "$\ltimes$" may be omitted. We can always consider AB as $A \ltimes B$, when A and B meet the equal-dimension requirement, the product becomes conventional matrix product automatically. In the sequel, unless we want to emphasize the product is left STP, the symbol "$\ltimes$" is mostly omitted.

In Chapter 1, in addition to conventional matrix product, some other matrix products have been introduced, which contain Kronecker product, Hadamard product, and Khatri-Rao product. One sees from there that all the different matrix products satisfy two fundamental properties: associative law and distributive law. These two properties may be considered as fundamental requirements for a matrix product. Without them, the fundamental matrix algebraic structure will be destroyed. So we have to show that the left STP also satisfies these two laws. First, we give a lemma.

Lemma 2.1. *Given $A \in \mathscr{M}_{m\times n}$ and $B \in \mathscr{M}_{p\times q}$, where $n|p$ or $p|n$. Then*

$$\begin{aligned} A \ltimes B &= \begin{bmatrix} \mathrm{Row}_1(A) \ltimes B \\ \vdots \\ \mathrm{Row}_m(A) \ltimes B \end{bmatrix} \\ &= \begin{bmatrix} A \ltimes \mathrm{Col}_1(B) & \cdots & A \ltimes \mathrm{Col}_q(B) \end{bmatrix}. \end{aligned} \tag{2.10}$$

Proof. According to the definition, a straightforward computation, starting from the first (second) form of (2.10) by expanding B column by column (expanding A row by row), (2.9) follows immediately. □

Theorem 2.1. *Assume the dimensions of the matrices involved in the following equations (2.11) and (2.12) meet the requirement such that the $\ltimes$ is well defined. Then we have*

1. (Distributive Law)

$$\begin{cases} F \ltimes (aG \pm bH) = aF \ltimes G \pm bF \ltimes H, \\ (aF \pm bG) \ltimes H = aF \ltimes H \pm bG \ltimes H, \quad a,b \in R. \end{cases} \tag{2.11}$$

2. (Associative Law)

$$(F \ltimes G) \ltimes H = F \ltimes (G \ltimes H). \tag{2.12}$$

Proof. Equation (2.11) can be proved by a straightforward computation. We leave to the reader to check it. In the following we prove (2.12).

First of all, we have to show that if F, G and H have feasible dimensions for $(F \ltimes G) \ltimes H$ the dimensions are also feasible for $F \ltimes (G \ltimes H)$.

Case 1: $F \succ G$ and $G \succ H$. So the dimensions of F, G and H can be assumed as $m \times np$, $p \times qr$ and $r \times s$ respectively.

Now the dimension of $F \ltimes G$ is $m \times nqr$. It is good for $(F \ltimes G) \ltimes H$. On the other hand the dimension of $G \ltimes H$ is $p \times qs$. It is good for $F \ltimes (G \ltimes H)$.

Case 2: $F \prec G$ and $G \prec H$. So the dimensions of F, G and H can be assumed as $m \times n$, $np \times q$ and $rq \times s$ respectively.

Now the dimension of $F \ltimes G$ is $mp \times q$. It is good for $(F \ltimes G) \ltimes H$. On the other hand the dimension of $G \ltimes H$ is $npr \times s$. It is good for $F \ltimes (G \ltimes H)$.

Case 3: $F \prec G$ and $G \succ H$. So the dimensions of F, G and H can be assumed as $m \times n$, $np \times qr$ and $r \times s$ respectively.

Now the dimension of $F \ltimes G$ is $mp \times qr$. It is good for $(F \ltimes G) \ltimes H$. On the other hand the dimension of $G \ltimes H$ is $np \times qs$. It is good for $F \ltimes (G \ltimes H)$.

Case 4: $F \succ G$ and $G \prec H$. So the dimensions of F, G and H can be assumed as $m \times np$, $p \times q$ and $rq \times s$ respectively.

Now the dimension of $F \ltimes G$ is $m \times nq$. To make it feasible for $(F \ltimes G) \ltimes H$, we need

Case 4.1: $(F \ltimes G) \succ H$, i.e., $n = n'r$. It is good for $F \ltimes (G \ltimes H)$;

Case 4.2: $(F \ltimes G) \prec H$, i.e., $r = nr'$. It is good for $F \ltimes (G \ltimes H)$;

The dimension of $G \ltimes H$ is $pr \times s$. To make it feasible for $(F \ltimes G) \ltimes H$, we need

Case 4.3: $F \succ (G \ltimes H)$, i.e., $n = n'r$. It is good for $(F \ltimes G) \ltimes H$;

Case 4.4: $F \prec (G \ltimes H)$, i.e., $r = nr'$. It is good for $(F \ltimes G) \ltimes H$;

Next, we prove the associativity. We have to prove it case by case. But Case 1–3 are similar, we prove only Case 1. i.e. $F \succ G$ and $G \succ H$.

Let $F_{m\times np}$, $G_{p\times qr}$ and $H_{r\times s}$ be given. Based on Lemma 2.1 we can, without loss of generality, assume $m = 1$ and $s = 1$. Split F as

$$F = [F_1, \cdots, F_p],$$

where F_i, $i = 1, \cdots, p$ are $1 \times n$ blocks. Then

$$\begin{aligned} F \ltimes G &= \left(F_1, \cdots, F_p\right) \ltimes \begin{pmatrix} g_{11}^1 & \cdots & g_{1q}^1 & \cdots & g_{r1}^1 & \cdots & g_{rq}^1 \\ \vdots & & & & & & \\ g_{11}^p & \cdots & g_{1q}^p & \cdots & g_{r1}^p & \cdots & g_{rq}^p \end{pmatrix} \\ &= \left(\sum_{i=1}^p F_i g_{11}^i, \cdots, \sum_{i=1}^p F_i g_{1q}^i, \cdots, \sum_{i=1}^p F_i g_{r1}^i, \cdots, \sum_{i=1}^p F_i g_{rq}^i \right). \end{aligned}$$

Then we have

$$\begin{aligned} (F \ltimes G) \ltimes H &= (F \ltimes G) \ltimes \begin{pmatrix} h_1 \\ \vdots \\ h_r \end{pmatrix} \\ &= \left(\sum_{j=1}^r \sum_{i=1}^p F_i g_{j1}^i h_j, \cdots, \sum_{j=1}^r \sum_{i=1}^p F_i g_{jq}^i h_j \right). \end{aligned} \tag{2.13}$$

