量子計算的數學基礎 MA5501*

Chapter 3. Mathematical Backgrounds

§3.1 Vector Spaces and Linear Maps

§3.2 Direct Sum of Vector Spaces and Multi-linear Maps

§3.3 Inner Product Spaces and Hilbert Spaces

§3.4 Dual Spaces and Adjoint Operators

§3.5 Unitary Operators and Unitary Matrices

§3.6 Tensor Product of Vector Spaces

§3.7 Unitary Decomposition

§3.8 Implementation of Multi-Controlled Rotation Gates

§3.1 Vector Spaces and Linear Maps

§3.1.1 Vector spaces

Definition

A ***vector space*** $\mathbb{V}$ over a scalar field $\mathbb{F}$ is a collection of elements called vectors, with given operations of vector addition $+ : \mathbb{V} \times \mathbb{V} \to \mathbb{V}$ and scalar multiplication $\cdot : \mathbb{F} \times \mathbb{V} \to \mathbb{V}$ such that

1. $\boldsymbol{v} + \boldsymbol{w} = \boldsymbol{w} + \boldsymbol{v}$ for all $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{V}$.
2. $(\boldsymbol{v} + \boldsymbol{w}) + \boldsymbol{u} = \boldsymbol{v} + (\boldsymbol{u} + \boldsymbol{w})$ for all $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in \mathbb{V}$.
3. there exists $\mathbf{0}$, the zero vector, such that $\boldsymbol{v} + \mathbf{0} = \boldsymbol{v}$ for all $\boldsymbol{v} \in \mathbb{V}$.
4. for each $\boldsymbol{v} \in \mathbb{V}$ there exists $\boldsymbol{w} \in \mathbb{V}$ such that $\boldsymbol{v} + \boldsymbol{w} = \mathbf{0}$.
5. $\lambda \cdot (\boldsymbol{v} + \boldsymbol{w}) = \lambda \cdot \boldsymbol{v} + \lambda \cdot \boldsymbol{w}$ for all $\lambda \in \mathbb{F}$ and $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{V}$.
6. $(\lambda + \mu) \cdot \boldsymbol{v} = \lambda \cdot \boldsymbol{v} + \mu \cdot \boldsymbol{v}$ for all $\lambda, \mu \in \mathbb{F}$ and $\boldsymbol{v} \in \mathbb{V}$.
7. $(\lambda \cdot \mu) \cdot \boldsymbol{v} = \lambda \cdot (\mu \cdot \boldsymbol{v})$ for all $\lambda, \mu \in \mathbb{F}$ and $\boldsymbol{v} \in \mathbb{V}$.
8. $1 \cdot \boldsymbol{v} = \boldsymbol{v}$ for all $\boldsymbol{v} \in \mathbb{V}$.

§3.1 Vector Spaces and Linear Maps

Remark: In the following we always assume that the scalar field $\mathbb{F}$ under consideration is $\mathbb{R}$ or $\mathbb{C}$.

Remark: In property 4 of the definition above, it is easy to see that for each $\boldsymbol{v}$, there is only one vector $\boldsymbol{w}$ such that $\boldsymbol{v}+\boldsymbol{w}=\mathbf{0}$. We often denote this $\boldsymbol{w}$ by $-\boldsymbol{v}$, and the vector substraction $-:\mathbb{V}\times\mathbb{V}\to\mathbb{V}$ is then defined (or understood) as $\boldsymbol{v}-\boldsymbol{w}=\boldsymbol{v}+(-\boldsymbol{w})$.

Example

Let $\mathbb{F}$ be a scalar field. The space $\mathbb{F}^n$ is the collection of n-tuple $\boldsymbol{v}=(\mathrm{v}_1,\mathrm{v}_2,\cdots,\mathrm{v}_n)$ with $\mathrm{v}_i\in\mathbb{F}$ with addition $+$ and scalar multiplication $\cdot$ defined by

$$(\mathrm{v}_1,\cdots,\mathrm{v}_n)+(\mathrm{w}_1,\cdots,\mathrm{w}_n)\equiv(\mathrm{v}_1+\mathrm{w}_1,\cdots,\mathrm{v}_n+\mathrm{w}_n),$$
$$\alpha(\mathrm{v}_1,\cdots,\mathrm{v}_n)\equiv(\alpha\mathrm{v}_1,\cdots,\alpha\mathrm{v}_n).$$

Then $\mathbb{F}^n$ is a vector space over $\mathbb{F}$.

§3.1 Vector Spaces and Linear Maps

Remark: In the following we always assume that the scalar field $\mathbb{F}$ under consideration is $\mathbb{R}$ or $\mathbb{C}$.

Remark: In property 4 of the definition above, it is easy to see that for each $\boldsymbol{v}$, there is only one vector $\boldsymbol{w}$ such that $\boldsymbol{v}+\boldsymbol{w}=\mathbf{0}$. We often denote this $\boldsymbol{w}$ by $-\boldsymbol{v}$, and the vector substraction $-:\mathbb{V}\times\mathbb{V}\to\mathbb{V}$ is then defined (or understood) as $\boldsymbol{v}-\boldsymbol{w}=\boldsymbol{v}+(-\boldsymbol{w})$.

Example

Let $\mathbb{F}$ be a scalar field. The space $\mathbb{F}^n$ is the collection of n-tuple $\boldsymbol{v}=(\mathrm{v}_1,\mathrm{v}_2,\cdots,\mathrm{v}_n)$ with $\mathrm{v}_i\in\mathbb{F}$ with addition $+$ and scalar multiplication $\cdot$ defined by

$$(\mathrm{v}_1,\cdots,\mathrm{v}_n)+(\mathrm{w}_1,\cdots,\mathrm{w}_n)\equiv(\mathrm{v}_1+\mathrm{w}_1,\cdots,\mathrm{v}_n+\mathrm{w}_n)\,,$$
$$\alpha(\mathrm{v}_1,\cdots,\mathrm{v}_n)\equiv(\alpha\mathrm{v}_1,\cdots,\alpha\mathrm{v}_n)\,.$$

Then $\mathbb{F}^n$ is a vector space over $\mathbb{F}$.

§3.1 Vector Spaces and Linear Maps

Remark: In the following we always assume that the scalar field $\mathbb{F}$ under consideration is $\mathbb{R}$ or $\mathbb{C}$.

Remark: In property 4 of the definition above, it is easy to see that for each $\boldsymbol{v}$, there is only one vector $\boldsymbol{w}$ such that $\boldsymbol{v}+\boldsymbol{w}=\mathbf{0}$. We often denote this $\boldsymbol{w}$ by $-\boldsymbol{v}$, and the vector substraction $-:\mathbb{V}\times\mathbb{V}\to\mathbb{V}$ is then defined (or understood) as $\boldsymbol{v}-\boldsymbol{w}=\boldsymbol{v}+(-\boldsymbol{w})$.

Example

Let $\mathbb{F}$ be a scalar field. The space $\mathbb{F}^n$ is the collection of n-tuple $\boldsymbol{v}=(\mathrm{v}_1,\mathrm{v}_2,\cdots,\mathrm{v}_n)$ with $\mathrm{v}_i\in\mathbb{F}$ with addition $+$ and scalar multiplication $\cdot$ defined by

$$(\mathrm{v}_1,\cdots,\mathrm{v}_n)+(\mathrm{w}_1,\cdots,\mathrm{w}_n)\equiv(\mathrm{v}_1+\mathrm{w}_1,\cdots,\mathrm{v}_n+\mathrm{w}_n),$$
$$\alpha(\mathrm{v}_1,\cdots,\mathrm{v}_n)\equiv(\alpha\mathrm{v}_1,\cdots,\alpha\mathrm{v}_n).$$

Then $\mathbb{F}^n$ is a vector space over $\mathbb{F}$.

§3.1 Vector Spaces and Linear Maps

Example

Let $\mathbb{F}$ be a scalar field. The collection of $m \times n$ matrices with entries in $\mathbb{F}$ is denoted by $\mathcal{M}(m, n; \mathbb{F})$ or $\mathbb{F}^{m\times n}$; that is, $A \in \mathcal{M}(m, n; \mathbb{F})$ if and only if $A = [a_{ij}]_{1\leqslant i\leqslant m, 1\leqslant j\leqslant n}$ for some $a_{ij} \in \mathbb{F}$. Define the addition $+$ and scalar multiplication $\cdot$ on $\mathcal{M}(m, n; \mathbb{F})$ by

$$A + B = [a_{ij} + b_{ij}] \qquad \text{if} \quad A = [a_{ij}] \text{ and } B = [b_{ij}]$$

and

$$c \cdot A = [c \cdot a_{ij}] \qquad \text{if} \quad A = [a_{ij}] .$$

Then $(\mathcal{M}(m, n; \mathbb{F}), +, \cdot)$ is a vector space over $\mathbb{F}$.

Definition (Vector subspace)

Let $\mathbb{V}$ be a vector space over scalar field $\mathbb{F}$. A subset $\mathbb{W} \subseteq \mathbb{V}$ is called a vector subspace of $\mathbb{V}$ if itself is a vector space over $\mathbb{F}$.

§3.1 Vector Spaces and Linear Maps

Example

Let $\mathbb{F}$ be a scalar field. The collection of $m \times n$ matrices with entries in $\mathbb{F}$ is denoted by $\mathcal{M}(m, n; \mathbb{F})$ or $\mathbb{F}^{m\times n}$; that is, $A \in \mathcal{M}(m, n; \mathbb{F})$ if and only if $A = [a_{ij}]_{1\leqslant i\leqslant m, 1\leqslant j\leqslant n}$ for some $a_{ij} \in \mathbb{F}$. Define the addition $+$ and scalar multiplication $\cdot$ on $\mathcal{M}(m, n; \mathbb{F})$ by

$$A + B = [a_{ij} + b_{ij}] \qquad \text{if} \quad A = [a_{ij}] \text{ and } B = [b_{ij}]$$

and

$$c \cdot A = [c \cdot a_{ij}] \qquad \text{if} \quad A = [a_{ij}] \,.$$

Then $(\mathcal{M}(m, n; \mathbb{F}), +, \cdot)$ is a vector space over $\mathbb{F}$.

Definition (Vector subspace)

Let $\mathbb{V}$ be a vector space over scalar field $\mathbb{F}$. A subset $\mathbb{W} \subseteq \mathbb{V}$ is called a vector subspace of $\mathbb{V}$ if itself is a vector space over $\mathbb{F}$.

§3.1 Vector Spaces and Linear Maps

Definition

Let $\mathbb{V}$ be a vector space over a scalar field $\mathbb{F}$. k vectors $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, $\cdots$, $\boldsymbol{v}_k$ in $\mathbb{V}$ is said to be ***linearly dependent*** if there exist $\alpha_1, \cdots, \alpha_k \in \mathbb{F}$, $(\alpha_1, \cdots, \alpha_k) \neq \mathbf{0}$ such that $\alpha_1 \boldsymbol{v}_1 + \alpha_2 \boldsymbol{v}_2 + \cdots + \alpha_k \boldsymbol{v}_k = \mathbf{0}$. k vectors $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, $\cdots$, $\boldsymbol{v}_k$ in $\mathbb{V}$ is said to be ***linearly independent*** if they are not linearly dependent. In other words, $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k\}$ are linearly independent if

$$\alpha_1 \boldsymbol{v}_1 + \alpha_2 \boldsymbol{v}_2 + \cdots + \alpha_k \boldsymbol{v}_k = \mathbf{0} \quad \Rightarrow \quad \alpha_1 = \alpha_2 = \cdots = \alpha_k = 0 \,.$$

Example

The k vectors $\{1, x, x^2, \cdots, x^{k-1}\}$ are linearly independent in the space of polynomials for all $k \in \mathbb{N}$.

§3.1 Vector Spaces and Linear Maps

Definition

Let $\mathbb{V}$ be a vector space over a scalar field $\mathbb{F}$. k vectors $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, $\cdots$, $\boldsymbol{v}_k$ in $\mathbb{V}$ is said to be ***linearly dependent*** if there exist $\alpha_1, \cdots, \alpha_k \in \mathbb{F}$, $(\alpha_1, \cdots, \alpha_k) \neq \mathbf{0}$ such that $\alpha_1 \boldsymbol{v}_1 + \alpha_2 \boldsymbol{v}_2 + \cdots + \alpha_k \boldsymbol{v}_k = \mathbf{0}$. k vectors $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, $\cdots$, $\boldsymbol{v}_k$ in $\mathbb{V}$ is said to be ***linearly independent*** if they are not linearly dependent. In other words, $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k\}$ are linearly independent if

$$\alpha_1 \boldsymbol{v}_1 + \alpha_2 \boldsymbol{v}_2 + \cdots + \alpha_k \boldsymbol{v}_k = \mathbf{0} \quad \Rightarrow \quad \alpha_1 = \alpha_2 = \cdots = \alpha_k = 0\,.$$

Example

The k vectors $\{1, x, x^2, \cdots, x^{k-1}\}$ are linearly independent in the space of polynomials for all $k \in \mathbb{N}$.

§3.1 Vector Spaces and Linear Maps

Definition

The ***dimension*** of a vector space $\mathbb{V}$ is the number of maximum linearly independent set in $\mathbb{V}$, and in such case $\mathbb{V}$ is called an n-dimensional vector space, where n is the dimension of $\mathbb{V}$. If for every number $n \in \mathbb{N}$ there exists n linearly independent vectors in $\mathbb{V}$, the vector space $\mathbb{V}$ is said to be infinitely dimensional.

Definition (Basis)

Let $\mathbb{V}$ be a vector space over $\mathbb{F}$. A collection of vectors $\{\boldsymbol{v}_i\}_{i\in\mathcal{I}}$ in $\mathbb{V}$ is called a ***basis*** of $\mathbb{V}$ if for every $\boldsymbol{v} \in \mathbb{V}$, there exists a unique $\{\alpha_i\}_{i\in\mathcal{I}} \subseteq \mathbb{F}$ such that

$$\boldsymbol{v} = \sum_{\alpha\in\mathcal{I}} \alpha_i \boldsymbol{v}_i \,.$$

For a given basis $\mathcal{B} = \{\boldsymbol{v}_i\}_{i\in\mathcal{I}}$, the coefficients $\{\alpha_i\}_{i\in\mathcal{I}}$ given in the above relation is denoted by $[\boldsymbol{v}]_{\mathcal{B}}$.

§3.1 Vector Spaces and Linear Maps

Definition

The ***dimension*** of a vector space $\mathbb{V}$ is the number of maximum linearly independent set in $\mathbb{V}$, and in such case $\mathbb{V}$ is called an n-dimensional vector space, where n is the dimension of $\mathbb{V}$. If for every number $n \in \mathbb{N}$ there exists n linearly independent vectors in $\mathbb{V}$, the vector space $\mathbb{V}$ is said to be infinitely dimensional.