On the other hand

$$\begin{pmatrix} g_{11}^1 & \cdots & g_{1q}^1 & \cdots & g_{r1}^1 & \cdots & g_{rq}^1 \\ \vdots & & & & & & \\ g_{11}^p & \cdots & g_{1q}^p & \cdots & g_{r1}^p & \cdots & g_{rq}^p \end{pmatrix} \ltimes \begin{pmatrix} h_1 \\ \vdots \\ h_r \end{pmatrix}$$
$$= \begin{pmatrix} \sum\limits_{j=1}^r g_{j1}^1 h_j & \cdots & \sum\limits_{j=1}^r g_{jq}^1 h_j \\ \vdots & & \\ \sum\limits_{j=1}^r g_{j1}^p h_j & \cdots & \sum\limits_{j=1}^r g_{jq}^p h_j \end{pmatrix}.$$

Then

$$\begin{aligned} F \ltimes (G \ltimes H) &= (F_1, \cdots, F_p) \ltimes (G \ltimes H) \\ &= \left(\sum_{j=1}^r \sum_{i=1}^p F_i g_{j1}^i h_j, \ \cdots, \ \sum_{j=1}^r \sum_{i=1}^p F_i g_{jq}^i h_j \right), \end{aligned}$$

which is the same as equation (2.12).

Since Case 4.1–4.4 are similar, we prove Case 4.1 only.

Let $F_{m\times npr}$, $G_{p\times q}$ and $H_{rq\times s}$ be given. We also assume $m = 1$ and $s = 1$. Split F as

$$F = (F_{11}, \cdots, F_{1r}, \cdots, F_{p1}, \cdots, F_{pr}),$$

where each F_{ij}, $i = 1, \cdots, p,\ j = 1, \cdots, r$ are $1 \times n$ blocks.

$$G = \begin{pmatrix} g_{11} & \cdots & g_{1q} \\ \vdots & & \\ g_{p1} & \cdots & g_{pq} \end{pmatrix}, \quad H = (h_{11}, \cdots, h_{1r}, \cdots, h_{q1}, \cdots, h_{qr})^T.$$

A careful computation shows that

$$(F \ltimes G) \ltimes H = F \ltimes (G \ltimes H) = \sum_{i=1}^p \sum_{j=1}^r \sum_{k=1}^q F_{ij} g_{ik} h_{kj}.$$

□

In the above proof, we leave some cases for the reader to verify.

Before exploring more properties of left STP, we give an example, which may provide a convincing reason for this generalization of matrix product.

Example 2.3. Let $X, Y, Z, W \in \mathbb{R}^n$. Then

$$A := (XY^T)(ZW^T) \in \mathscr{M}_{n\times n}$$

is a well defined matrix. Noticing that the matrix product is associative and $Y^T Z$ is a scalar, we can do the following computation:

$$A = XY^T ZW^T = X(Y^T Z)W^T = (Y^T Z)XW^T = Y^T(ZX)W^T. \tag{2.14}$$

Now we meet a puzzle: What is expression ZX in (2.14)? The puzzle shows that the conventional matrix product has a "bug". Because illegal item can be reached through legal procedure.

Since the conventional matrix product can be considered as a special case of left STP of matrices, we may ignore the conventional product and consider all the products involved in this example is the left STP. Then we leave for the reader to verify that

$$A = Y^T \ltimes (Z \ltimes X) \ltimes W^T. \tag{2.15}$$

2.3 Fundamental Properties

This section provides fundamental properties of left STP. For statement ease, we simply call it the STP. Through this book the default STP is the left STP.

Proposition 2.1. *1. Let $X \in \mathbb{R}^m$ and $Y \in \mathbb{R}^n$ be two column vectors. Then $X \ltimes Y$ is well defined. Moreover,*

$$X \ltimes Y = X \otimes Y. \tag{2.16}$$

2. Let $\omega \in \mathbb{R}^m$ and $\sigma \in \mathbb{R}^n$ be two row vectors. Then $\omega \ltimes \sigma$ is well defined. Moreover,

$$\omega \ltimes \sigma = \sigma \otimes \omega. \tag{2.17}$$

Proof. Both equalities can be verified by direct computations. □

This proposition is simple but useful. It converts Kronecker product of vectors into STP. We give some examples to show how to use it.

Example 2.4. 1. If X is a row vector or a column vector, then

$$X^k := \underbrace{X \ltimes X \ltimes \cdots \ltimes X}_{k}$$

is always well defined.

2. Recall the tensor formula (1.44) in Chapter 1. Using (2.16) and (2.17), we have

$$f(X_1, \cdots, X_s, \omega_1, \cdots, \omega_t) = \omega_t \ltimes \cdots \ltimes \omega_1 M_f X_1 \ltimes \cdots \ltimes X_s. \tag{2.18}$$

The advantage of (2.18) over (1.44) is, (2.18) involves only one matrix product, and since the STP has associative property, no parentheses are necessary any more. Later on, you can see that in the form of (2.18), a tensor will be much easier manipulated.

3. Let $X \in \mathbb{R}^n$, $Y \in \mathbb{R}^q$ be two columns and $A \in \mathcal{M}_{m\times n}$, $B \in \mathcal{M}_{p\times q}$ be two given matrices. Then

$$(AX) \ltimes (BY) = (A \otimes B)(X \ltimes Y). \tag{2.19}$$

In general let $X_i \in \mathbb{R}^{n_i}$ and $A_i \in \mathcal{M}_{m_i \times n_i}$, $i = 1, \cdots, k$. Then

$$\ltimes_{i=1}^{k} (A_i X_i) = \left(\otimes_{i=1}^{k} A_i\right) \left(\ltimes_{i=1}^{k} X_i\right). \tag{2.20}$$

Particularly,

$$(AX)^k = \left(\underbrace{A \otimes \cdots \otimes A}_{k}\right) X^k. \tag{2.21}$$

4. Let $\omega \in \mathbb{R}^m$, $\sigma \in \mathbb{R}^p$ be two rows and $A \in \mathcal{M}_{m\times n}$, $B \in \mathcal{M}_{p\times q}$ be two given matrices. Then

$$(\omega A) \ltimes (\sigma B) = (\omega \ltimes \sigma)(B \otimes A). \tag{2.22}$$

In general let $\omega_i \in \mathbb{R}^{m_i}$ be rows and $A_i \in \mathcal{M}_{m_i \times n_i}$, $i = 1, \cdots, k$. Then

$$\ltimes_{i=1}^{k} (\omega_i A_i) = \left(\ltimes_{i=1}^{k} \omega_i\right) \left(\otimes_{i=1}^{k} A_{k+1-i}\right). \tag{2.23}$$

Particularly,

$$(\omega A)^k = \omega^k \left(\underbrace{A \otimes \cdots \otimes A}_{k}\right). \tag{2.24}$$

The verification of equations (2.19)-(2.24) is left to the reader.

The following proposition is about the block-multiplication rule.

Proposition 2.2. *Assume $A \succ_t B$ (or $A \prec_t B$). Decompose A and B into blocks as*

$$A = \begin{pmatrix} A^{11} & \cdots & A^{1s} \\ \vdots & & \vdots \\ A^{r1} & \cdots & A^{rs} \end{pmatrix}, \quad B = \begin{pmatrix} B^{11} & \cdots & B^{1t} \\ \vdots & & \vdots \\ B^{s1} & \cdots & B^{st} \end{pmatrix}.$$

If $A^{ik} \succ_t B^{kj}$, $\forall i, j, k$ (correspondingly, $A^{ik} \prec_t B^{kj}$, $\forall i, j, k$), then

$$A \ltimes B = \begin{pmatrix} C^{11} & \cdots & C^{1t} \\ \vdots & & \vdots \\ C^{r1} & \cdots & C^{rt} \end{pmatrix}, \tag{2.25}$$

where

$$C^{ij} = \sum_{k=1}^{s} A^{ik} \ltimes B^{kj}.$$

Proof. Using the definition, a careful calculation leads to the result. □

Next, we consider the power of a matrix A.

Definition 2.4. Assume $A \in \mathcal{M}_{m\times n}$, where either $m|n$ or $n|m$. Then A^n is inductively defined as

$$\begin{cases} A^1 = A \\ A^{k+1} = A^k \ltimes A, \quad k = 1, 2, \cdots. \end{cases} \tag{2.26}$$

It is easy to see that A^k is well defined. Moreover, if $m = nt$, then $A^k \in \mathcal{M}_{t^k n\times n}$, and if $mt = n$, then $A^k \in \mathcal{M}_{m\times t^k m}$.

In the following remark we briefly discuss the dimension of the STP of matrices. It is easy to verify the facts in the following remark.

Remark 2.3. 1. The dimension of the STP of matrices can be easily obtained by the common fact elimination of the second index of the leading matrix with the first index of the following matrix. For instance

$$A_{p\times qr} \ltimes B_{r\times s} \ltimes C_{qst\times l} = (A \ltimes B)_{p\times qs} \ltimes C_{qst\times l} = (A \ltimes B \ltimes C)_{pt\times l}.$$

In the first equality r is canceled and in the second equality qs is canceled. This way is obviously the generalization of the conventional matrix product. For instance, $A_{p\times s}B_{s\times q} = (AB)_{p\times q}$. It can be considered as s has been canceled.

2. Unlike conventional multiplication, even if both $A \ltimes B$ and $B \ltimes C$ are well defined, $A \ltimes B \ltimes C = (A \ltimes B) \ltimes C$ may not be defined. For instance, a counter example is $A \in M_{3\times 4}$, $B \in M_{2\times 3}$ and $C \in M_{9\times 1}$. (A general version of STP will be introduced in Chapter 4, where the STP is defined for arbitrary factor matrices.)
3. If $A \succ_s B$ ($A \prec_s B$) and $B \succ_t C$ ($B \prec_t C$), then $A \ltimes B \succ_{st} C$ (Correspondingly, $A \ltimes B \prec_{st} C$). Hence if $A_1 \prec A_2 \prec \cdots \prec A_k$ or $A_1 \succ A_2 \succ \cdots \succ A_k$, then $\ltimes_{i=1}^k A_i$ is well defined.
4. Let $p \geq 2$ be an integer. Define a set of matrices as

$$\mathcal{M}^p := \cup_{i,j\in\mathbb{Z}_+} \mathcal{M}_{p^i\times p^j}.$$

Then it is easy to see that $\ltimes : \mathcal{M}^p \times \mathcal{M}^p \to \mathcal{M}^p$ is always well defined and closed. When the p-valued logic is considered ($p = 2$ is the standard logic), the matrices met there are of this form.

Proposition 2.3. *Assume $A \ltimes B$ is well defined, then*

$$(A \ltimes B)^{\mathrm{T}} = B^{\mathrm{T}} \ltimes A^{\mathrm{T}}. \tag{2.27}$$

Proof. Assume X is a row, Y is a column and $X \ltimes Y$ is well defined, then a straightforward computation shows that

$$X \ltimes Y = \left[Y^{\mathrm{T}} \ltimes X^{\mathrm{T}}\right]^{\mathrm{T}}. \tag{2.28}$$

Now consider $A \ltimes B$. By definition, the (i,j)th block of $A \ltimes B$ is

$$\mathrm{Row}_i(A) \ltimes \mathrm{Col}_j(B).$$

Meanwhile, the (j,i)th block of $B^{\mathrm{T}} \ltimes A^{\mathrm{T}}$ is

$$\mathrm{Row}_j(B^{\mathrm{T}}) \ltimes \mathrm{Col}_i(A^{\mathrm{T}}) = [\mathrm{Col}_j(B)]^{\mathrm{T}} \ltimes [\mathrm{Row}_i(A)]^{\mathrm{T}}.$$

According to (2.28),

$$\left[[\mathrm{Col}_j(B)]^{\mathrm{T}} \ltimes [\mathrm{Row}_i(A)]^{\mathrm{T}}\right]^{\mathrm{T}} = \mathrm{Row}_i(A) \ltimes \mathrm{Col}_j(B).$$

That is, the transpose of (i,j)th block $A \ltimes B$ is the (j,i)th block of $B^{\mathrm{T}} \ltimes A^{\mathrm{T}}$. □

The following proposition shows that the STP of two matrices can easily be realized by using conventional product plus Kronecker product.

Proposition 2.4. *1. If $A \succ_t B$, then*

$$A \ltimes B = A(B \otimes I_t). \tag{2.29}$$

2. If $A \prec_t B$, then

$$A \ltimes B = (A \otimes I_t)B. \tag{2.30}$$

Proof. We prove (2.29) only. The proof of (2.30) is similar. Say, $B \in \mathscr{M}_{p\times q}$. Then

$$B \otimes I_t = [\mathrm{Col}_1(B) \otimes I_t \ \mathrm{Col}_2(B) \otimes I_t \ \cdots \ \mathrm{Col}_q(B) \otimes I_t\ ].$$

Using this form and the Proposition 2.2, we can, without lose of generality, assume A is a row and B is a column. Then a straightforward computation verifies the equalities. □

Proposition 2.4 is of particular importance. Many properties can easily be obtained via (2.29) and (2.30). In fact, it can be considered as an alternative definition of left STP of matrices.