Definition (Basis)

Let $\mathbb{V}$ be a vector space over $\mathbb{F}$. A collection of vectors $\{\boldsymbol{v}_i\}_{i\in\mathcal{I}}$ in $\mathbb{V}$ is called a ***basis*** of $\mathbb{V}$ if for every $\boldsymbol{v} \in \mathbb{V}$, there exists a unique $\{\alpha_i\}_{i\in\mathcal{I}} \subseteq \mathbb{F}$ such that

$$\boldsymbol{v} = \sum_{\alpha\in\mathcal{I}} \alpha_i \boldsymbol{v}_i .$$

For a given basis $\mathcal{B} = \{\boldsymbol{v}_i\}_{i\in\mathcal{I}}$, the coefficients $\{\alpha_i\}_{i\in\mathcal{I}}$ given in the above relation is denoted by $[\boldsymbol{v}]_{\mathcal{B}}$.

§3.1 Vector Spaces and Linear Maps

§3.1.2 Linear Maps and their matrix representations

Definition

Let $\mathbb{V}, \mathbb{W}$ be vector spaces over a common scalar field $\mathbb{F}$. A map L from $\mathbb{V}$ to $\mathbb{W}$ is said to be linear if $L(c\boldsymbol{v}_1 + \boldsymbol{v}_2) = cL(\boldsymbol{v}_1) + L(\boldsymbol{v}_2)$ for all $\boldsymbol{v}_1, \boldsymbol{v}_2 \in \mathbb{V}$ and $c \in \mathbb{F}$. We often write $L\boldsymbol{v}$ instead of $L(\boldsymbol{v})$, and the collection of all linear maps from $\mathbb{V}$ to $\mathbb{W}$ is denoted by $\mathscr{L}(\mathbb{V};\mathbb{W})$. We also write $\mathscr{L}(\mathbb{V})$ instead of $\mathscr{L}(\mathbb{V};\mathbb{V})$ if $\mathbb{W} = \mathbb{V}$. An element in $\mathscr{L}(\mathbb{V};\mathbb{F})$ is called a linear functional on $\mathbb{V}$.

Proposition

Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces over a common scalar field $\mathbb{F}$. Then $\mathscr{L}(\mathbb{V};\mathbb{W})$ is a vector space over $\mathbb{F}$.

§3.1 Vector Spaces and Linear Maps

§3.1.2 Linear Maps and their matrix representations

Definition

Let $\mathbb{V}, \mathbb{W}$ be vector spaces over a common scalar field $\mathbb{F}$. A map L from $\mathbb{V}$ to $\mathbb{W}$ is said to be linear if $L(c\boldsymbol{v}_1 + \boldsymbol{v}_2) = cL(\boldsymbol{v}_1) + L(\boldsymbol{v}_2)$ for all $\boldsymbol{v}_1, \boldsymbol{v}_2 \in \mathbb{V}$ and $c \in \mathbb{F}$. We often write $L\boldsymbol{v}$ instead of $L(\boldsymbol{v})$, and the collection of all linear maps from $\mathbb{V}$ to $\mathbb{W}$ is denoted by $\mathscr{L}(\mathbb{V};\mathbb{W})$. We also write $\mathscr{L}(\mathbb{V})$ instead of $\mathscr{L}(\mathbb{V};\mathbb{V})$ if $\mathbb{W} = \mathbb{V}$. An element in $\mathscr{L}(\mathbb{V};\mathbb{F})$ is called a linear functional on $\mathbb{V}$.

Proposition

Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces over a common scalar field $\mathbb{F}$. Then $\mathscr{L}(\mathbb{V};\mathbb{W})$ is a vector space over $\mathbb{F}$.

§3.1 Vector Spaces and Linear Maps

Example

Let $\mathbb{F}$ be a scalar field, and $A = [a_{ij}]_{1\leqslant i\leqslant m, 1\leqslant j\leqslant n} \in \mathcal{M}(m,n;\mathbb{F})$ be an $m \times n$ matrix. Define a vector-valued function $L : \mathbb{F}^n \to \mathbb{F}^m$ by

$$L(x_1, \cdots, x_n) = \Big(\sum_{j=1}^{n} a_{1j}x_j, \sum_{j=1}^{n} a_{2j}x_j, \cdots, \sum_{j=1}^{n} a_{mj}x_j \Big).$$

Then $L \in \mathcal{L}(\mathbb{F}^n, \mathbb{F}^m)$.

From the example above, we see that any $m \times n$ matrix is associated with a linear map. Now suppose that $\mathbb{V}$ and $\mathbb{W}$ are vector spaces over a common scalar field $\mathbb{F}$, $\mathbb{V}$ is a n-dimensional vector space with basis $\mathcal{B} = \{\boldsymbol{v}_j\}_{j=1}^n$, and $\mathbb{W}$ is a m-dimensional vector space with basis $\widetilde{\mathcal{B}} = \{\boldsymbol{w}_i\}_{i=1}^m$.

§3.1 Vector Spaces and Linear Maps

Example

Let $\mathbb{F}$ be a scalar field, and $A = [a_{ij}]_{1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n} \in \mathcal{M}(m, n; \mathbb{F})$ be an $m \times n$ matrix. Define a vector-valued function $L : \mathbb{F}^n \to \mathbb{F}^m$ by

$$L(x_1, \cdots, x_n) = \Big(\sum_{j=1}^{n} a_{1j} x_j, \sum_{j=1}^{n} a_{2j} x_j, \cdots, \sum_{j=1}^{n} a_{mj} x_j \Big) .$$

Then $L \in \mathcal{L}(\mathbb{F}^n, \mathbb{F}^m)$.

From the example above, we see that any $m \times n$ matrix is associated with a linear map. Now suppose that $\mathbb{V}$ and $\mathbb{W}$ are vector spaces over a common scalar field $\mathbb{F}$, $\mathbb{V}$ is a n-dimensional vector space with basis $\mathcal{B} = \{\boldsymbol{v}_j\}_{j=1}^{n}$, and $\mathbb{W}$ is a m-dimensional vector space with basis $\widetilde{\mathcal{B}} = \{\boldsymbol{w}_i\}_{i=1}^{m}$.

§3.1 Vector Spaces and Linear Maps

Example

Let $\mathbb{F}$ be a scalar field, and $A = [a_{ij}]_{1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n} \in \mathcal{M}(m, n; \mathbb{F})$ be an $m \times n$ matrix. Define a vector-valued function $L : \mathbb{F}^n \to \mathbb{F}^m$ by

$$L(x_1, \cdots, x_n) = \Big(\sum_{j=1}^{n} a_{1j} x_j, \sum_{j=1}^{n} a_{2j} x_j, \cdots, \sum_{j=1}^{n} a_{mj} x_j \Big).$$

Then $L \in \mathscr{L}(\mathbb{F}^n, \mathbb{F}^m)$.

From the example above, we see that any $m \times n$ matrix is associated with a linear map. Now suppose that $\mathbb{V}$ and $\mathbb{W}$ are vector spaces over a common scalar field $\mathbb{F}$, $\mathbb{V}$ is a n-dimensional vector space with basis $\mathcal{B} = \{\boldsymbol{v}_j\}_{j=1}^{n}$, and $\mathbb{W}$ is a m-dimensional vector space with basis $\widetilde{\mathcal{B}} = \{\boldsymbol{w}_i\}_{i=1}^{m}$.

§3.1 Vector Spaces and Linear Maps

Let $L \in \mathcal{L}(\mathbb{V};\mathbb{W})$. Since $\widetilde{B} = \{\boldsymbol{w}_i\}_{i=1}^{m}$ is a basis of $\mathbb{W}$, for each $\boldsymbol{v}_j \in \mathcal{B}$ there exist unique a_{1j}, a_{2j}, $\cdots$, $a_{mj} \in \mathbb{F}$ such that $L\boldsymbol{v}_j = \sum\limits_{i=1}^{m} a_{ij}\boldsymbol{w}_i$.

Moreover, if $\boldsymbol{u} \in \mathbb{V}$, then there exist $c_1, \cdots, c_n \in \mathbb{F}$ such that

$$\boldsymbol{u} = \sum_{j=1}^{n} c_j \boldsymbol{v}_j \qquad \text{or} \qquad \boldsymbol{c} = [\boldsymbol{u}]_{\mathcal{B}}\,,$$

and by the linearity of L,

$$L\boldsymbol{u} = L\Big(\sum_{j=1}^{n} c_j \boldsymbol{v}_j\Big) = \sum_{j=1}^{n} c_j L\boldsymbol{v}_j = \sum_{j=1}^{n} c_j \Big(\sum_{i=1}^{m} a_{ij}\boldsymbol{w}_i\Big) = \sum_{i=1}^{m} \Big(\sum_{j=1}^{n} a_{ij}c_j\Big)\boldsymbol{w}_i\,.$$

Let $b_i = \sum\limits_{j=1}^{n} a_{ij}c_j$, and $\boldsymbol{b} = [b_1, \cdots, b_m]^{\mathrm{T}}$. Then with A denoting the $m \times n$ matrix $[a_{ij}]_{1\leqslant i\leqslant m, 1\leqslant j\leqslant n}$,

$$[L\boldsymbol{u}]_{\widetilde{\mathcal{B}}} = \boldsymbol{b} = A\boldsymbol{c} = A[\boldsymbol{u}]_{\mathcal{B}}\,.$$

§3.1 Vector Spaces and Linear Maps

Let $L \in \mathscr{L}(\mathbb{V};\mathbb{W})$. Since $\widetilde{\mathcal{B}} = \{\boldsymbol{w}_i\}_{i=1}^m$ is a basis of $\mathbb{W}$, for each $\boldsymbol{v}_j \in \mathcal{B}$ there exist unique $a_{1j}, a_{2j}, \cdots, a_{mj} \in \mathbb{F}$ such that $L\boldsymbol{v}_j = \sum\limits_{i=1}^m a_{ij}\boldsymbol{w}_i$. Moreover, if $\boldsymbol{u} \in \mathbb{V}$, then there exist $c_1, \cdots, c_n \in \mathbb{F}$ such that

$$\boldsymbol{u} = \sum_{j=1}^n c_j \boldsymbol{v}_j \qquad \text{or} \qquad \boldsymbol{c} = [\boldsymbol{u}]_{\mathcal{B}}\,,$$

and by the linearity of L,

$$L\boldsymbol{u} = L\Big(\sum_{j=1}^n c_j\boldsymbol{v}_j\Big) = \sum_{j=1}^n c_j L\boldsymbol{v}_j = \sum_{j=1}^n c_j\Big(\sum_{i=1}^m a_{ij}\boldsymbol{w}_i\Big) = \sum_{i=1}^m\Big(\sum_{j=1}^n a_{ij}c_j\Big)\boldsymbol{w}_i\,.$$

Let $b_i = \sum\limits_{j=1}^n a_{ij}c_j$, and $\boldsymbol{b} = [b_1, \cdots, b_m]^{\mathrm{T}}$. Then with A denoting the $m \times n$ matrix $[a_{ij}]_{1\leqslant i\leqslant m, 1\leqslant j\leqslant n}$,

$$[L\boldsymbol{u}]_{\widetilde{\mathcal{B}}} = \boldsymbol{b} = A\boldsymbol{c} = A[\boldsymbol{u}]_{\mathcal{B}}\,.$$

§3.1 Vector Spaces and Linear Maps

Let $L \in \mathscr{L}(\mathbb{V};\mathbb{W})$. Since $\widetilde{\mathcal{B}} = \{\boldsymbol{w}_i\}_{i=1}^m$ is a basis of $\mathbb{W}$, for each $\boldsymbol{v}_j \in \mathcal{B}$ there exist unique $a_{1j}, a_{2j}, \cdots, a_{mj} \in \mathbb{F}$ such that $L\boldsymbol{v}_j = \sum\limits_{i=1}^m a_{ij}\boldsymbol{w}_i$. Moreover, if $\boldsymbol{u} \in \mathbb{V}$, then there exist $c_1, \cdots, c_n \in \mathbb{F}$ such that

$$\boldsymbol{u} = \sum_{j=1}^n c_j\boldsymbol{v}_j \qquad \text{or} \qquad \boldsymbol{c} = [\boldsymbol{u}]_{\mathcal{B}}\,,$$

and by the linearity of L,

$$L\boldsymbol{u} = L\Big(\sum_{j=1}^n c_j\boldsymbol{v}_j\Big) = \sum_{j=1}^n c_j L\boldsymbol{v}_j = \sum_{j=1}^n c_j\Big(\sum_{i=1}^m a_{ij}\boldsymbol{w}_i\Big) = \sum_{i=1}^m\Big(\sum_{j=1}^n a_{ij}c_j\Big)\boldsymbol{w}_i\,.$$

Let $b_i = \sum\limits_{j=1}^n a_{ij}c_j$, and $\boldsymbol{b} = [b_1, \cdots, b_m]^{\mathrm{T}}$. Then with A denoting the $m \times n$ matrix $[a_{ij}]_{1\leqslant i\leqslant m, 1\leqslant j\leqslant n}$,

$$[L\boldsymbol{u}]_{\widetilde{\mathcal{B}}} = \boldsymbol{b} = A\boldsymbol{c} = A[\boldsymbol{u}]_{\mathcal{B}}\,.$$

§3.1 Vector Spaces and Linear Maps

The discussion above induces the following

Definition

Let $\mathbb{V}, \mathbb{W}$ be two vector spaces over a common scalar field $\mathbb{F}$, $\dim(\mathbb{V}) = n$ and $\dim(\mathbb{W}) = m$, and $\mathcal{B}, \widetilde{\mathcal{B}}$ be basis of $\mathbb{V}, \mathbb{W}$, respectively. For $L \in \mathscr{L}(\mathbb{V}; \mathbb{W})$, the ***matrix representation*** of L relative to bases $\mathcal{B}$ and $\widetilde{\mathcal{B}}$, denoted by $[L]_{\widetilde{\mathcal{B}},\mathcal{B}}$, is the matrix in $\mathcal{M}(m, n; \mathbb{F})$ satisfying

$$[L\boldsymbol{u}]_{\widetilde{\mathcal{B}}} = [L]_{\widetilde{\mathcal{B}},\mathcal{B}}[\boldsymbol{u}]_{\mathcal{B}} \qquad \forall\, \boldsymbol{u} \in \mathbb{V}\,.$$

If $L \in \mathscr{L}(\mathbb{V}; \mathbb{V})$, we simply use $[L]_{\mathcal{B}}$ to denote $[L]_{\mathcal{B},\mathcal{B}}$.

Remark: If $\mathbb{V}$ has a standard basis $\mathcal{B}$ or $\mathcal{B}$ is well-known, we also use $[L]$ instead of $[L]_{\mathcal{B}}$ to simplify the notation.

§3.1 Vector Spaces and Linear Maps

The discussion above induces the following

Definition

Let $\mathbb{V}, \mathbb{W}$ be two vector spaces over a common scalar field $\mathbb{F}$, $\dim(\mathbb{V}) = n$ and $\dim(\mathbb{W}) = m$, and $\mathcal{B}, \widetilde{\mathcal{B}}$ be basis of $\mathbb{V}, \mathbb{W}$, respectively. For $L \in \mathcal{L}(\mathbb{V}; \mathbb{W})$, the ***matrix representation*** of L relative to bases $\mathcal{B}$ and $\widetilde{\mathcal{B}}$, denoted by $[L]_{\widetilde{\mathcal{B}},\mathcal{B}}$, is the matrix in $\mathcal{M}(m, n; \mathbb{F})$ satisfying

$$[L\boldsymbol{u}]_{\widetilde{B}} = [L]_{\widetilde{\mathcal{B}},\mathcal{B}}[\boldsymbol{u}]_{\mathcal{B}} \qquad \forall\, \boldsymbol{u} \in \mathbb{V}\,.$$

If $L \in \mathcal{L}(\mathbb{V}; \mathbb{V})$, we simply use $[L]_{\mathcal{B}}$ to denote $[L]_{\mathcal{B},\mathcal{B}}$.

Remark: If $\mathbb{V}$ has a standard basis $\mathcal{B}$ or $\mathcal{B}$ is well-known, we also use $[L]$ instead of $[L]_{\mathcal{B}}$ to simplify the notation.

§3.1 Vector Spaces and Linear Maps

Example

Let $\mathbb{V} = \text{span}(1, x, \cdots, x^{n-1})$ and $\mathbb{W} = \text{span}(1, x, \cdots, x^{m-1})$ with $m \geqslant n-1$. Then $\dfrac{d}{dx} : \mathbb{V} \to \mathbb{W}$ defined by

$$\frac{d}{dx}\Big(\sum_{k=1}^{n} a_k x^{k-1}\Big) = \sum_{k=1}^{n} a_k (k-1) x^{k-2}$$

is linear, and the matrix representation of $\dfrac{d}{dx}$ (relative to the standard basis of $\mathbb{V}$ and $\mathbb{W}$) is

$$m\text{-rows}\left\{\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & 0 & 2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & 0 & (n-1) \\ \vdots & & & \ddots & 0 \\ \vdots & & & & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 \end{bmatrix}\right. .$$

§3.1 Vector Spaces and Linear Maps

Theorem

Let $\mathbb{V}_1, \mathbb{V}_2, \mathbb{V}_3$ be finite dimensional vector spaces, and $\mathcal{B}_1$, $\mathcal{B}_2$, $\mathcal{B}_3$ be basis of $\mathbb{V}_1$, $\mathbb{V}_2$, $\mathbb{V}_3$, respectively. Then

$$[TS]_{\mathcal{B}_3,\mathcal{B}_1} = [T]_{\mathcal{B}_3,\mathcal{B}_2}[S]_{\mathcal{B}_2,\mathcal{B}_1} \qquad \forall\, S \in \mathcal{L}(\mathbb{V}_1;\mathbb{V}_2),\, T \in \mathcal{L}(\mathbb{V}_2;\mathbb{V}_3)\,.$$

Proof.

Let $S \in \mathcal{L}(\mathbb{V}_1;\mathbb{V}_2)$ and $T \in \mathcal{L}(\mathbb{V}_2;\mathbb{V}_3)$ be given. For all $\boldsymbol{v} \in \mathbb{V}_1$,

$$[TS\boldsymbol{v}]_{\mathcal{B}_3} = [T]_{\mathcal{B}_3,\mathcal{B}_2}[S\boldsymbol{v}]_{\mathcal{B}_2} = [T]_{\mathcal{B}_3,\mathcal{B}_2}[S]_{\mathcal{B}_2,\mathcal{B}_1}[\boldsymbol{v}]_{\mathcal{B}_1}\,,$$

$$[TS\boldsymbol{v}]_{\mathcal{B}_3} = [TS]_{\mathcal{B}_3,\mathcal{B}_1}[\boldsymbol{v}]_{\mathcal{B}_1}\,.$$

Therefore,

$$[T]_{\mathcal{B}_3,\mathcal{B}_2}[S]_{\mathcal{B}_2,\mathcal{B}_1}[\boldsymbol{v}]_{\mathcal{B}_1} = [TS]_{\mathcal{B}_3,\mathcal{B}_1}[\boldsymbol{v}]_{\mathcal{B}_1} \qquad \forall\, \boldsymbol{v} \in \mathbb{V}_1\,;$$

thus letting $\boldsymbol{u}$ be any basis vector implies that

$$[TS]_{\mathcal{B}_3,\mathcal{B}_1} = [T]_{\mathcal{B}_3,\mathcal{B}_2}[S]_{\mathcal{B}_2,\mathcal{B}_1}\,.$$ □

§3.1 Vector Spaces and Linear Maps

Theorem

Let $\mathbb{V}_1, \mathbb{V}_2, \mathbb{V}_3$ *be finite dimensional vector spaces, and* $\mathcal{B}_1$, $\mathcal{B}_2$, $\mathcal{B}_3$ *be basis of* $\mathbb{V}_1$, $\mathbb{V}_2$, $\mathbb{V}_3$, *respectively. Then*

$$[TS]_{\mathcal{B}_3,\mathcal{B}_1} = [T]_{\mathcal{B}_3,\mathcal{B}_2}[S]_{\mathcal{B}_2,\mathcal{B}_1} \qquad \forall\, S \in \mathcal{L}(\mathbb{V}_1;\mathbb{V}_2), T \in \mathcal{L}(\mathbb{V}_2;\mathbb{V}_3)\,.$$

Proof.

Let $S \in \mathcal{L}(\mathbb{V}_1;\mathbb{V}_2)$ and $T \in \mathcal{L}(\mathbb{V}_2;\mathbb{V}_3)$ be given. For all $\boldsymbol{v} \in \mathbb{V}_1$,

$$[TS\boldsymbol{v}]_{\mathcal{B}_3} = [T]_{\mathcal{B}_3,\mathcal{B}_2}[S\boldsymbol{v}]_{\mathcal{B}_2} = [T]_{\mathcal{B}_3,\mathcal{B}_2}[S]_{\mathcal{B}_2,\mathcal{B}_1}[\boldsymbol{v}]_{\mathcal{B}_1}\,,$$

$$[TS\boldsymbol{v}]_{\mathcal{B}_3} = [TS]_{\mathcal{B}_3,\mathcal{B}_1}[\boldsymbol{v}]_{\mathcal{B}_1}\,.$$

Therefore,

$$[T]_{\mathcal{B}_3,\mathcal{B}_2}[S]_{\mathcal{B}_2,\mathcal{B}_1}[\boldsymbol{v}]_{\mathcal{B}_1} = [TS]_{\mathcal{B}_3,\mathcal{B}_1}[\boldsymbol{v}]_{\mathcal{B}_1} \qquad \forall\, \boldsymbol{v} \in \mathbb{V}_1\,;$$

thus letting $\boldsymbol{u}$ be any basis vector implies that

$$[TS]_{\mathcal{B}_3,\mathcal{B}_1} = [T]_{\mathcal{B}_3,\mathcal{B}_2}[S]_{\mathcal{B}_2,\mathcal{B}_1}\,.$$

□

§3.1 Vector Spaces and Linear Maps

§3.1.3 Algebraic dual space

For a given vector space $\mathbb{V}$ over scalar field $\mathbb{F}$, the *algebraic* dual space of $\mathbb{V}$ is $\mathcal{L}(\mathbb{V};\mathbb{F})$ (also denoted by $\mathbb{V}'$).

Example

Let $\mathbb{V} = \mathbb{R}^n$ and $\mathbb{F} = \mathbb{R}$. From Linear Algebra we know that $\mathbb{V}'$ has one-to-one correspondence with $\mathbb{V}$: every $f \in \mathbb{V}'$ corresponds to a unique matrix $\boldsymbol{a} \equiv [a_1, \cdots, a_n] \in \mathbb{R}^n$ such that

$$f(\boldsymbol{x}) = \boldsymbol{a} \cdot \boldsymbol{x}$$

and vice versa. We write $\mathbb{V}'$ "$=$" $\mathbb{V}$. A bit more generalized version of the result above is that $(\mathbb{C}^n)' = \mathbb{C}^n$ in the sense that every $f \in (\mathbb{C}^n)'$ corresponds to a unique vector $\boldsymbol{a} \equiv (c_1, \cdots, c_n) \in \mathbb{C}^n$ such that

$$f(\boldsymbol{z}) = \boldsymbol{c} \cdot \boldsymbol{z} = \sum_{j=1}^{n} \bar{c}_j z_j \,,$$

where $\bar{c}_j$ is the complex conjugate of c_j.

§3.1 Vector Spaces and Linear Maps

§3.1.3 Algebraic dual space

For a given vector space $\mathbb{V}$ over scalar field $\mathbb{F}$, the *algebraic* dual space of $\mathbb{V}$ is $\mathcal{L}(\mathbb{V};\mathbb{F})$ (also denoted by $\mathbb{V}'$).

Example

Let $\mathbb{V} = \mathbb{R}^n$ and $\mathbb{F} = \mathbb{R}$. From Linear Algebra we know that $\mathbb{V}'$ has one-to-one correspondence with $\mathbb{V}$: every $f \in \mathbb{V}'$ corresponds to a unique matrix $\boldsymbol{a} \equiv [a_1, \cdots, a_n] \in \mathbb{R}^n$ such that

$$f(\boldsymbol{x}) = \boldsymbol{a} \cdot \boldsymbol{x}$$

and vice versa. We write $\mathbb{V}'$ "$=$" $\mathbb{V}$. A bit more generalized version of the result above is that $(\mathbb{C}^n)' = \mathbb{C}^n$ in the sense that every $f \in (\mathbb{C}^n)'$ corresponds to a unique vector $\boldsymbol{a} \equiv (c_1, \cdots, c_n) \in \mathbb{C}^n$ such that

$$f(\boldsymbol{z}) = \boldsymbol{c} \cdot \boldsymbol{z} = \sum_{j=1}^{n} \bar{c}_j z_j \,,$$

where $\bar{c}_j$ is the complex conjugate of c_j.

§3.1 Vector Spaces and Linear Maps

§3.1.3 Algebraic dual space

For a given vector space $\mathbb{V}$ over scalar field $\mathbb{F}$, the *algebraic* dual space of $\mathbb{V}$ is $\mathscr{L}(\mathbb{V};\mathbb{F})$ (also denoted by $\mathbb{V}'$).

Example

Let $\mathbb{V}=\mathbb{R}^n$ and $\mathbb{F}=\mathbb{R}$. From Linear Algebra we know that $\mathbb{V}'$ has one-to-one correspondence with $\mathbb{V}$: every $f\in\mathbb{V}'$ corresponds to a unique matrix $\boldsymbol{a}\equiv[a_1,\cdots,a_n]\in\mathbb{R}^n$ such that

$$f(\boldsymbol{x})=\boldsymbol{a}\cdot\boldsymbol{x}$$

and vice versa. We write $\mathbb{V}'$ "$=$" $\mathbb{V}$. A bit more generalized version of the result above is that $(\mathbb{C}^n)'=\mathbb{C}^n$ in the sense that every $f\in(\mathbb{C}^n)'$ corresponds to a unique vector $\boldsymbol{a}\equiv(c_1,\cdots,c_n)\in\mathbb{C}^n$ such that

$$f(\boldsymbol{z})=\boldsymbol{c}\cdot\boldsymbol{z}=\sum_{j=1}^{n}\bar{c}_j z_j\,,$$

where $\bar{c}_j$ is the complex conjugate of c_j.

§3.1 Vector Spaces and Linear Maps

Proposition

Let $\mathbb{V}$ be a finite dimensional vector space over field $\mathbb{F}$. Then $\dim(\mathbb{V}')=\dim(\mathbb{V})$.

Proof.

Let $\{\mathbf{e}_1,\cdots,\mathbf{e}_n\}$ be a basis of $\mathbb{V}$. Define $\varphi_1,\cdots\varphi_n$ by

$$\varphi_i\Big(\sum_{j=1}^{n} c_j\,\mathbf{e}_j\Big)=\sum_{j=1}^{n} c_j\,\delta_{ij}\quad\forall\,1\leqslant i\leqslant n\text{ and }c_j\in\mathbb{F}\,. \tag{1}$$

Then $\varphi_1,\cdots,\varphi_n\in\mathbb{V}'$. Moreover, the collection $\{\varphi_1,\cdots,\varphi_n\}$ are linearly independent since if $\alpha_1,\cdots,\alpha_n\in\mathbb{F}$ verify that

$$\alpha_1\varphi_1+\alpha_2\varphi_2+\cdots+\alpha_n\varphi_n=0\quad(\text{the zero function})\,,$$

we must have

$$(\alpha_1\varphi_1+\alpha_2\varphi_2+\cdots+\alpha_n\varphi_n)(\mathbf{e}_j)=0\qquad\forall\,1\leqslant j\leqslant n$$

which implies that $\alpha_j=0$ for all $1\leqslant j\leqslant n$; thus $\dim(\mathbb{V}')\geqslant n$. □

§3.1 Vector Spaces and Linear Maps

PROPOSITION

Let $\mathbb{V}$ *be a finite dimensional vector space over field* $\mathbb{F}$*. Then* $\dim(\mathbb{V}') = \dim(\mathbb{V})$.

Proof.

Let $\{\mathbf{e}_1, \cdots, \mathbf{e}_n\}$ be a basis of $\mathbb{V}$. Define $\varphi_1, \cdots \varphi_n$ by

$$\varphi_i\Big(\sum_{j=1}^{n} c_j \mathbf{e}_j\Big) = \sum_{j=1}^{n} c_j \delta_{ij} \quad \forall\, 1 \leqslant i \leqslant n \text{ and } c_j \in \mathbb{F}\,. \tag{1}$$

Then $\varphi_1, \cdots, \varphi_n \in \mathbb{V}'$. Moreover, the collection $\{\varphi_1, \cdots, \varphi_n\}$ are linearly independent since if $\alpha_1, \cdots, \alpha_n \in \mathbb{F}$ verify that

$$\alpha_1\varphi_1 + \alpha_2\varphi_2 + \cdots + \alpha_n\varphi_n = 0 \quad \text{(the zero function)}\,,$$

we must have

$$(\alpha_1\varphi_1 + \alpha_2\varphi_2 + \cdots + \alpha_n\varphi_n)(\mathbf{e}_j) = 0 \qquad \forall\, 1 \leqslant j \leqslant n$$

which implies that $\alpha_j = 0$ for all $1 \leqslant j \leqslant n$; thus $\dim(\mathbb{V}') \geq n$. □

§3.1 Vector Spaces and Linear Maps

Proposition

Let $\mathbb{V}$ *be a finite dimensional vector space over field* $\mathbb{F}$*. Then* $\dim(\mathbb{V}') = \dim(\mathbb{V})$*.*

Proof.