Proposition 2.5. *Assume A and B are square matrices and both $A \ltimes B$ and $B \ltimes A$ are well defined, then*

1. $A \ltimes B$ and $B \ltimes A$ have the same characteristic functions.
2.

$$\mathrm{tr}(A \ltimes B) = \mathrm{tr}(B \ltimes A). \tag{2.31}$$

3. If at least one of A and B is invertible, then

$$A \ltimes B \sim B \ltimes A, \tag{2.32}$$

where "$\sim$" stands for the similarity of two matrices.

4. *If both A and B are upper triangular (lower triangular, diagonal, orthogonal), then $A \ltimes B$ is also upper triangular (lower triangular, diagonal, orthogonal).*
5. *Assume both A and B are invertible and $A \ltimes B$ is well defined, then*

$$(A \ltimes B)^{-1} = B^{-1} \ltimes A^{-1}. \tag{2.33}$$

6. *If $A \prec_t B$, then*

$$\det(A \ltimes B) = [\det(A)]^t \det(B). \tag{2.34}$$

If $A \succ_t B$, then

$$\det(A \ltimes B) = \det(A)[\det(B)]^t. \tag{2.35}$$

Proof. Using (2.29) and (2.30) to convert the STP into conventional product plus Kronecker product, then the above properties can be easily obtained via known properties of either conventional or Kronecker products. As an example, we show item 5: Assume $A \prec_t B$, then

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

□

The STP of a matrix with an identity matrix has some special properties. Roughly speaking, if the size of I_k is larger than the size of matrix M (comparing the column number of the first factor with the row number of the second factor), then it will enlarge M, otherwise, it keeps M unchanged.

Proposition 2.6. *1. Let $M \in \mathcal{M}_{m \times pn}$. Then*

$$M \ltimes I_n = M. \tag{2.36}$$

2. Let $M \in \mathcal{M}_{m \times n}$. Then

$$M \ltimes I_{pn} = M \otimes I_p. \tag{2.37}$$

3. Let $M \in \mathcal{M}_{pm \times n}$. Then

$$I_p \ltimes M = M. \tag{2.38}$$

4. Let $M \in \mathcal{M}_{m \times n}$. Then

$$I_{pm} \ltimes M = I_p \otimes M. \tag{2.39}$$

Proof. All the equations follow from Proposition 2.4 immediately. □

2.4 Swapping Product Factors

One major inferior of matrix product to scalar product is that it is not commutative. Using swap matrix etc., the STP can change the order of its factors in certain sense. We call these properties the pseudo-commutativity. It is very useful in applications. The following proposition is a re-statement of Corollary **??**.

Proposition 2.7. *1. Let $X \in \mathbb{R}^m$ and $Y \in \mathbb{R}^n$ be two column vectors. Then*

$$W_{[m,n]}XY = YX. \tag{2.40}$$

2. Let $\omega \in \mathbb{R}^m$ and $\sigma \in \mathbb{R}^n$ be two row vectors. Then

$$\omega\sigma W_{[m,n]} = \sigma\omega. \tag{2.41}$$

Equations (2.40) and (2.41) may tell you why $W_{[m,n]}$ is called a swap matrix. Later on, you will see that (2.40) and (2.41) are extremely useful, because they are used to overcome the non-commutative shortage of the matrix product.

The following is a generalization of Proposition 2.7.

Proposition 2.8. *1. Let $X_i \in \mathbb{R}^{n_i}$, $i = 1, \cdots, k$ be k column vectors. Then we can use swap matrix to swap the factors X_t with X_{t+1} in the product in $\ltimes_{i=1}^{k} X_i$.*

$$\begin{aligned} &\left[I_{\prod_{j=1}^{t-1} n_j} \otimes W_{[n_t,n_{t+1}]} \otimes I_{\prod_{j=t+1}^{k} n_j} \right] X_1 X_2 \cdots X_k \\ &= X_1 \cdots X_{t-1} X_{t+1} X_t X_{t+2} \cdots X_k. \end{aligned} \tag{2.42}$$

2. Similarly, let $\omega_i \in \mathbb{R}^{n_i}$, $i = 1, \cdots, k$ be k row vectors. Then we have

$$\begin{aligned} &\omega_1 \omega_2 \cdots \omega_k \left[I_{\prod_{j=t+1}^{k} n_j} \otimes W_{[n_t,n_{t+1}]} \otimes I_{\prod_{j=1}^{t-1} n_j} \right] \\ &= \omega_1 \cdots \omega_{t-1} \omega_{t+1} \omega_t \omega_{t+2} \cdots \omega_k. \end{aligned} \tag{2.43}$$

Proof. We prove (2.42) only. The proof of (2.43) is similar.

$$\begin{aligned} LHS &= \left[I_{\prod_{j=1}^{t-1} n_j} \otimes W_{[n_t,n_{t+1}]} \otimes I_{\prod_{j=t+1}^{k} n_j} \right] \\ &\quad \times [(X_1 \cdots X_{t-1}) \otimes X_t X_{t+1} \otimes (X_{t+2} \cdots X_k)] \\ &= \left[I_{\prod_{j=1}^{t-1} n_j} \times (X_1 \cdots X_{t-1}) \right] \otimes \left[W_{[n_t,n_{t+1}]} \times X_t X_{t+1} \right] \\ &\quad \otimes \left[I_{\prod_{j=t+1}^{k} n_j} \times (X_{t+2} \cdots X_k) \right] \\ &= (X_1 \cdots X_{t-1}) \otimes (X_{t+1} X_t) \otimes (X_{t+2} \cdots X_k) \\ &= (X_1 \cdots X_{t-1}) \ltimes (X_{t+1} X_t) \ltimes (X_{t+2} \cdots X_k) = RHS. \end{aligned}$$

□

In fact, the swap matrix can also be used to exchange the order of blocks in a matrix. The following is a further generalization of Proposition 2.7, or a generalization of Proposition 2.8.

Proposition 2.9. *1. Assume a matrix A is split as a block-row as*

$$A = [A_{11}, \cdots, A_{1n}, \cdots, A_{m1}, \cdots, A_{mn}],$$

where each block has the same size. Moreover, the blocks are ordered by multi-index $\mathrm{id}(i, j; m, n)$. *Then*

$$AW_{[n,m]} = [A_{11}, \cdots, A_{m1}, \cdots, A_{1n}, \cdots, A_{mn}], \tag{2.44}$$

in which the blocks are ordered by multi-index $\mathrm{id}(j, i; n, m)$.