Let $\{\mathbf{e}_1, \cdots, \mathbf{e}_n\}$ be a basis of $\mathbb{V}$. Define $\varphi_1, \cdots \varphi_n$ by

$$\varphi_i\Big(\sum_{j=1}^{n} c_j \mathbf{e}_j\Big) = \sum_{j=1}^{n} c_j \delta_{ij} \quad \forall\, 1 \leqslant i \leqslant n \text{ and } c_j \in \mathbb{F}\,. \tag{1}$$

Then $\varphi_1, \cdots, \varphi_n \in \mathbb{V}'$. Moreover, the collection $\{\varphi_1, \cdots, \varphi_n\}$ are linearly independent since if $\alpha_1, \cdots, \alpha_n \in \mathbb{F}$ verify that

$$\alpha_1\varphi_1 + \alpha_2\varphi_2 + \cdots + \alpha_n\varphi_n = 0 \quad \text{(the zero function)}\,,$$

we must have

$$(\alpha_1\varphi_1 + \alpha_2\varphi_2 + \cdots + \alpha_n\varphi_n)(\mathbf{e}_j) = 0 \qquad \forall\, 1 \leqslant j \leqslant n$$

which implies that $\alpha_j = 0$ for all $1 \leqslant j \leqslant n$; thus $\dim(\mathbb{V}') \geqslant n$. □

§3.1 Vector Spaces and Linear Maps

Proposition

Let $\mathbb{V}$ be a finite dimensional vector space over field $\mathbb{F}$. Then $\dim(\mathbb{V}') = \dim(\mathbb{V})$.

Proof.

Let $\{\mathbf{e}_1, \cdots, \mathbf{e}_n\}$ be a basis of $\mathbb{V}$. Define $\varphi_1, \cdots \varphi_n$ by

$$\varphi_i\Big(\sum_{j=1}^{n} c_j \mathbf{e}_j\Big) = \sum_{j=1}^{n} c_j \delta_{ij} \quad \forall\, 1 \leqslant i \leqslant n \text{ and } c_j \in \mathbb{F}\,. \tag{1}$$

Then $\varphi_1, \cdots, \varphi_n \in \mathbb{V}'$. Moreover, the collection $\{\varphi_1, \cdots, \varphi_n\}$ are linearly independent since if $\alpha_1, \cdots, \alpha_n \in \mathbb{F}$ verify that

$$\alpha_1\varphi_1 + \alpha_2\varphi_2 + \cdots + \alpha_n\varphi_n = 0 \quad \text{(the zero function)}\,,$$

we must have

$$(\alpha_1\varphi_1 + \alpha_2\varphi_2 + \cdots + \alpha_n\varphi_n)(\mathbf{e}_j) = 0 \qquad \forall\, 1 \leqslant j \leqslant n$$

which implies that $\alpha_j = 0$ for all $1 \leqslant j \leqslant n$; thus $\dim(\mathbb{V}') \geqslant n$. □

§3.1 Vector Spaces and Linear Maps

Proof (cont.)

On the other hand, suppose that $g \in \mathbb{V}'$ and $g(\mathbf{e}_j) = d_j$. If $\boldsymbol{x} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n$, the linearity of g implies that

$$\begin{aligned} g(\boldsymbol{x}) &= g(x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n) = x_1 g(\mathbf{e}_1) + \cdots + x_n g(\mathbf{e}_n) \\ &= d_1 x_1 + \cdots + d_n x_n \end{aligned}$$

and (1) shows that

$$\begin{aligned} &(d_1\varphi_1 + \cdots + d_n\varphi_n)(\boldsymbol{x}) \\ &\quad = (d_1\varphi_1 + \cdots + d_n\varphi_n)(x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n) = \sum_{i=1}^{n} d_i\varphi_i\Big(\sum_{j=1}^{n} x_j\mathbf{e}_j\Big) \\ &\quad = \sum_{i=1}^{n}\sum_{j=1}^{n} d_i x_j \delta_{ij} = \sum_{i=1}^{n} d_i x_i = d_1 x_1 + \cdots + d_n x_n\,. \end{aligned}$$

Therefore, $g = d_1\varphi_1 + \cdots + d_n\varphi_n$ which implies that $\mathbb{V}' = \text{span}(\varphi_1, \cdots, \varphi_n)$. This establishes that $\dim(\mathbb{V}') = n$. □

§3.2 Direct Sum and Multi-Linear Maps

§3.2.1 Direct sum of vector spaces

Definition

Given sets A and B, the Cartesian product of A and B, denoted by $A \times B$, is the set of all ordered pairs (a, b) with $a \in A$ and $b \in B$; that is, $A \times B = \big\{(a, b) \,\big|\, a \in A \text{ and } b \in B\big\}$. The Cartesian of three or more sets are defined similarly.

Let X and Y be vector spaces over a common scalar field $\mathbb{F}$. The ***direct sum*** of X and Y, denoted by $X \oplus Y$, is $X \times Y$ with the following vector space structure: $\forall\, \lambda \in \mathbb{F}, \boldsymbol{x}_1, \boldsymbol{x}_2 \in X$ and $\boldsymbol{y}_1, \boldsymbol{y}_2 \in Y$,

$$\lambda \cdot (\boldsymbol{x}_1, \boldsymbol{y}_1) + (\boldsymbol{x}_2, \boldsymbol{y}_2) = (\lambda \cdot \boldsymbol{x}_1 + \boldsymbol{x}_2, \lambda \cdot \boldsymbol{y}_1 + \boldsymbol{y}_2)\,.$$

For $\boldsymbol{x} \in X$ and $\boldsymbol{y} \in Y$, the ordered pair $(\boldsymbol{x}, \boldsymbol{y})$ is also written as $\boldsymbol{x} \oplus \boldsymbol{y}$.

§3.2 Direct Sum and Multi-Linear Maps

Remark:

1. The direct sum is a way of getting a new big vector space from two (or more) smaller vector spaces in the simplest way one can imagine: you just **line them up**.
2. Let X, Y be finite dimensional vector spaces over a scalar field $\mathbb{F}$, where $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. Then $X \oplus Y$ is a finite dimensional vector space over $\mathbb{F}$ and $\dim(X \oplus Y) = \dim(X) + \dim(Y)$. In fact, if $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_m\}$ is a basis of X and $\{\boldsymbol{y}_1, \cdots, \boldsymbol{y}_n\}$ is a basis of Y, then $X \oplus Y$ has a basis
$$\{\boldsymbol{x}_1 \oplus \mathbf{0}, \boldsymbol{x}_2 \oplus \mathbf{0}, \cdots, \boldsymbol{x}_m \oplus \mathbf{0}, \mathbf{0} \oplus \boldsymbol{y}_1, \mathbf{0} \oplus \boldsymbol{y}_2, \cdots, \mathbf{0} \oplus \boldsymbol{y}_n\}.$$
This basis is called the **induced basis** of basis $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_m\}$ of X and basis $\{\boldsymbol{y}_1, \cdots, \boldsymbol{y}_n\}$ of Y.

§3.2 Direct Sum and Multi-Linear Maps

Remark:

1. The direct sum is a way of getting a new big vector space from two (or more) smaller vector spaces in the simplest way one can imagine: you just **line them up**.
2. Let X, Y be finite dimensional vector spaces over a scalar field $\mathbb{F}$, where $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. Then $X \oplus Y$ is a finite dimensional vector space over $\mathbb{F}$ and $\dim(X \oplus Y) = \dim(X) + \dim(Y)$. In fact, if $\{x_1, \cdots, x_m\}$ is a basis of X and $\{y_1, \cdots, y_n\}$ is a basis of Y, then $X \oplus Y$ has a basis
$$\{x_1 \oplus \mathbf{0}, x_2 \oplus \mathbf{0}, \cdots, x_m \oplus \mathbf{0}, \mathbf{0} \oplus y_1, \mathbf{0} \oplus y_2, \cdots, \mathbf{0} \oplus y_n\}.$$
This basis is called the **induced basis** of basis $\{x_1, \cdots, x_m\}$ of X and basis $\{y_1, \cdots, y_n\}$ of Y.

§3.2 Direct Sum and Multi-Linear Maps

Remark:

1. The direct sum is a way of getting a new big vector space from two (or more) smaller vector spaces in the simplest way one can imagine: you just **line them up**.
2. Let X, Y be finite dimensional vector spaces over a scalar field $\mathbb{F}$, where $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. Then $X \oplus Y$ is a finite dimensional vector space over $\mathbb{F}$ and $\dim(X \oplus Y) = \dim(X) + \dim(Y)$. In fact, if $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_m\}$ is a basis of X and $\{\boldsymbol{y}_1, \cdots, \boldsymbol{y}_n\}$ is a basis of Y, then $X \oplus Y$ has a basis

$$\left\{\boldsymbol{x}_1 \oplus \mathbf{0}, \boldsymbol{x}_2 \oplus \mathbf{0}, \cdots, \boldsymbol{x}_m \oplus \mathbf{0}, \mathbf{0} \oplus \boldsymbol{y}_1, \mathbf{0} \oplus \boldsymbol{y}_2, \cdots, \mathbf{0} \oplus \boldsymbol{y}_n\right\}.$$

This basis is called the **induced basis** of basis $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_m\}$ of X and basis $\{\boldsymbol{y}_1, \cdots, \boldsymbol{y}_n\}$ of Y.

§3.2 Direct Sum and Multi-Linear Maps

Remark:

1. The direct sum is a way of getting a new big vector space from two (or more) smaller vector spaces in the simplest way one can imagine: you just **line them up**.
2. Let X, Y be finite dimensional vector spaces over a scalar field $\mathbb{F}$, where $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. Then $X \oplus Y$ is a finite dimensional vector space over $\mathbb{F}$ and $\dim(X \oplus Y) = \dim(X) + \dim(Y)$. In fact, if $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_m\}$ is a basis of X and $\{\boldsymbol{y}_1, \cdots, \boldsymbol{y}_n\}$ is a basis of Y, then $X \oplus Y$ has a basis
$$\left\{\boldsymbol{x}_1 \oplus \mathbf{0}, \boldsymbol{x}_2 \oplus \mathbf{0}, \cdots, \boldsymbol{x}_m \oplus \mathbf{0}, \mathbf{0} \oplus \boldsymbol{y}_1, \mathbf{0} \oplus \boldsymbol{y}_2, \cdots, \mathbf{0} \oplus \boldsymbol{y}_n\right\}.$$
This basis is called the **induced basis** of basis $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_m\}$ of X and basis $\{\boldsymbol{y}_1, \cdots, \boldsymbol{y}_n\}$ of Y.

§3.2 Direct Sum and Multi-Linear Maps

Definition

Let X, Y, Z, W be vector spaces over a common scalar field $\mathbb{F}$, and $A \in \mathcal{L}(X;Z)$, $B \in \mathcal{L}(Y;W)$. The direct sum of A and B, denoted by $A \oplus B$, is a linear map in $\mathcal{L}(X \oplus Y; Z \oplus W)$ satisfying that

$$(A \oplus B)(\boldsymbol{x} \oplus \boldsymbol{y}) = (A\boldsymbol{x}) \oplus (B\boldsymbol{y}) \quad \forall\, \boldsymbol{x} \in X, \boldsymbol{y} \in Y.$$

Theorem

Let X_1, X_2, X_3, Y_1, Y_2, Y_3 be vectors spaces over a common scalar field $\mathbb{F}$, and $A_1 \in \mathcal{L}(X_1;X_2)$, $A_2 \in \mathcal{L}(X_2;X_3)$, $B_1 \in \mathcal{L}(Y_1;Y_2)$, $B_2 \in \mathcal{L}(Y_2;Y_3)$. Then $(A_2 \oplus B_2)(A_1 \oplus B_1) = (A_2A_1) \oplus (B_2B_1)$.

Proof.

Let $\boldsymbol{x} \in X_1$ and $\boldsymbol{y} \in Y_1$. Then

$$\begin{aligned}(A_2 \oplus B_2)(A_1 \oplus B_1)(\boldsymbol{x} \oplus \boldsymbol{y}) &= (A_2 \oplus B_2)(A_1\boldsymbol{x} \oplus B_1\boldsymbol{y}) \\ &= (A_2A_1\boldsymbol{x} \oplus B_2B_1\boldsymbol{y}) = (A_2A_1 \oplus B_2B_1)(\boldsymbol{x} \oplus \boldsymbol{y}).\end{aligned}$$

□

§3.2 Direct Sum and Multi-Linear Maps

Definition

Let X, Y, Z, W be vector spaces over a common scalar field $\mathbb{F}$, and $A \in \mathcal{L}(X;Z)$, $B \in \mathcal{L}(Y;W)$. The direct sum of A and B, denoted by $A \oplus B$, is a linear map in $\mathcal{L}(X \oplus Y; Z \oplus W)$ satisfying that

$$(A \oplus B)(\boldsymbol{x} \oplus \boldsymbol{y}) = (A\boldsymbol{x}) \oplus (B\boldsymbol{y}) \qquad \forall\, \boldsymbol{x} \in X, \boldsymbol{y} \in Y\,.$$

Theorem

Let X_1, X_2, X_3, Y_1, Y_2, Y_3 be vectors spaces over a common scalar field $\mathbb{F}$, and $A_1 \in \mathcal{L}(X_1;X_2)$, $A_2 \in \mathcal{L}(X_2;X_3)$, $B_1 \in \mathcal{L}(Y_1;Y_2)$, $B_2 \in \mathcal{L}(Y_2;Y_3)$. Then $(A_2 \oplus B_2)(A_1 \oplus B_1) = (A_2A_1) \oplus (B_2B_1)$.

Proof.

Let $\boldsymbol{x} \in X_1$ and $\boldsymbol{y} \in Y_1$. Then

$$\begin{aligned}(A_2 \oplus B_2)(A_1 \oplus B_1)(\boldsymbol{x} \oplus \boldsymbol{y}) &= (A_2 \oplus B_2)(A_1\boldsymbol{x} \oplus B_1\boldsymbol{y}) \\ &= (A_2A_1\boldsymbol{x} \oplus B_2B_1\boldsymbol{y}) = (A_2A_1 \oplus B_2B_1)(\boldsymbol{x} \oplus \boldsymbol{y})\,.\end{aligned}$$

□

§3.2 Direct Sum and Multi-Linear Maps

Definition

Let X, Y, Z, W be vector spaces over a common scalar field $\mathbb{F}$, and $A \in \mathscr{L}(X; Z)$, $B \in \mathscr{L}(Y; W)$. The direct sum of A and B, denoted by $A \oplus B$, is a linear map in $\mathscr{L}(X \oplus Y; Z \oplus W)$ satisfying that

$$(A \oplus B)(\boldsymbol{x} \oplus \boldsymbol{y}) = (A\boldsymbol{x}) \oplus (B\boldsymbol{y}) \qquad \forall\, \boldsymbol{x} \in X, \boldsymbol{y} \in Y.$$

Theorem

Let X_1, X_2, X_3, Y_1, Y_2, Y_3 be vectors spaces over a common scalar field $\mathbb{F}$, and $A_1 \in \mathscr{L}(X_1; X_2)$, $A_2 \in \mathscr{L}(X_2; X_3)$, $B_1 \in \mathscr{L}(Y_1; Y_2)$, $B_2 \in \mathscr{L}(Y_2; Y_3)$. Then $(A_2 \oplus B_2)(A_1 \oplus B_1) = (A_2A_1) \oplus (B_2B_1)$.

Proof.

Let $\boldsymbol{x} \in X_1$ and $\boldsymbol{y} \in Y_1$. Then

$$\begin{aligned}(A_2 \oplus B_2)(A_1 \oplus B_1)(\boldsymbol{x} \oplus \boldsymbol{y}) &= (A_2 \oplus B_2)(A_1\boldsymbol{x} \oplus B_1\boldsymbol{y}) \\ &= (A_2A_1\boldsymbol{x} \oplus B_2B_1\boldsymbol{y}) = (A_2A_1 \oplus B_2B_1)(\boldsymbol{x} \oplus \boldsymbol{y}).\end{aligned}$$

□

§3.2 Direct Sum and Multi-Linear Maps

The following theorem concerns with the matrix representation of $A \oplus B$ if A, B are linear maps.

Theorem

Let X_1, X_2, Y_1, Y_2 be finite dimensional vector spaces over field $\mathbb{F}$, and $A \in \mathcal{L}(X_1; X_2)$, $B \in \mathcal{L}(Y_1; Y_2)$. Suppose that relative to given basis of X_1, X_2, Y_1, Y_2, the matrix representations of A and B are $[A]$ and $[B]$, respectively. Then relative to the induced basis of $X_1 \oplus Y_1$ and $X_2 \oplus Y_2$ of given basis of X_1, X_2, Y_1, Y_2, the matrix representation of $A \oplus B$ is $\begin{bmatrix} [A] & \mathbf{0} \\ \mathbf{0} & [B] \end{bmatrix}$.

§3.2 Direct Sum and Multi-Linear Maps

§3.2.2 Multi-Linear Maps

Definition

Let $\mathbb{V}_1, \mathbb{V}_2, \mathbb{W}$ be vector spaces over a common scalar field $\mathbb{F}$. A map $L : \mathbb{V}_1 \oplus \mathbb{V}_2 \to \mathbb{W}$ is said to be bilinear provided that

$$L(c\boldsymbol{u} + \boldsymbol{v}, \boldsymbol{w}) = cL(\boldsymbol{u}, \boldsymbol{w}) + L(\boldsymbol{v}, \boldsymbol{w}) \quad \forall\, \boldsymbol{u}, \boldsymbol{v} \in \mathbb{V}_1, \boldsymbol{w} \in \mathbb{V}_2 \text{ and } c \in \mathbb{F},$$

$$L(\boldsymbol{u}, c\boldsymbol{v} + \boldsymbol{w}) = cL(\boldsymbol{u}, \boldsymbol{v}) + L(\boldsymbol{u}, \boldsymbol{w}) \quad \forall\, \boldsymbol{u} \in \mathbb{V}_1, \boldsymbol{v}, \boldsymbol{w} \in \mathbb{V}_2 \text{ and } c \in \mathbb{F}.$$

The collection of all maps $L : \mathbb{V}_1 \oplus \mathbb{V}_2 \to \mathbb{W}$ satisfying two identities above is denoted by $\mathscr{L}(\mathbb{V}_1, \mathbb{V}_2; \mathbb{W})$.

§3.2 Direct Sum and Multi-Linear Maps

The extension of the bilinearity is the multi-linearity given by

Definition

Let $\mathbb{V}_1, \cdots, \mathbb{V}_n, \mathbb{W}$ be vector spaces over a common scalar field $\mathbb{F}$. A map $L : \mathbb{V}_1 \oplus \cdots \oplus \mathbb{V}_n \to \mathbb{W}$ is said to be multi-linear, denoted by $L \in \mathcal{L}(\mathbb{V}_1, \cdots, \mathbb{V}_n; \mathbb{W})$, provided that

$$\begin{aligned} & L(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_{j-1}, c\boldsymbol{v}_j + \boldsymbol{w}_j, \boldsymbol{u}_{j+1}, \cdots, \boldsymbol{u}_n) \\ & \quad = c\, L(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_{j-1}, \boldsymbol{v}_j, \boldsymbol{u}_{j+1}, \cdots, \boldsymbol{u}_n) + L(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_{j-1}, \boldsymbol{w}_j, \boldsymbol{u}_{j+1}, \cdots, \boldsymbol{u}_n) \end{aligned}$$

for all $1 \leqslant j \leqslant n$ and $c \in \mathbb{F}$, and $\boldsymbol{u}_\ell \in \mathbb{V}_\ell$ for all $\ell \neq j$, $\boldsymbol{v}_j, \boldsymbol{w}_j \in \mathbb{V}_j$.

Remark: One can think of the space of linear maps $\mathcal{L}(\mathbb{V}; \mathbb{W})$ as collection of “one”-linear maps.

§3.2 Direct Sum and Multi-Linear Maps

The extension of the bilinearity is the multi-linearity given by

Definition

Let $\mathbb{V}_1, \cdots, \mathbb{V}_n, \mathbb{W}$ be vector spaces over a common scalar field $\mathbb{F}$. A map $L : \mathbb{V}_1 \oplus \cdots \oplus \mathbb{V}_n \to \mathbb{W}$ is said to be multi-linear, denoted by $L \in \mathcal{L}(\mathbb{V}_1, \cdots, \mathbb{V}_n; \mathbb{W})$, provided that

$$\begin{aligned} & L(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_{j-1}, c\boldsymbol{v}_j + \boldsymbol{w}_j, \boldsymbol{u}_{j+1}, \cdots, \boldsymbol{u}_n) \\ & \quad = c\,L(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_{j-1}, \boldsymbol{v}_j, \boldsymbol{u}_{j+1}, \cdots, \boldsymbol{u}_n) + L(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_{j-1}, \boldsymbol{w}_j, \boldsymbol{u}_{j+1}, \cdots, \boldsymbol{u}_n) \end{aligned}$$

for all $1 \leqslant j \leqslant n$ and $c \in \mathbb{F}$, and $\boldsymbol{u}_\ell \in \mathbb{V}_\ell$ for all $\ell \neq j$, $\boldsymbol{v}_j, \boldsymbol{w}_j \in \mathbb{V}_j$.

Remark: One can think of the space of linear maps $\mathcal{L}(\mathbb{V}; \mathbb{W})$ as collection of "one"-linear maps.

§3.2 Direct Sum and Multi-Linear Maps

Theorem

Let $\mathbb{V}_1, \cdots, \mathbb{V}_n, \mathbb{W}$ be vector spaces over a common scalar field $\mathbb{F}$. Then $\mathscr{L}(\mathbb{V}_1, \cdots, \mathbb{V}_n; \mathbb{W})$ is a vector space over $\mathbb{F}$.

Proof.

Let $f, g \in \mathscr{L}(\mathbb{V}_1, \cdots, \mathbb{V}_n)$, and $\alpha \in \mathbb{F}$. Then if $1 \leqslant j \leqslant n$, $c \in \mathbb{F}$, and $\boldsymbol{u}_\ell \in \mathbb{V}_\ell$ for all $\ell \neq j$, $\boldsymbol{v}_j, \boldsymbol{w}_j \in \mathbb{V}_j$, we have

$$\begin{aligned}
&(\alpha f + g)(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_{j-1}, c\boldsymbol{v}_j + \boldsymbol{w}_j, \boldsymbol{u}_{j+1}, \cdots, \boldsymbol{u}_n) \\
&\quad = \alpha f(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_{j-1}, c\boldsymbol{v}_j + \boldsymbol{w}_j, \boldsymbol{u}_{j+1}, \cdots, \boldsymbol{u}_n) \\
&\qquad + g(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_{j-1}, c\boldsymbol{v}_j + \boldsymbol{w}_j, \boldsymbol{u}_{j+1}, \cdots, \boldsymbol{u}_n) \\
&\quad = \alpha\big[cf(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_{j-1}, \boldsymbol{v}_j, \boldsymbol{u}_{j+1}, \cdots, \boldsymbol{u}_n) + f(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_j, \boldsymbol{w}_j, \boldsymbol{u}_{j+1}, \cdots, \boldsymbol{u}_n)\big] \\
&\qquad + cg(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_{j-1}, \boldsymbol{v}_j, \boldsymbol{u}_{j+1}, \cdots, \boldsymbol{u}_n) + g(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_j, \boldsymbol{w}_j, \boldsymbol{u}_{j+1}, \cdots, \boldsymbol{u}_n) \\
&\quad = c(\alpha f + g)(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_{j-1}, \boldsymbol{v}_j, \boldsymbol{u}_{j+1}, \cdots, \boldsymbol{u}_n) \\
&\qquad + (\alpha f + g)(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_{j-1}, \boldsymbol{w}_j, \boldsymbol{u}_{j+1}, \cdots, \boldsymbol{u}_n).
\end{aligned}$$

□

§3.2 Direct Sum and Multi-Linear Maps

Theorem

Let $\mathbb{V}_1, \cdots, \mathbb{V}_n, \mathbb{W}$ *be vector spaces over a common scalar field* $\mathbb{F}$. *Then* $\mathcal{L}(\mathbb{V}_1, \cdots, \mathbb{V}_n; \mathbb{W})$ *is a vector space over* $\mathbb{F}$.

Proof.

Let $f, g \in \mathcal{L}(\mathbb{V}_1, \cdots, \mathbb{V}_n)$, and $\alpha \in \mathbb{F}$. Then if $1 \leqslant j \leqslant n$, $c \in \mathbb{F}$, and $\boldsymbol{u}_\ell \in \mathbb{V}_\ell$ for all $\ell \neq j$, $\boldsymbol{v}_j, \boldsymbol{w}_j \in \mathbb{V}_j$, we have

$$\begin{aligned}
&(\alpha f + g)(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_{j-1}, c\boldsymbol{v}_j + \boldsymbol{w}_j, \boldsymbol{u}_{j+1}, \cdots, \boldsymbol{u}_n) \\
&\quad = \alpha f(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_{j-1}, c\boldsymbol{v}_j + \boldsymbol{w}_j, \boldsymbol{u}_{j+1}, \cdots, \boldsymbol{u}_n) \\
&\qquad + g(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_{j-1}, c\boldsymbol{v}_j + \boldsymbol{w}_j, \boldsymbol{u}_{j+1}, \cdots, \boldsymbol{u}_n) \\
&\quad = \alpha\big[cf(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_{j-1}, \boldsymbol{v}_j, \boldsymbol{u}_{j+1}, \cdots, \boldsymbol{u}_n) + f(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_j, \boldsymbol{w}_j, \boldsymbol{u}_{j+1}, \cdots, \boldsymbol{u}_n)\big] \\
&\qquad + cg(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_{j-1}, \boldsymbol{v}_j, \boldsymbol{u}_{j+1}, \cdots, \boldsymbol{u}_n) + g(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_j, \boldsymbol{w}_j, \boldsymbol{u}_{j+1}, \cdots, \boldsymbol{u}_n) \\
&\quad = c\,(\alpha f + g)(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_{j-1}, \boldsymbol{v}_j, \boldsymbol{u}_{j+1}, \cdots, \boldsymbol{u}_n) \\
&\qquad + (\alpha f + g)(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_{j-1}, \boldsymbol{w}_j, \boldsymbol{u}_{j+1}, \cdots, \boldsymbol{u}_n)\,.
\end{aligned}$$

□

§3.3 Inner Product Spaces and Hilbert Spaces

Definition

An ***inner product space*** $(\mathbb{V}, \langle \cdot, \cdot \rangle)$ is a vector space $\mathbb{V}$ over a scalar field $\mathbb{F}$ associated with a function $\langle \cdot, \cdot \rangle : \mathbb{V} \times \mathbb{V} \to \mathbb{F}$ such that

1. $\langle \boldsymbol{x}, \boldsymbol{x} \rangle \geqslant 0, \ \forall \, \boldsymbol{x} \in \mathbb{V}$.
2. $\langle \boldsymbol{x}, \boldsymbol{x} \rangle = 0$ if and only if $\boldsymbol{x} = 0$.
3. $\langle \boldsymbol{x}, \boldsymbol{y} + \boldsymbol{z} \rangle = \langle \boldsymbol{x}, \boldsymbol{y} \rangle + \langle \boldsymbol{x}, \boldsymbol{z} \rangle$ for all $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathbb{V}$.
4. $\langle \boldsymbol{x}, \lambda \boldsymbol{y} \rangle = \lambda \langle \boldsymbol{x}, \boldsymbol{y} \rangle$ for all $\lambda \in \mathbb{F}$ and $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{V}$.
5. $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \overline{\langle \boldsymbol{y}, \boldsymbol{x} \rangle}$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{V}$, where $\bar{c}$ denotes the complex conjugate of c.

A function $\langle \cdot, \cdot \rangle$ satisfying ①-⑤ is called an ***inner product*** on $\mathbb{V}$.

Remark: Properties ③ and ④ are called the **right-linearity** of inner products, while ⑤ is called the **left-antilinearity** of inner products.

§3.3 Inner Product Spaces and Hilbert Spaces

Definition

An ***inner product space*** $(\mathbb{V}, \langle\cdot,\cdot\rangle)$ is a vector space $\mathbb{V}$ over a scalar field $\mathbb{F}$ associated with a function $\langle\cdot,\cdot\rangle : \mathbb{V}\times\mathbb{V}\to\mathbb{F}$ such that

1. $\langle \boldsymbol{x}, \boldsymbol{x}\rangle \geqslant 0, \ \forall\, \boldsymbol{x} \in \mathbb{V}$.
2. $\langle \boldsymbol{x}, \boldsymbol{x}\rangle = 0$ if and only if $\boldsymbol{x} = 0$.
3. $\langle \boldsymbol{x}, \boldsymbol{y} + \boldsymbol{z}\rangle = \langle \boldsymbol{x}, \boldsymbol{y}\rangle + \langle \boldsymbol{x}, \boldsymbol{z}\rangle$ for all $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathbb{V}$.
4. $\langle \boldsymbol{x}, \lambda\boldsymbol{y}\rangle = \lambda\langle \boldsymbol{x}, \boldsymbol{y}\rangle$ for all $\lambda \in \mathbb{F}$ and $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{V}$.
5. $\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \overline{\langle \boldsymbol{y}, \boldsymbol{x}\rangle}$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{V}$, where $\bar{c}$ denotes the complex conjugate of c.

A function $\langle\cdot,\cdot\rangle$ satisfying ①-⑤ is called an ***inner product*** on $\mathbb{V}$.

Remark: Properties ③ and ④ are called the **right-linearity** of inner products, while ⑤ is called the **left-antilinearity** of inner products.

§3.3 Inner Product Spaces and Hilbert Spaces

Proposition

Let $\langle\cdot,\cdot\rangle$ be an inner product on a vector space $\mathbb{V}$ over a scalar field $\mathbb{F}$. Then

1. $\langle \boldsymbol{u}, \lambda\boldsymbol{v} + \mu\boldsymbol{w}\rangle = \lambda\langle \boldsymbol{u}, \boldsymbol{v}\rangle + \mu\langle \boldsymbol{u}, \boldsymbol{w}\rangle$ *for all* $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in \mathbb{V}$, $\lambda, \mu \in \mathbb{F}$.
2. $\langle \lambda\boldsymbol{v} + \mu\boldsymbol{w}, \boldsymbol{u}\rangle = \bar{\lambda}\langle \boldsymbol{v}, \boldsymbol{u}\rangle + \bar{\mu}\langle \boldsymbol{w}, \boldsymbol{u}\rangle$ *for all* $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in \mathbb{V}$, $\lambda, \mu \in \mathbb{F}$.
3. $\langle \mathbf{0}, \boldsymbol{w}\rangle = \langle \boldsymbol{w}, \mathbf{0}\rangle = 0$ *for all* $\boldsymbol{w} \in \mathbb{V}$.