2. Let

$$B = \left[B_{11}^T, \cdots, B_{1n}^T, \cdots, B_{m1}^T, \cdots, B_{mn}^T\right]^T$$

be a block-column, in which the equal-size blocks are ordered by the multi-index $\mathrm{id}(i, j; m, n)$. *Then*

$$W_{[m,n]}B = \left[B_{11}^T, \cdots, B_{m1}^T, \cdots, B_{1n}^T, \cdots, M_{mn}^T\right]^T, \tag{2.45}$$

in which the blocks are ordered by multi-index $\mathrm{id}(j, i; n, m)$.

We leave the proof to the reader.

In the following we consider the swap of matrices with vectors. we need some auxiliary properties.

Lemma 2.2. *1. Let Z be a t-dimensional row vector and* $A \in \mathscr{M}_{m\times n}$. *Then*

$$ZW_{[m,t]}A = AZW_{[n,t]} = A \otimes Z. \tag{2.46}$$

2. Let Y be a t-dimensional column vector and $A \in \mathscr{M}_{m\times n}$. *Then*

$$AW_{[t,n]}Y = W_{[t,m]}YA = A \otimes Y. \tag{2.47}$$

Proof. 1. Using Equation (1.60) in Exercise 1, a direct computation shows that

$$ZW_{[m,t]} = \sum_{j=1}^{t} z_j I_m \otimes (\delta_n^j)^T = I_m \otimes Z. \tag{2.48}$$

Using Proposition 2.4, we have

$$ZW_{[m,t]}A = (I_m \otimes Z)A = (I_m \otimes Z)(A \otimes I_t) = A \otimes Z.$$

Similarly, we have

$$AZW_{[n,t]} = A(I_n \otimes Z) = (A \otimes I_1)(I_n \otimes Z) = A \otimes Z.$$

2. Starting from (2.46), we replace A by A^{T} and replace Z by Y^{T}, and then take transpose on both sides. Noting that $W_{[m,n]}^{\mathrm{T}} = W_{[n,m]}$, (2.47) follows immediately. □

Lemma 2.3. *Let $A \in \mathcal{M}_{m\times n}$ and $X \in \mathcal{M}_{n\times q}$. Then*

$$V_r(AX) = A \ltimes V_r(X); \tag{2.49}$$

and

$$V_c(AX) = (I_q \otimes A)V_c(X). \tag{2.50}$$

We leave the proves of (2.49) and (2.50) to the reader.

The following lemma is very useful.

Lemma 2.4. *Let $A \in \mathcal{M}_{m\times n}$. Then*

$$W_{[m,q]} \ltimes A \ltimes W_{[q,n]} = I_q \otimes A. \tag{2.51}$$

Proof. Let $X \in M_{n\times q}$. According to (2.49) we have

$$V_r(AX) = A \ltimes V_r(X) = A \ltimes W_{[q,n]}V_c(X). \tag{2.52}$$

Multiplying both sides of (2.52) by $W_{[m,q]}$ yields

$$V_c(AX) = (W_{[m,q]} \ltimes A \ltimes W_{[q,n]})V_c(X). \tag{2.53}$$

Comparing (2.50) with (2.53) and taking into consideration that the entries of X are arbitrary, it is clear that (2.51) is true. □

Now we are ready to present the following result, which may be considered as the pseudo-commutativity between matrices and vectors.

Proposition 2.10. *Given $A \in \mathcal{M}_{m\times n}$.*

1. Let $\omega \in \mathbb{R}^t$ be a row vector. Then

$$A\omega = \omega W_{[m,t]}AW_{[t,n]} = \omega(I_t \otimes A). \tag{2.54}$$

2. Let $Z \in \mathbb{R}^t$ be a column vector. Then

$$ZA = W_{[m,t]}AW_{[t,n]}Z = (I_t \otimes A)Z. \tag{2.55}$$

3. Let $X \in \mathbb{R}^m$ be a column vector. Then

$$X^{\mathrm{T}}A = [V_r(A)]^{\mathrm{T}}X. \tag{2.56}$$

4. Let $Y \in \mathbb{R}^n$ be a column vector. Then

$$AY = Y^{\mathrm{T}}V_c(A). \tag{2.57}$$

5. *Let $X \in \mathbb{R}^m$ be a column vector and $\omega \in \mathbb{R}^n$ be a row vector. Then*

$$X\omega = \omega W_{[m,n]} X. \tag{2.58}$$

Proof. Right multiplying both sides of the first equality of (2.46) by $W_{[t,n]}$, and using the fact that $W_{[n,t]}^{-1} = W_{[t,n]}$ yield the first equality of (2.54). From this first equality and using (2.51), we have the second equality.

Similarly, left multiplying both sides of the first equality of (2.47) by $W_{[m,t]}$ yields the first equality of (2.55). Applying (2.51) to the first equality yields the second equality.

We leave the proves of (2.56), (2.57), and (2.58) to the reader.

Note that $W_{[1,t]} = W_{[t,1]} = I_t$. Then (2.58) is an immediate consequence of (2.54) or (2.55).

The following result "swaps" two factors of a Kronecker product.

Proposition 2.11. *Let $A \in \mathscr{M}_{m\times n}$ and $B \in \mathscr{M}_{s\times t}$. Then*

$$A \otimes B = W_{[s,m]} \ltimes B \ltimes W_{[m,t]} \ltimes A = (I_m \otimes B) \ltimes A. \tag{2.59}$$

Proof. Denoting $A^i := \mathrm{Row}_i(A)$, $i = 1,\cdots,m$, and $B^j := \mathrm{Row}_j(B)$, $j = 1,\cdots,s$, a straightforward computation shows

$$A \otimes B = \begin{bmatrix} a_{11}B^1 & \cdots & a_{1n}B^1 \\ a_{11}B^2 & \cdots & a_{1n}B^2 \\ \vdots & & \\ a_{11}B^s & \cdots & a_{1n}B^s \\ \vdots & & \\ a_{m1}B^1 & \cdots & a_{mn}B^1 \\ a_{m1}B^2 & \cdots & a_{mn}B^2 \\ \vdots & & \\ a_{m1}B^s & \cdots & a_{mn}B^s \end{bmatrix} = \begin{bmatrix} B^1 \ltimes A^1 \\ B^2 \ltimes A^1 \\ \vdots \\ B^s \ltimes A^1 \\ \vdots \\ B^1 \ltimes A^m \\ B^2 \ltimes A^m \\ \vdots \\ B^s \ltimes A^m \end{bmatrix}. \tag{2.60}$$

Applying (2.46) to each row of B yields

$$B \ltimes W_{[m,t]} \ltimes A = \begin{bmatrix} A \otimes B^1 \\ A \otimes B^2 \\ \vdots \\ A \otimes B^s \end{bmatrix} = \begin{bmatrix} B^1 \ltimes A^1 \\ B^1 \ltimes A^2 \\ \vdots \\ B^1 \ltimes A^m \\ \vdots \\ B^s \ltimes A^1 \\ B^s \ltimes A^2 \\ \vdots \\ B^s \ltimes A^m \end{bmatrix}. \tag{2.61}$$

Comparing (2.60) with (2.61), one sees easily that the blocks of

$$\{B^i \ltimes A^j \,|\, i = 1, \cdots, s;\ j = 1, \cdots, m\}$$

are arranged in (2.61) by the order of $\mathrm{id}(i, j; s, m)$ and arranged in (2.60) by the order of $\mathrm{id}(j, i; m, s)$. Using Proposition 2.9, the first equality of (2.59) follows.