Theorem

The inner product $\langle\cdot,\cdot\rangle$ on a vector space $\mathbb{V}$ over scalar field $\mathbb{F}$ satisfies the ***Cauchy-Schwarz inequality***

$$\big|\langle \boldsymbol{x}, \boldsymbol{y}\rangle\big| \leqslant \sqrt{\langle \boldsymbol{x}, \boldsymbol{x}\rangle}\sqrt{\langle \boldsymbol{y}, \boldsymbol{y}\rangle} \qquad \forall\, \boldsymbol{x}, \boldsymbol{y} \in \mathbb{V}\,.$$

Moreover, for non-zero vectors $\boldsymbol{x}, \boldsymbol{y}$, the equality holds if and only if there exists $\gamma \in \mathbb{F}$ such that $\boldsymbol{x} = \gamma\boldsymbol{y}$.

§3.3 Inner Product Spaces and Hilbert Spaces

Proposition

Let $\langle\cdot,\cdot\rangle$ be an inner product on a vector space $\mathbb{V}$ over a scalar field $\mathbb{F}$. Then

1. $\langle \boldsymbol{u}, \lambda\boldsymbol{v}+\mu\boldsymbol{w}\rangle = \lambda\langle\boldsymbol{u},\boldsymbol{v}\rangle + \mu\langle\boldsymbol{u},\boldsymbol{w}\rangle$ *for all* $\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}\in\mathbb{V}$, $\lambda,\mu\in\mathbb{F}$.
2. $\langle \lambda\boldsymbol{v}+\mu\boldsymbol{w}, \boldsymbol{u}\rangle = \bar{\lambda}\langle\boldsymbol{v},\boldsymbol{u}\rangle + \bar{\mu}\langle\boldsymbol{w},\boldsymbol{u}\rangle$ *for all* $\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}\in\mathbb{V}$, $\lambda,\mu\in\mathbb{F}$.
3. $\langle \mathbf{0}, \boldsymbol{w}\rangle = \langle\boldsymbol{w},\mathbf{0}\rangle = 0$ *for all* $\boldsymbol{w}\in\mathbb{V}$.

Theorem

*The inner product $\langle\cdot,\cdot\rangle$ on a vector space $\mathbb{V}$ over scalar field $\mathbb{F}$ satisfies **the Cauchy-Schwarz inequality***

$$\big|\langle\boldsymbol{x},\boldsymbol{y}\rangle\big| \leqslant \sqrt{\langle\boldsymbol{x},\boldsymbol{x}\rangle}\sqrt{\langle\boldsymbol{y},\boldsymbol{y}\rangle} \qquad \forall\,\boldsymbol{x},\boldsymbol{y}\in\mathbb{V}\,.$$

Moreover, for non-zero vectors $\boldsymbol{x},\boldsymbol{y}$, the equality holds if and only if there exists $\gamma\in\mathbb{F}$ such that $\boldsymbol{x}=\gamma\boldsymbol{y}$.

§3.3 Inner Product Spaces and Hilbert Spaces

Definition

A ***normed vector space*** (or simply ***normed space***) $(\mathbb{V}, \|\cdot\|)$ is a vector space $\mathbb{V}$ over a scalar field $\mathbb{F}$ associated with a function $\|\cdot\| : \mathbb{V} \to \mathbb{R}$ such that

1. $\|\boldsymbol{x}\| \geqslant 0$ for all $\boldsymbol{x} \in \mathbb{V}$.
2. $\|\boldsymbol{x}\| = 0$ if and only if $\boldsymbol{x} = \mathbf{0}$.
3. $\|\lambda \cdot \boldsymbol{x}\| = |\lambda| \cdot \|\boldsymbol{x}\|$ for all $\lambda \in \mathbb{F}$ and $\boldsymbol{x} \in \mathbb{V}$.
4. $\|\boldsymbol{x} + \boldsymbol{y}\| \leqslant \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{V}$.

A function $\|\cdot\|$ satisfying ①-④ is called a ***norm*** on $\mathbb{V}$.

Theorem

The inner product $\langle \cdot, \cdot \rangle$ on a vector space $\mathbb{V}$ (over scalar field $\mathbb{F}$) induces a norm $\|\cdot\|$ given by $\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$.

§3.3 Inner Product Spaces and Hilbert Spaces

Definition

A ***normed vector space*** (or simply ***normed space***) $(\mathbb{V}, \|\cdot\|)$ is a vector space $\mathbb{V}$ over a scalar field $\mathbb{F}$ associated with a function $\|\cdot\| : \mathbb{V} \to \mathbb{R}$ such that

1. $\|\boldsymbol{x}\| \geqslant 0$ for all $\boldsymbol{x} \in \mathbb{V}$.
2. $\|\boldsymbol{x}\| = 0$ if and only if $\boldsymbol{x} = \mathbf{0}$.
3. $\|\lambda \cdot \boldsymbol{x}\| = |\lambda| \cdot \|\boldsymbol{x}\|$ for all $\lambda \in \mathbb{F}$ and $\boldsymbol{x} \in \mathbb{V}$.
4. $\|\boldsymbol{x} + \boldsymbol{y}\| \leqslant \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{V}$.

A function $\|\cdot\|$ satisfying ①-④ is called a ***norm*** on $\mathbb{V}$.

Theorem

The inner product $\langle \cdot, \cdot \rangle$ on a vector space $\mathbb{V}$ (over scalar field $\mathbb{F}$) induces a norm $\|\cdot\|$ given by $\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$.

§3.3 Inner Product Spaces and Hilbert Spaces

Definition

A ***Banach space*** is a complete normed vector space, and a ***Hilbert space*** is a complete inner product space (that is, a Banach space whose norm is induced by the inner product).

Remark: In the definition above, the completeness of a normed vector space is defined as follows.

1. A sequence $\{x_n\}_{n=1}^{\infty}$ is called a **Cauchy sequence** in a normed space $(\mathbb{V}, \|\cdot\|)$ if
$$(\forall\,\varepsilon>0)(\exists\,N>0)\big(n,m\geqslant N\Rightarrow \|x_n-x_m\|<\varepsilon\big)\,.$$
2. A normed space $(\mathbb{V}, \|\cdot\|)$ is complete if every Cauchy sequence in $\mathbb{V}$ converges; that is, if $\{x_n\}_{n=1}^{\infty}$ is a Cauchy sequence in $\mathbb{V}$, then there exists $x\in\mathbb{V}$ such that $\lim\limits_{n\to\infty}\|x_n-x\|=0$.

§3.3 Inner Product Spaces and Hilbert Spaces

Definition

A ***Banach space*** is a complete normed vector space, and a ***Hilbert space*** is a complete inner product space (that is, a Banach space whose norm is induced by the inner product).

Remark: In the definition above, the completeness of a normed vector space is defined as follows.

1. A sequence $\{\boldsymbol{x}_n\}_{n=1}^{\infty}$ is called a **Cauchy sequence** in a normed space $(\mathbb{V}, \|\cdot\|)$ if
$$(\forall\, \varepsilon > 0)(\exists\, N > 0)\big(n, m \geqslant N \Rightarrow \|x_n - x_m\| < \varepsilon\big)\,.$$
2. A normed space $(\mathbb{V}, \|\cdot\|)$ is complete if every Cauchy sequence in $\mathbb{V}$ converges; that is, if $\{\boldsymbol{x}_n\}_{n=1}^{\infty}$ is a Cauchy sequence in $\mathbb{V}$, then there exists $\boldsymbol{x} \in \mathbb{V}$ such that $\lim\limits_{n\to\infty} \|\boldsymbol{x}_n - \boldsymbol{x}\| = 0$.

§3.3 Inner Product Spaces and Hilbert Spaces

Definition

A ***Banach space*** is a complete normed vector space, and a ***Hilbert space*** is a complete inner product space (that is, a Banach space whose norm is induced by the inner product).

Remark: In the definition above, the completeness of a normed vector space is defined as follows.

1. A sequence $\{\boldsymbol{x}_n\}_{n=1}^{\infty}$ is called a **Cauchy sequence** in a normed space $(\mathbb{V}, \|\cdot\|)$ if
$$(\forall\,\varepsilon > 0)(\exists\,N > 0)\big(n, m \geqslant N \Rightarrow \|x_n - x_m\| < \varepsilon\big)\,.$$
2. A normed space $(\mathbb{V}, \|\cdot\|)$ is complete if every Cauchy sequence in $\mathbb{V}$ converges; that is, if $\{\boldsymbol{x}_n\}_{n=1}^{\infty}$ is a Cauchy sequence in $\mathbb{V}$, then there exists $\boldsymbol{x} \in \mathbb{V}$ such that $\lim\limits_{n\to\infty} \|\boldsymbol{x}_n - \boldsymbol{x}\| = 0$.

§3.3 Inner Product Spaces and Hilbert Spaces

Definition

A ***Banach space*** is a complete normed vector space, and a ***Hilbert space*** is a complete inner product space (that is, a Banach space whose norm is induced by the inner product).

Remark: In the definition above, the completeness of a normed vector space is defined as follows.

1. A sequence $\{\boldsymbol{x}_n\}_{n=1}^{\infty}$ is called a **Cauchy sequence** in a normed space $(\mathbb{V}, \|\cdot\|)$ if
$$(\forall\,\varepsilon > 0)(\exists\, N > 0)\big(n, m \geqslant N \Rightarrow \|x_n - x_m\| < \varepsilon\big)\,.$$
2. A normed space $(\mathbb{V}, \|\cdot\|)$ is complete if every Cauchy sequence in $\mathbb{V}$ converges; that is, if $\{\boldsymbol{x}_n\}_{n=1}^{\infty}$ is a Cauchy sequence in $\mathbb{V}$, then there exists $\boldsymbol{x} \in \mathbb{V}$ such that $\lim\limits_{n\to\infty} \|\boldsymbol{x}_n - \boldsymbol{x}\| = 0$.

§3.3 Inner Product Spaces and Hilbert Spaces

Definition

Let $(\mathbb{V}, \langle\cdot,\cdot\rangle)$ be a finite dimensional inner product space. A basis $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ of $\mathbb{V}$ is said to be ***orthonormal*** if $\langle \mathbf{v}_i, \mathbf{v}_j\rangle = \delta_{ij}$ for all $1 \leqslant i, j \leqslant n$, where δ_{ij} is the Kronecker delta.

Let $(\mathbb{V}, \langle\cdot,\cdot\rangle_{\mathbb{V}})$ and $(\mathbb{W}, \langle\cdot,\cdot\rangle_{\mathbb{W}})$ be finite dimensional inner product spaces over $\mathbb{C}$, $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ and $\widetilde{\mathcal{B}} = \{\mathbf{w}_1, \cdots, \mathbf{w}_m\}$ be orthonormal basis of $\mathbb{V}$ and $\mathbb{W}$, respectively, and $L \in \mathscr{L}(\mathbb{V}; \mathbb{W})$. If $A = [L]_{\widetilde{\mathcal{B}},\mathcal{B}} = [a_{ij}]_{m\times n}$ be the matrix representation of L relative to B and $\widetilde{\mathcal{B}}$, then

$$\begin{aligned}\langle \boldsymbol{w}, L\boldsymbol{v}\rangle_{\mathbb{W}} &= \Big\langle \sum_{k=1}^{m} w_k \mathbf{w}_k, \sum_{i=1}^{m}\Big(\sum_{j=1}^{n} a_{ij} v_j\Big)\mathbf{w}_i \Big\rangle_{\mathbb{W}} = \sum_{j=1}^{n}\sum_{i,k=1}^{m} a_{ij} v_j \overline{w_k} \langle \mathbf{w}_k, \mathbf{w}_i\rangle_{\mathbb{W}} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij} v_j \bar{w}_i = \big\langle [\boldsymbol{w}]_{\widetilde{\mathcal{B}}}, A[\boldsymbol{v}]_{\mathcal{B}} \big\rangle_{\mathbb{C}^m} = \big\langle [\boldsymbol{w}]_{\widetilde{\mathcal{B}}}, [L]_{\widetilde{\mathcal{B}},\mathcal{B}}[\boldsymbol{v}]_{\mathcal{B}} \big\rangle_{\mathbb{C}^m}\end{aligned}$$

§3.3 Inner Product Spaces and Hilbert Spaces

Definition

Let $(\mathbb{V}, \langle\cdot,\cdot\rangle)$ be a finite dimensional inner product space. A basis $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ of $\mathbb{V}$ is said to be ***orthonormal*** if $\langle \mathbf{v}_i, \mathbf{v}_j\rangle = \delta_{ij}$ for all $1 \leqslant i, j \leqslant n$, where δ_{ij} is the Kronecker delta.

Let $(\mathbb{V}, \langle\cdot,\cdot\rangle_{\mathbb{V}})$ and $(\mathbb{W}, \langle\cdot,\cdot\rangle_{\mathbb{W}})$ be finite dimensional inner product spaces over $\mathbb{C}$, $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ and $\widetilde{\mathcal{B}} = \{\mathbf{w}_1, \cdots, \mathbf{w}_m\}$ be orthonormal basis of $\mathbb{V}$ and $\mathbb{W}$, respectively, and $L \in \mathscr{L}(\mathbb{V};\mathbb{W})$. If $A = [L]_{\widetilde{\mathcal{B}},\mathcal{B}} = [a_{ij}]_{m\times n}$ be the matrix representation of L relative to B and $\widetilde{\mathcal{B}}$, then

$$\begin{aligned}\langle \boldsymbol{w}, L\boldsymbol{v}\rangle_{\mathbb{W}} &= \Big\langle \sum_{k=1}^{m} w_k \mathbf{w}_k, \sum_{i=1}^{m}\Big(\sum_{j=1}^{n} a_{ij} v_j\Big)\mathbf{w}_i \Big\rangle_{\mathbb{W}} = \sum_{j=1}^{n}\sum_{i,k=1}^{m} a_{ij} v_j \overline{w_k}\langle \mathbf{w}_k, \mathbf{w}_i\rangle_{\mathbb{W}} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij} v_j \bar{w}_i = \Big\langle [\boldsymbol{w}]_{\widetilde{\mathcal{B}}}, A[\boldsymbol{v}]_{\mathcal{B}}\Big\rangle_{\mathbb{C}^m} = \Big\langle [\boldsymbol{w}]_{\widetilde{\mathcal{B}}}, [L]_{\widetilde{\mathcal{B}},\mathcal{B}}[\boldsymbol{v}]_{\mathcal{B}}\Big\rangle_{\mathbb{C}^m}\end{aligned}$$

§3.3 Inner Product Spaces and Hilbert Spaces

Definition

Let $(\mathbb{V}, \langle\cdot,\cdot\rangle)$ be a finite dimensional inner product space. A basis $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ of $\mathbb{V}$ is said to be ***orthonormal*** if $\langle \mathbf{v}_i, \mathbf{v}_j\rangle = \delta_{ij}$ for all $1 \leqslant i, j \leqslant n$, where δ_{ij} is the Kronecker delta.