Applying (2.51) to the first equality of (2.59), its second equality is obtained. □

We give an example for this.

Example 2.5. Assume

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix},$$

then $m = n = 2$, $s = 3$, $t = 2$, and hence we have

$$W_{[32]} = \begin{bmatrix} 1&0&0&0&0&0 \\ 0&0&1&0&0&0 \\ 0&0&0&0&1&0 \\ 0&1&0&0&0&0 \\ 0&0&0&1&0&0 \\ 0&0&0&0&0&1 \end{bmatrix}, \quad W_{[22]} = \begin{bmatrix} 1&0&0&0 \\ 0&0&1&0 \\ 0&1&0&0 \\ 0&0&0&1 \end{bmatrix},$$

$$\begin{aligned}
&W_{[32]} \ltimes B \ltimes W_{[22]} \ltimes A \\
&= \begin{bmatrix} b_{11} & b_{12} & 0 & 0 \\ b_{21} & b_{22} & 0 & 0 \\ b_{31} & b_{32} & 0 & 0 \\ 0 & 0 & b_{11} & b_{12} \\ 0 & 0 & b_{21} & b_{22} \\ 0 & 0 & b_{31} & b_{32} \end{bmatrix} \ltimes \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \\
&= \begin{bmatrix} a_{11}b_{11} & a_{11}b_{12} & a_{12}b_{11} & a_{12}b_{12} \\ a_{11}b_{21} & a_{11}b_{22} & a_{12}b_{21} & a_{12}b_{22} \\ a_{11}b_{31} & a_{11}b_{32} & a_{12}b_{31} & a_{12}b_{32} \\ a_{21}b_{11} & a_{21}b_{12} & a_{22}b_{11} & a_{22}b_{12} \\ a_{21}b_{21} & a_{21}b_{22} & a_{22}b_{21} & a_{22}b_{22} \\ a_{21}b_{31} & a_{21}b_{32} & a_{22}b_{31} & a_{22}b_{32} \end{bmatrix} = A \otimes B.
\end{aligned}$$

Finally, we show some factorization properties of the swap matrix. Later on, you will see that they are useful in data order manipulation.

Proposition 2.12. *Consider* $W_{[m,n]}$.

1. *When* n *is a composite number,* $W_{[m,n]}$ *can be factorized as*

$$W_{[p,qr]} = (I_q \otimes W_{[p,r]})(W_{[p,q]} \otimes I_r) = (I_r \otimes W_{[p,q]})(W_{[p,r]} \otimes I_q). \tag{2.62}$$

2. *When* m *is a composite number,* $W_{[m,n]}$ *can be factorized as*

$$W_{[pq,r]} = (W_{[p,r]} \otimes I_q)(I_p \otimes W_{[q,r]}) = (W_{[q,r]} \otimes I_p)(I_q \otimes W_{[p,r]}). \tag{2.63}$$

Proof. We prove (2.62) only. Let $X_1 \in \mathbb{R}^p$, $X_2 \in \mathbb{R}^q$, and $X_3 \in \mathbb{R}^r$. Then

$$W_{[p,qr]} X_1 X_2 X_3 = X_2 X_3 X_1.$$

Meanwhile,

$$(W_{[p,q]} \otimes I_r) X_1 X_2 X_3 = X_2 X_1 X_3,$$

and

$$(I_q \otimes W_{[p,r]})(W_{[p,q]} \otimes I_r) X_1 X_2 X_3 = (I_q \otimes W_{[p,r]}) X_2 X_1 X_3 = X_2 X_3 X_1.$$

That is,

$$W_{[p,qr]} X_1 X_2 X_3 = (I_q \otimes W_{[p,r]})(W_{[p,q]} \otimes I_r) X_1 X_2 X_3.$$

Since X_1, X_2 and X_3 are arbitrarily chosen, the above equation shows the first equality in (2.62). Exchanging q and r yields the second equality of (2.62).

The proof of (2.63) is similar. □

2.5 Semi-tensor Product as Bilinear Mapping

When the STP is applied to two vectors, then it becomes a bilinear mapping. This mapping has particular importance in further applications. This section is devoted to explore this.

Definition 2.5. Let E, F, G be 3 vector spaces. A mapping $\phi: E \times F \to G$ is called a bilinear mapping, if

$$\begin{aligned} \phi(aX_1+bX_2,Y) &= a\phi(X_1,Y)+b\phi(X_2,Y); \\ \phi(X,cY_1+dY_2) &= c\phi(X,Y_1)+d\phi(X,Y_2), \end{aligned} \tag{2.64}$$

where $a,b,c,d \in \mathbb{R}$, $X,X_1,X_2 \in E$, $Y,Y_1,Y_2 \in F$.

Let $\{e_1,\cdots,e_m\}$ and $\{f_1,\cdots,f_n\}$ be bases of E and F respectively. Denote by

$$t_{i,j} = \phi(e_i,f_j), \quad 1 \le i \le m; 1 \le j \le n,$$

and let

$$T = \mathrm{Span}\{t_{i,j}\} \subset G.$$

Then T is the smallest subspace containing $\mathrm{im}(\phi)$. Assume

$$\left\{ t_{i,j} \,\middle|\, 1 \le i \le m; 1 \le j \le n \right\}$$

are linearly independent, then they form a basis of T. Arrange them into a matrix form by using multi-index $\mathrm{id}(i,j;m,n)$ as

$$B_T = (t_{11},t_{12},\cdots,t_{1n},\cdots,t_{m1},\cdots,t_{mn}).$$

Let

$$X = \sum_{i=1}^{m} x_i e_i \in E; \quad Y = \sum_{j=1}^{n} y_j f_j \in F.$$

Then

$$\phi(X,Y) = \sum_{i=1}^{m} x_i \sum_{j=1}^{n} y_j \phi(e_i,f_j) = B_t(x_i,\cdots,x_m)^T \ltimes (y_1,\cdots,y_n)^T.$$

Simply express a vector by its coefficients as $X \sim (x_i,\cdots,x_m)^t$ etc. Then we have $\phi: E \times F \to T$ is described as

$$\phi(X,Y) = B_t \ltimes X \ltimes Y. \tag{2.65}$$

Note that $\mathrm{im}(\phi) \ne T$. Particularly, we would like to emphasize that $\mathrm{im}(\phi)$ is not a vector space. For instance, assume $E = F = \mathbb{R}^2$ with their canonical basis $\{\delta_2^1, \delta_2^2\}$. Then

$$t_{i,j} = \delta_2^i \ltimes \delta_2^j, \quad i,j = 1,2.$$

Note that $\{t_{i,j} | i = 1,2,\ j = 1,2\}$ are linearly independent, and

$$T = \text{Span}\left\{t_{i,j} | i = 1,2,\ j = 1,2\right\} = \mathbb{R}^4.$$

Then it is easy to verify that

$$\text{im}(\phi) = \left\{ z \in \mathbb{R}^4 \,\middle|\, z_1 z_4 = z_2 z_3 \right\}.$$

Particularly, we use the STP as a mapping $\ltimes : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^{mn}$, defined in a natural way as in (2.65). Denote by $Z = X \ltimes Y \in \mathbb{R}^{mn}$ and label the entries of Z by $\text{id}(i,j;m,n)$. Then the image is

$$\text{im}(\ltimes) = \left\{ Z = (z_{11}, \cdots, z_{1n}, \cdots, z_{mn})^T \,\middle|\, z_{i,p} z_{j,q} = z_{i,q} z_{j,p} \right\}.$$

That is, there are

$$\binom{m}{2}\binom{n}{2} = \frac{m(m-1)n(n-1)}{4}$$

constrained equations, which are

$$z_{i,p} z_{j,q} = z_{i,q} z_{j,p}, \quad 1 \le i \ne j \le m;\ 1 \le p \ne q \le n.$$

Definition 2.6. Let E and F be two finite dimensional vector spaces. A bilinear mapping $\otimes : E \times F \to T$, where $T \supset \text{im}(\otimes)$, is called a universal mapping, if for any bilinear mapping $\phi : E \times F \to H$, there exists a unique mapping $f : T \to H$, such that the graph 2.1 is commutative. That is,

$$\phi = f \circ \otimes.$$

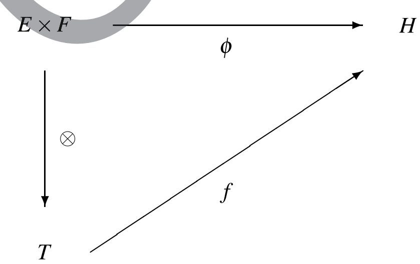

Fig. 2.1 Universal Bilinear Mapping

Proposition 2.13. *The STP* $\ltimes : E \times F \to T$ *is a universal bilinear mapping, there* T *is the space generated by the image of* $\ltimes$.

Proof. Let $\phi : E \times F \to H$ be a bilinear mapping. Then

$$\phi(X,Y) = M_\phi XY, \quad X \in E,\ Y \in F.$$

where M_ϕ is the structure matrix of ϕ. Define $f : T \to H$ as a linear mapping $Z \mapsto M_\phi Z$, Then $\phi = f \circ \ltimes$. The uniqueness of f comes from the fact that the structure matrix of a bilinear mapping is unique. □

In the following we prove a property concerning about the STP as a bilinear mapping.

Proposition 2.14. *Let $X_1, \cdots, X_k \in E$, $Y_1, \cdots, Y_k \in F$, and*

$$\sum_{i=1}^{k} X_i \ltimes Y_i = 0. \tag{2.66}$$

1. *If $Y_1, \cdots, Y_k$ are linearly independent, then $X_1 = \cdots = X_k = 0$.*
2. *If $X_1, \cdots, X_k$ are linearly independent, then $Y_1 = \cdots = Y_k = 0$.*

Proof. 1. Denote

$$X_i = (x_i^1, \cdots, x_i^m)^T, \quad i = 1, \cdots, k.$$

Then

$$\sum_{i=1}^{k} X_i \ltimes Y_i = \begin{bmatrix} \sum_{i=1}^{k} x_i^1 Y_i \\ \vdots \\ \sum_{i=1}^{k} x_i^m Y_i \end{bmatrix} = 0.$$

The conclusion follows.

2.

$$\sum_{i=1}^{k} Y_i \ltimes X_i = W_{[m,n]} \sum_{i=1}^{k} X_i \ltimes Y_i = 0.$$

The above conclusion implies this one.

□

We refer to [3] for a completed description of multi-linear mappings.

Exercise 2

1. Let $f : V_1 \times V_2 \times \cdots \times V_k \to V_0$ be a multi-linear mapping. Denote by $(e_1^i, \cdots, e_{n_i}^i)$ the basis of V_i, $i = 0, 1, \cdots, k$.
(i) Give the matrix expression of f. That is, find a matrix M_f, which is called the structure matrix of f, such that

$$X_0 = f(X_1, \cdots, X_k) = M_f \ltimes_{i=1}^{k} X_i, \quad X_i \in V_i,\ i = 0, 1, \cdots, k.$$

(ii) Let $(\tilde{e}^i_1, \cdots, \tilde{e}^i_{n_i})$ be another basis of V_i, satisfying

$$\begin{bmatrix} \tilde{e}^i_1 \\ \vdots \\ \tilde{e}^i_{n_i} \end{bmatrix} = A_i \begin{bmatrix} e^i_1 \\ \vdots \\ e^i_{n_i} \end{bmatrix}, \quad i = 0, 1, \cdots, k.$$

Find the structure matrix of f under the new basis.