Let $(\mathbb{V}, \langle\cdot,\cdot\rangle_{\mathbb{V}})$ and $(\mathbb{W}, \langle\cdot,\cdot\rangle_{\mathbb{W}})$ be finite dimensional inner product spaces over $\mathbb{C}$, $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ and $\widetilde{\mathcal{B}} = \{\mathbf{w}_1, \cdots, \mathbf{w}_m\}$ be orthonormal basis of $\mathbb{V}$ and $\mathbb{W}$, respectively, and $L \in \mathcal{L}(\mathbb{V};\mathbb{W})$. If $A = [L]_{\widetilde{\mathcal{B}},\mathcal{B}} = [a_{ij}]_{m\times n}$ be the matrix representation of L relative to B and $\widetilde{\mathcal{B}}$, then

$$\langle \boldsymbol{w}, L\boldsymbol{v}\rangle_{\mathbb{W}} = \Big\langle \sum_{k=1}^{m} w_k \mathbf{w}_k, \sum_{i=1}^{m}\Big(\sum_{j=1}^{n} a_{ij} v_j\Big)\mathbf{w}_i \Big\rangle_{\mathbb{W}} = \sum_{j=1}^{n}\sum_{i,k=1}^{m} a_{ij} v_j \overline{w_k} \langle \mathbf{w}_k, \mathbf{w}_i\rangle_{\mathbb{W}}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij} v_j \bar{w}_i = \big\langle [\boldsymbol{w}]_{\widetilde{\mathcal{B}}}, A[\boldsymbol{v}]_{\mathcal{B}} \big\rangle_{\mathbb{C}^m} = \big\langle [\boldsymbol{w}]_{\widetilde{\mathcal{B}}}, [L]_{\widetilde{\mathcal{B}},\mathcal{B}}[\boldsymbol{v}]_{\mathcal{B}} \big\rangle_{\mathbb{C}^m}$$

§3.3 Inner Product Spaces and Hilbert Spaces

so that

$$\langle \boldsymbol{w}, L\boldsymbol{v}\rangle_{\mathbb{W}} = \Big\langle [\boldsymbol{w}]_{\widetilde{\mathcal{B}}}, [L]_{\widetilde{\mathcal{B}},\mathcal{B}}[\boldsymbol{v}]_{\mathcal{B}}\Big\rangle_{\mathbb{C}^m} = \Big\langle [\boldsymbol{w}]_{\widetilde{\mathcal{B}}}, [L\boldsymbol{v}]_{\widetilde{\mathcal{B}}}\Big\rangle_{\mathbb{C}^m}, .$$

The identity above converts the computation of the inner product of $\boldsymbol{w}$ and $L\boldsymbol{v}$ in $\mathbb{W}$ in terms of the inner product of $[\boldsymbol{w}]_{\widetilde{\mathcal{B}}}$ and $[L\boldsymbol{v}]_{\widetilde{\mathcal{B}}}(= [L]_{\widetilde{\mathcal{B}},\mathcal{B}}[\boldsymbol{v}]_{\mathcal{B}})$ in $\mathbb{C}^m$ using the matrix representation of L and matrix multiplications.

In general, if $L_1, L_2 \in \mathcal{L}(\mathbb{V})$ and $\mathcal{B}$ is an orthonormal basis of $\mathbb{V}$, then

$$\langle L_1\boldsymbol{u}, L_2\boldsymbol{v}\rangle_{\mathbb{V}} = \Big\langle [L_1]_{\mathcal{B}}[\boldsymbol{u}]_{\mathcal{B}}, [L_2]_{\mathcal{B}}[\boldsymbol{v}]_{\mathcal{B}}\Big\rangle_{\mathbb{C}^n} \qquad \forall\, \boldsymbol{u}, \boldsymbol{v} \in \mathbb{V}$$

since

$$\langle L_1\boldsymbol{u}, L_2\boldsymbol{v}\rangle_{\mathbb{V}} = \big\langle [L_1\boldsymbol{u}]_{\mathcal{B}}, [L_2]_{\mathcal{B}}[\boldsymbol{v}]_{\mathcal{B}}\big\rangle_{\mathbb{C}^n} = \big\langle [L_1]_{\mathcal{B}}[\boldsymbol{u}]_{\mathcal{B}}, [L_2]_{\mathcal{B}}[\boldsymbol{v}]_{\mathcal{B}}\big\rangle_{\mathbb{C}^n} .$$

§3.3 Inner Product Spaces and Hilbert Spaces

so that

$$\langle \boldsymbol{w}, L\boldsymbol{v}\rangle_{\mathbb{W}} = \big\langle [\boldsymbol{w}]_{\widetilde{\mathcal{B}}}, [L]_{\widetilde{\mathcal{B}},\mathcal{B}}[\boldsymbol{v}]_{\mathcal{B}}\big\rangle_{\mathbb{C}^m} = \big\langle [\boldsymbol{w}]_{\widetilde{\mathcal{B}}}, [L\boldsymbol{v}]_{\widetilde{\mathcal{B}}}\big\rangle_{\mathbb{C}^m}, .$$

The identity above converts the computation of the inner product of $\boldsymbol{w}$ and $L\boldsymbol{v}$ in $\mathbb{W}$ in terms of the inner product of $[\boldsymbol{w}]_{\widetilde{\mathcal{B}}}$ and $[L\boldsymbol{v}]_{\widetilde{\mathcal{B}}}(= [L]_{\widetilde{\mathcal{B}},\mathcal{B}}[\boldsymbol{v}]_{\mathcal{B}})$ in $\mathbb{C}^m$ using the matrix representation of L and matrix multiplications.

In general, if $L_1, L_2 \in \mathcal{L}(\mathbb{V})$ and $\mathcal{B}$ is an orthonormal basis of $\mathbb{V}$, then

$$\langle L_1\boldsymbol{u}, L_2\boldsymbol{v}\rangle_{\mathbb{V}} = \big\langle [L_1]_{\mathcal{B}}[\boldsymbol{u}]_{\mathcal{B}}, [L_2]_{\mathcal{B}}[\boldsymbol{v}]_{\mathcal{B}}\big\rangle_{\mathbb{C}^n} \qquad \forall\, \boldsymbol{u}, \boldsymbol{v} \in \mathbb{V}$$

since

$$\langle L_1\boldsymbol{u}, L_2\boldsymbol{v}\rangle_{\mathbb{V}} = \big\langle [L_1\boldsymbol{u}]_{\mathcal{B}}, [L_2]_{\mathcal{B}}[\boldsymbol{v}]_{\mathcal{B}}\rangle_{\mathbb{C}^n} = \big\langle [L_1]_{\mathcal{B}}[\boldsymbol{u}]_{\mathcal{B}}, [L_2]_{\mathcal{B}}[\boldsymbol{v}]_{\mathcal{B}}\rangle_{\mathbb{C}^n} .$$

§3.4 Dual Spaces and Adjoint Operators

Definition

Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces over a common scalar field $\mathbb{F}$ equipped with norms $\|\cdot\|_{\mathbb{V}}$ and $\|\cdot\|_{\mathbb{W}}$, respectively. A linear map $L : \mathbb{V} \to \mathbb{W}$ is said to be bounded if the number

$$\|L\|_{\mathcal{B}(\mathbb{V},\mathbb{W})} \equiv \sup_{\|\boldsymbol{x}\|_{\mathbb{V}}=1} \|L\boldsymbol{x}\|_{\mathbb{W}} < \infty\,.$$

The collection of all bounded linear maps from $\mathbb{V}$ to $\mathbb{W}$ is denoted by $\mathcal{B}(\mathbb{V},\mathbb{W})$. When $\mathbb{W} = \mathbb{V}$, we write $\mathcal{B}(\mathbb{V})$ instead of $\mathcal{B}(\mathbb{V},\mathbb{V})$.

Definition

Let X be a Banach space over scalar field $\mathbb{F}$. The (continuous) dual space of X, denoted by X^*, is the collection of all bounded linear functionals on X; that is,

$$X^* = \left\{ L \in \mathcal{L}(X,\mathbb{F}) \,\middle|\, \sup_{\|\boldsymbol{x}\|_X=1} \big|L(\boldsymbol{x})\big| < \infty \right\}.$$

§3.4 Dual Spaces and Adjoint Operators

Definition

Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces over a common scalar field $\mathbb{F}$ equipped with norms $\|\cdot\|_{\mathbb{V}}$ and $\|\cdot\|_{\mathbb{W}}$, respectively. A linear map $L : \mathbb{V} \to \mathbb{W}$ is said to be bounded if the number

$$\|L\|_{\mathcal{B}(\mathbb{V},\mathbb{W})} \equiv \sup_{\|\boldsymbol{x}\|_{\mathbb{V}}=1} \|L\boldsymbol{x}\|_{\mathbb{W}} < \infty\,.$$

The collection of all bounded linear maps from $\mathbb{V}$ to $\mathbb{W}$ is denoted by $\mathcal{B}(\mathbb{V},\mathbb{W})$. When $\mathbb{W}=\mathbb{V}$, we write $\mathcal{B}(\mathbb{V})$ instead of $\mathcal{B}(\mathbb{V},\mathbb{V})$.

Definition

Let X be a Banach space over scalar field $\mathbb{F}$. The (continuous) dual space of X, denoted by X^*, is the collection of all bounded linear functionals on X; that is,

$$X^* = \Big\{ L \in \mathcal{L}(X,\mathbb{F}) \,\Big|\, \sup_{\|\boldsymbol{x}\|_X=1} \big|L(\boldsymbol{x})\big| < \infty \Big\}\,.$$

§3.4 Dual Spaces and Adjoint Operators

Definition

Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces over a common scalar field $\mathbb{F}$ equipped with norms $\|\cdot\|_{\mathbb{V}}$ and $\|\cdot\|_{\mathbb{W}}$, respectively. A linear map $L : \mathbb{V} \to \mathbb{W}$ is said to be bounded if the number

$$\|L\|_{\mathscr{B}(\mathbb{V},\mathbb{W})} \equiv \sup_{\|\boldsymbol{x}\|_{\mathbb{V}}=1} \|L\boldsymbol{x}\|_{\mathbb{W}} < \infty\,.$$

The collection of all bounded linear maps from $\mathbb{V}$ to $\mathbb{W}$ is denoted by $\mathscr{B}(\mathbb{V},\mathbb{W})$. When $\mathbb{W} = \mathbb{V}$, we write $\mathscr{B}(\mathbb{V})$ instead of $\mathscr{B}(\mathbb{V},\mathbb{V})$.

Definition

Let X be a Banach space over scalar field $\mathbb{F}$. The (continuous) dual space of X, denoted by X^*, is the collection of all bounded linear maps from X to $\mathbb{F}$; that is,

$$X^* = \Big\{ L \in \mathscr{L}(X,\mathbb{F}) \,\Big|\, \sup_{\|\boldsymbol{x}\|_X=1} \big|L(\boldsymbol{x})\big| < \infty \Big\}\,.$$

§3.4 Dual Spaces and Adjoint Operators

Definition

Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces over a common scalar field $\mathbb{F}$ equipped with norms $\|\cdot\|_{\mathbb{V}}$ and $\|\cdot\|_{\mathbb{W}}$, respectively. A linear map $L:\mathbb{V}\to\mathbb{W}$ is said to be bounded if the number

$$\|L\|_{\mathscr{B}(\mathbb{V},\mathbb{W})} \equiv \sup_{\|\boldsymbol{x}\|_{\mathbb{V}}=1} \|L\boldsymbol{x}\|_{\mathbb{W}} < \infty\,.$$

The collection of all bounded linear maps from $\mathbb{V}$ to $\mathbb{W}$ is denoted by $\mathscr{B}(\mathbb{V},\mathbb{W})$. When $\mathbb{W}=\mathbb{V}$, we write $\mathscr{B}(\mathbb{V})$ instead of $\mathscr{B}(\mathbb{V},\mathbb{V})$.

Definition

Let X be a Banach space over scalar field $\mathbb{F}$. The (continuous) dual space of X, denoted by X^*, is the collection of all bounded linear maps from X to $\mathbb{F}$; that is,

$$X^* = \Big\{ L\in\mathscr{L}(X,\mathbb{F}) \,\Big|\, \sup_{\|\boldsymbol{x}\|_X=1} \big|L(\boldsymbol{x})\big| < \infty \Big\}\,.$$

§3.4 Dual Spaces and Adjoint Operators

Theorem

If $(\mathbb{H}, \langle \cdot, \cdot \rangle)$ *is a finite dimensional Hilbert space over field* $\mathbb{F}$*, then* $\mathbb{H}^*$ *is also finite dimensional and* $\dim(\mathbb{H}) = \dim(\mathbb{H}^*)$.

Proof.

Let $\{\mathrm{e}_1, \mathrm{e}_2, \cdots, \mathrm{e}_n\}$ be an orthonormal basis of $\mathbb{H}$ (one can always find an orthonormal basis through the Gram-Schmidt process). For each $1 \leqslant k \leqslant n$, define $\varphi_k : \mathbb{H} \to \mathbb{F}$ by

$$\varphi_k(\boldsymbol{x}) = \langle \mathrm{e}_k, \boldsymbol{x} \rangle .$$

The Cauchy-Schwarz inequality then implies that

$$\left|\varphi_k(\boldsymbol{x})\right| \leqslant \|\mathrm{e}_k\| \cdot \|\boldsymbol{x}\| = \|\boldsymbol{x}\| \qquad \forall\, \boldsymbol{x} \in \mathbb{H};$$

thus $\varphi_k \in \mathbb{H}^*$ for each $1 \leqslant k \leqslant n$. □

§3.4 Dual Spaces and Adjoint Operators

Theorem

If $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ *is a finite dimensional Hilbert space over field* $\mathbb{F}$*, then* $\mathbb{H}^*$ *is also finite dimensional and* $\dim(\mathbb{H}) = \dim(\mathbb{H}^*)$.

Proof.

Let $\{\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n\}$ be an orthonormal basis of $\mathbb{H}$ (one can always find an orthonormal basis through the Gram-Schmidt process). For each $1 \leqslant k \leqslant n$, define $\varphi_k : \mathbb{H} \to \mathbb{F}$ by

$$\varphi_k(\boldsymbol{x}) = \langle \mathbf{e}_k, \boldsymbol{x} \rangle .$$

The Cauchy-Schwarz inequality then implies that

$$\big|\varphi_k(\boldsymbol{x})\big| \leqslant \|\mathbf{e}_k\| \cdot \|\boldsymbol{x}\| = \|\boldsymbol{x}\| \qquad \forall\, \boldsymbol{x} \in \mathbb{H}\,;$$

thus $\varphi_k \in \mathbb{H}^*$ for each $1 \leqslant k \leqslant n$. □

§3.4 Dual Spaces and Adjoint Operators

Proof (cont.)