2. Prove the following alternative expression of swap matrix.

$$W_{[m,n]} = \left[\delta_n^1 \ltimes \delta_m^1 \cdots \delta_n^n \ltimes \delta_m^1 \cdots \delta_n^1 \ltimes \delta_m^m \cdots \delta_n^n \ltimes \delta_m^m\right]. \tag{2.67}$$

3. Complete the proof of Theorem 2.1.
4. Prove equations (2.19)-(2.24).
5. Let $X \in V$. Define a mapping, which reduce the covariant order of a tensor by one, denoted by $i_X: \mathscr{T}_s^r \to T_s^{r-1}$, and defined as [2]

$$i_X(\omega) = \omega(X, \cdots; \cdots), \quad \omega \in \mathscr{T}_s^r. \tag{2.68}$$

Prove that the structure matrix of $i_X(\omega)$ is

$$M_{i_X(\omega)} = M_\omega \ltimes X. \tag{2.69}$$

6. Let $\sigma \in V^*$. A mapping $i_\sigma: T_s^r \to T_{s-1}^r$ is defined as

$$i_\sigma(\omega) = \omega(\cdots; \sigma, \cdots). \tag{2.70}$$

Prove that the structure matrix of $i_\sigma(\omega)$ is

$$M_{i_\sigma(\omega)} = \sigma \ltimes M_\omega. \tag{2.71}$$

7. Let η be a multi-linear mapping:

$$\eta \in L(V_1 \times V_2 \times \cdots \times V_k; W),$$

where $\dim(V_i) = n_i$, $\dim(W) = n_0$. Given $X \in V_t$, we define a mapping

$$i_X^t(\eta): L(V_1 \times \cdots \times V_k; W) \to L(V_1 \times \cdots V_{t-1} \times V_{t+1} \times \cdots \times V_k; W)$$

as

$$\begin{aligned} i_X^t(\eta)(Y_1, \cdots, Y_{k-1}) = \eta(Y_1, \cdots, Y_{t-1}, X, Y_t, \cdots, Y_{k-1}), \\ \forall Y_i \in V_i,\ i < t; Y_i \in V_{i+1},\ i \geq t. \end{aligned} \tag{2.72}$$

Denote by M_η and M_ζ the structure matrices of η and $\zeta = i_X^t(\eta)$ respectively. Then

$$M_\zeta = M_\eta \ltimes (I_{n_1 + \cdots + n_{t-1}} \otimes X). \tag{2.73}$$

8. Prove the Proposition 2.5.
9. Prove Equations (2.49) and (2.49).
10. Prove Equations (2.56), (2.57), and (2.58).
11. Let $\sigma \in \mathbf{S}_k$ be a kth-order permutation. $n = n_1 \times \cdots \times n_k$, $n_i \geq 1$. Define an $n \times n$ matrix W_σ as follows: Label its columns by k indices $i_1, \cdots, i_k$ in the order of $\mathrm{id}(i_1, \cdots, i_k; n_1, \cdots, n_k)$, and label its rows by k indices $j_1, \cdots, j_k$ in the order of $\mathrm{id}(j_{\sigma_1}, \cdots, j_{\sigma_k}; n_{\sigma_1}, \cdots, n_{\sigma_k})$. Then set its entry at $i_1 \cdots i_k$ column and $j_1 \cdots j_k$ row as
$$\omega^{i_1,\cdots,i_k}_{j_1,\cdots,j_k} = \begin{cases} 1, & i_1 = j_1, \cdots, i_k = j_k \\ 0, & \text{otherwise.} \end{cases}$$
Call the matrix W_σ the permutation matrix of σ.

   a. Given column vectors $X_i \in \mathbb{R}^{n_i}$, $i = 1, \cdots, k$, prove that
   $$W_\sigma X_1 \cdots X_k = X_{\sigma_1} \cdots X_{\sigma_k}. \tag{2.74}$$
   b. Given row vectors $\omega_i \in \mathbb{R}^{n_i}$, $i = 1, \cdots, k$, prove that
   $$\omega_1 \cdots \omega_k W_\sigma = \omega_{\sigma_1} \cdots \omega_{\sigma_k}. \tag{2.75}$$

12. Let $\sigma = \sigma_k \sigma_{k-1} \cdots \sigma_1 \in \mathbf{S}_n$. Then
$$W_\sigma = W_{\sigma_k} W_{\sigma_{k-1}} \cdots W_1. \tag{2.76}$$
Prove it.
13. Assume $n_1 = n_2 = n_3 = 2$, $\sigma = (1,2,3) \in \mathbf{S}_3$.

   a. Construct the permutation matrix W_σ.
   b. Let $\sigma_1 = (12)$, $\sigma_2 = (13)$. Then $\sigma = \sigma_2 \sigma_1$. Construct the swap matrices W_{σ_1} and W_{σ_2}. Then check that
   $$W_\sigma = W_{\sigma_2} W_{\sigma_1}.$$

14. Consider sets of vectors as $X_i \in \mathbb{R}^u$, $Y_{ij} \in \mathbb{R}^v$, $Z_{ijk} \in \mathbb{R}^s$, $W_{ijk} \in \mathbb{R}^t$. Assume $\{X_i \,|\, 1 \leq i \leq \alpha\}$ are linearly independent, $\{Y_{ij} \,|\, 1 \leq i \leq \alpha; 1 \leq j \leq \beta\}$ are linearly independent, $\{Z_{ijk} \,|\, 1 \leq i \leq \alpha; 1 \leq j \leq \beta; 1 \leq k \leq \gamma\}$ are also linearly independent. Moreover,
$$\sum_{i=1}^{\alpha} \sum_{i=1}^{\alpha} \sum_{j=1}^{\beta} \sum_{k=1}^{\gamma} X_i Y_{ij} Z_{ijk} W_{ijk} = 0.$$
Then
$$W_{ijk} = 0, \quad 1 \leq i \leq \alpha; 1 \leq j \leq \beta; 1 \leq k \leq \gamma.$$
Prove it.
15. Let X be a square matrix and $p(x)$ a polynomial. Moreover, the polynomial $p(x)$ is expressed as $p(x) = q(x)x + p_0$.

   a. Prove that

$$V_r(p(X)) = q(X)V_r(X) + p_0 V_r(I). \tag{2.77}$$

b. Give the column stacking form expression of (2.77).

16. Let $A \prec_t B$. Prove that

$$\text{rank}(AB) \leq \min(t \times \text{rank}(A), \text{rank}(B)).$$

Similarly, let $A \succ_t B$. Prove that

$$\text{rank}(AB) \leq \min(\text{rank}(A), t \times \text{rank}(B)).$$

References

1. Bates, D., Watts, D.: Relative curvature measures of nonlinearity. Journal of the Royal Statistical Society. Series B (Methodological) **42**, 1–25 (1980)
2. Boothby, W.: An Introduction to Differentiable Manifolds and Riemannian Geometry, 2nd edn. Academic Press, Inc., Orlando (1986)
3. Greub, W.: Multilinear Algebra, 2nd edn. Springer-Verlag, New York (1978)
4. Tsai, C.: Contributions to the Design and Analysis of Nonlinear Models. Ph.D. thesis, Univ. of Minisota (1983)
5. Wang, X.: Parameter Estimate of Nonlinear Models — Theory and Applications. Wuhan Univ. Press, Wuhan (2002). (in Chinese)
6. Zhang, Y.: Theory of Multi-Edge Matrix. Chinese Statistics Press, Henan (1993). (in Chinese)