Moreover, if $\alpha_1, \cdots, \alpha_n$ are numbers in $\mathbb{F}$ and

$$\alpha_1\varphi_1(\boldsymbol{x}) + \alpha_2\varphi_2(\boldsymbol{x}) + \cdots + \alpha_n\varphi_n(\boldsymbol{x}) = 0 \qquad \forall\, \boldsymbol{x} \in \mathbb{H}\,,$$

then for each $1 \leqslant j \leqslant n$,

$$0 = \alpha_1\varphi_1(\mathbf{e}_j) + \alpha_2\varphi_2(\mathbf{e}_j) + \cdots + \alpha_n\varphi_n(\mathbf{e}_j) = \sum_{k=1}^{n} \alpha_k\delta_{jk} = \alpha_k\,.$$

Therefore, $\{\varphi_1, \varphi_2, \cdots, \varphi_n\}$ is a linearly independent set.

Finally, by the fact that $\boldsymbol{x} = \sum_{k=1}^{n} \langle \mathbf{e}_k, \boldsymbol{x}\rangle \mathbf{e}_k$ for all $\boldsymbol{x} \in \mathbb{H}$, we find that for $f \in \mathbb{H}^*$,

$$f(\boldsymbol{x}) = f\Big(\sum_{k=1}^{n} \underline{\underline{\langle \mathbf{e}_k, \boldsymbol{x}\rangle}}\, \mathbf{e}_k\Big) = \sum_{k=1}^{n} f(\mathbf{e}_k)\underline{\underline{\varphi_k(\boldsymbol{x})}} \qquad \forall\, \boldsymbol{x} \in \mathbb{H}\,.$$

This shows that $\{\varphi_1, \varphi_2, \cdots, \varphi_n\}$ is a basis of $\mathbb{H}^*$; thus $\dim(\mathbb{H}^*) = n$. □

§3.4 Dual Spaces and Adjoint Operators

Proof (cont.)

Moreover, if $\alpha_1, \cdots, \alpha_n$ are numbers in $\mathbb{F}$ and

$$\alpha_1\varphi_1(\boldsymbol{x}) + \alpha_2\varphi_2(\boldsymbol{x}) + \cdots + \alpha_n\varphi_n(\boldsymbol{x}) = 0 \qquad \forall\, \boldsymbol{x} \in \mathbb{H}\,,$$

then for each $1 \leqslant j \leqslant n$,

$$0 = \alpha_1\varphi_1(\mathbf{e}_j) + \alpha_2\varphi_2(\mathbf{e}_j) + \cdots + \alpha_n\varphi_n(\mathbf{e}_j) = \sum_{k=1}^{n} \alpha_k\delta_{jk} = \alpha_k\,.$$

Therefore, $\{\varphi_1, \varphi_2, \cdots, \varphi_n\}$ is a linearly independent set.

Finally, by the fact that $\boldsymbol{x} = \sum\limits_{k=1}^{n} \langle \mathbf{e}_k, \boldsymbol{x}\rangle \mathbf{e}_k$ for all $\boldsymbol{x} \in \mathbb{H}$, we find that for $f \in \mathbb{H}^*$,

$$f(\boldsymbol{x}) = f\Big(\sum_{k=1}^{n} \underline{\underline{\langle \mathbf{e}_k, \boldsymbol{x}\rangle}}\, \mathbf{e}_k\Big) = \sum_{k=1}^{n} f(\mathbf{e}_k)\underline{\underline{\varphi_k(\boldsymbol{x})}} \qquad \forall\, \boldsymbol{x} \in \mathbb{H}\,.$$

This shows that $\{\varphi_1, \varphi_2, \cdots, \varphi_n\}$ is a basis of $\mathbb{H}^*$; thus $\dim(\mathbb{H}^*) = n$. □

§3.4 Dual Spaces and Adjoint Operators

Theorem (Riesz Representation)

*Let $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ be a Hilbert space. Then every $L \in \mathbb{H}^*$ corresponds to a **unique** $\boldsymbol{y} \in \mathbb{H}$ such that $L(\boldsymbol{x}) = \langle \boldsymbol{y}, \boldsymbol{x}\rangle$ for all $\boldsymbol{x} \in \mathbb{H}$. In other words, there exists a bijection $\varphi : \mathbb{H}^* \to \mathbb{H}$ such that*

$$L(\boldsymbol{x}) = \langle \varphi(L), \boldsymbol{x}\rangle \qquad \forall\, \boldsymbol{x} \in \mathbb{H}\,.$$

Moreover, $\|\varphi(L)\| = \|L\|_{\mathscr{B}(\mathbb{H},\mathbb{F})}$ for all $L \in \mathbb{H}^*$.

Proof.

W.L.O.G., we assume that L is not the zero map.

Let N be the null space of L; that is, $N = L^{-1}(\{0\})$. Then $N^\perp$, the orthogonal complement of N, has a non-zero element $\boldsymbol{z}$ with $\|\boldsymbol{z}\| = 1$ (**details required**). Such $\boldsymbol{z}$ verifies the identity that

$$L\big(L(\boldsymbol{x})\boldsymbol{z} - L(\boldsymbol{z})\boldsymbol{x}\big) = L(\boldsymbol{x})L(\boldsymbol{z}) - L(\boldsymbol{z})L(\boldsymbol{x}) = 0 \qquad \forall\, \boldsymbol{x} \in \mathbb{H}\,. \qquad \square$$

§3.4 Dual Spaces and Adjoint Operators

Theorem (Riesz Representation)

Let $(\mathbb{H}, \langle \cdot, \cdot \rangle)$ be a Hilbert space. Then every $L \in \mathbb{H}^*$ corresponds to a ***unique*** *$\boldsymbol{y} \in \mathbb{H}$ such that $L(\boldsymbol{x}) = \langle \boldsymbol{y}, \boldsymbol{x} \rangle$ for all $\boldsymbol{x} \in \mathbb{H}$. In other words, there exists a bijection $\varphi : \mathbb{H}^* \to \mathbb{H}$ such that*

$$L(\boldsymbol{x}) = \langle \varphi(L), \boldsymbol{x} \rangle \qquad \forall\, \boldsymbol{x} \in \mathbb{H}\,.$$

Moreover, $\|\varphi(L)\| = \|L\|_{\mathcal{B}(\mathbb{H},\mathbb{F})}$ for all $L \in \mathbb{H}^*$.

Proof.

W.L.O.G., we assume that L is not the zero map.

Let N be the null space of L; that is, $N = L^{-1}(\{0\})$. Then $N^{\perp}$, the orthogonal complement of N, has a non-zero element $\boldsymbol{z}$ with $\|\boldsymbol{z}\| = 1$ (**details required**). Such $\boldsymbol{z}$ verifies the identity that

$$L\big(L(\boldsymbol{x})\boldsymbol{z} - L(\boldsymbol{z})\boldsymbol{x}\big) = L(\boldsymbol{x})L(\boldsymbol{z}) - L(\boldsymbol{z})L(\boldsymbol{x}) = 0 \qquad \forall\, \boldsymbol{x} \in \mathbb{H}\,. \qquad \square$$

§3.4 Dual Spaces and Adjoint Operators

Theorem (Riesz Representation)

Let $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ be a Hilbert space. Then every $L \in \mathbb{H}^*$ corresponds to a ***unique*** *$\boldsymbol{y} \in \mathbb{H}$ such that $L(\boldsymbol{x}) = \langle \boldsymbol{y}, \boldsymbol{x}\rangle$ for all $\boldsymbol{x} \in \mathbb{H}$. In other words, there exists a bijection $\varphi : \mathbb{H}^* \to \mathbb{H}$ such that*

$$L(\boldsymbol{x}) = \langle \varphi(L), \boldsymbol{x}\rangle \qquad \forall\, \boldsymbol{x} \in \mathbb{H}\,.$$

Moreover, $\|\varphi(L)\| = \|L\|_{\mathscr{B}(\mathbb{H},\mathbb{F})}$ for all $L \in \mathbb{H}^*$.

Proof.

W.L.O.G., we assume that L is not the zero map.

Let N be the null space of L; that is, $N = L^{-1}(\{0\})$. Then $N^{\perp}$, the orthogonal complement of N, has a non-zero element $\boldsymbol{z}$ with $\|\boldsymbol{z}\| = 1$ (**details required**). Such z verifies the identity that

$$L\big(L(\boldsymbol{x})\boldsymbol{z} - L(\boldsymbol{z})\boldsymbol{x}\big) = L(\boldsymbol{x})L(\boldsymbol{z}) - L(\boldsymbol{z})L(\boldsymbol{x}) = 0 \qquad \forall\, \boldsymbol{x} \in \mathbb{H}\,. \qquad \square$$

§3.4 Dual Spaces and Adjoint Operators

Proof (cont.)

Therefore, the vector $L(\boldsymbol{x})\boldsymbol{z} - L(\boldsymbol{z})\boldsymbol{x} \in N$ for all $\boldsymbol{x} \in \mathbb{H}$; thus for each $\boldsymbol{x} \in \mathbb{H}$,

$$\begin{aligned} 0 = \langle \boldsymbol{z}, L(\boldsymbol{x})\boldsymbol{z} - L(\boldsymbol{z})\boldsymbol{x} \rangle &= L(\boldsymbol{x})\|\boldsymbol{z}\|^2 - L(\boldsymbol{z})\langle \boldsymbol{z}, \boldsymbol{x} \rangle \\ &= L(\boldsymbol{x}) - L(\boldsymbol{z})\langle \boldsymbol{z}, \boldsymbol{x} \rangle \end{aligned}$$

so that letting $\boldsymbol{y} = \overline{L(\boldsymbol{z})}\boldsymbol{z}$, we have

$$L(\boldsymbol{x}) = \langle \boldsymbol{y}, \boldsymbol{x} \rangle \qquad \forall\, \boldsymbol{x} \in \mathbb{H}\,.$$

Suppose that $\boldsymbol{y}_1, \boldsymbol{y}_2 \in \mathbb{H}$ satisfy $L(\boldsymbol{x}) = \langle \boldsymbol{y}_1, \boldsymbol{x} \rangle = \langle \boldsymbol{y}_2, \boldsymbol{x} \rangle$ for all $\boldsymbol{x} \in \mathbb{H}$. Then

$$\langle \boldsymbol{y}_1 - \boldsymbol{y}_2, \boldsymbol{x} \rangle = 0 \qquad \forall\, \boldsymbol{x} \in \mathbb{H}\,.$$

In particular, letting $\boldsymbol{x} = \boldsymbol{y}_1 - \boldsymbol{y}_2$ shows that $\|\boldsymbol{y}_1 - \boldsymbol{y}_2\| = 0$; thus $\boldsymbol{y}_1 = \boldsymbol{y}_2$. Therefore, each $L \in \mathbb{H}^*$ corresponds to a unique $\boldsymbol{y} \in \mathbb{H}$ satisfying $L(\boldsymbol{x}) = \langle \boldsymbol{y}, \boldsymbol{x} \rangle$ for all $\boldsymbol{x} \in \mathbb{H}$. □

§3.4 Dual Spaces and Adjoint Operators

Proof (cont.)

Therefore, the vector $L(\boldsymbol{x})\boldsymbol{z} - L(\boldsymbol{z})\boldsymbol{x} \in N$ for all $\boldsymbol{x} \in \mathbb{H}$; thus for each $\boldsymbol{x} \in \mathbb{H}$,

$$\begin{aligned} 0 = \langle \boldsymbol{z}, L(\boldsymbol{x})\boldsymbol{z} - L(\boldsymbol{z})\boldsymbol{x} \rangle &= L(\boldsymbol{x})\|\boldsymbol{z}\|^2 - L(\boldsymbol{z})\langle \boldsymbol{z}, \boldsymbol{x} \rangle \\ &= L(\boldsymbol{x}) - L(\boldsymbol{z})\langle \boldsymbol{z}, \boldsymbol{x} \rangle \end{aligned}$$

so that letting $\boldsymbol{y} = \overline{L(\boldsymbol{z})}\boldsymbol{z}$, we have

$$L(\boldsymbol{x}) = \langle \boldsymbol{y}, \boldsymbol{x} \rangle \qquad \forall\, \boldsymbol{x} \in \mathbb{H}\,.$$

Suppose that $\boldsymbol{y}_1, \boldsymbol{y}_2 \in \mathbb{H}$ satisfy $L(\boldsymbol{x}) = \langle \boldsymbol{y}_1, \boldsymbol{x} \rangle = \langle \boldsymbol{y}_2, \boldsymbol{x} \rangle$ for all $\boldsymbol{x} \in \mathbb{H}$. Then

$$\langle \boldsymbol{y}_1 - \boldsymbol{y}_2, \boldsymbol{x} \rangle = 0 \qquad \forall\, \boldsymbol{x} \in \mathbb{H}\,.$$

In particular, letting $\boldsymbol{x} = \boldsymbol{y}_1 - \boldsymbol{y}_2$ shows that $\|\boldsymbol{y}_1 - \boldsymbol{y}_2\| = 0$; thus $\boldsymbol{y}_1 = \boldsymbol{y}_2$. Therefore, each $L \in \mathbb{H}^*$ corresponds to a unique $\boldsymbol{y} \in \mathbb{H}$ satisfying $L(\boldsymbol{x}) = \langle \boldsymbol{y}, \boldsymbol{x} \rangle$ for all $\boldsymbol{x} \in \mathbb{H}$. □

§3.4 Dual Spaces and Adjoint Operators

Proof (cont.)

Therefore, the vector $L(\boldsymbol{x})\boldsymbol{z} - L(\boldsymbol{z})\boldsymbol{x} \in N$ for all $\boldsymbol{x} \in \mathbb{H}$; thus for each $\boldsymbol{x} \in \mathbb{H}$,

$$\begin{aligned} 0 = \langle \boldsymbol{z}, L(\boldsymbol{x})\boldsymbol{z} - L(\boldsymbol{z})\boldsymbol{x} \rangle &= L(\boldsymbol{x})\|\boldsymbol{z}\|^2 - L(\boldsymbol{z})\langle \boldsymbol{z}, \boldsymbol{x} \rangle \\ &= L(\boldsymbol{x}) - L(\boldsymbol{z})\langle \boldsymbol{z}, \boldsymbol{x} \rangle \end{aligned}$$

so that letting $\boldsymbol{y} = \overline{L(\boldsymbol{z})}\boldsymbol{z}$, we have

$$L(\boldsymbol{x}) = \langle \boldsymbol{y}, \boldsymbol{x} \rangle \qquad \forall\, \boldsymbol{x} \in \mathbb{H}\,.$$

Suppose that $\boldsymbol{y}_1, \boldsymbol{y}_2 \in \mathbb{H}$ satisfy $L(\boldsymbol{x}) = \langle \boldsymbol{y}_1, \boldsymbol{x} \rangle = \langle \boldsymbol{y}_2, \boldsymbol{x} \rangle$ for all $\boldsymbol{x} \in \mathbb{H}$. Then

$$\langle \boldsymbol{y}_1 - \boldsymbol{y}_2, \boldsymbol{x} \rangle = 0 \qquad \forall\, \boldsymbol{x} \in \mathbb{H}\,.$$

In particular, letting $\boldsymbol{x} = \boldsymbol{y}_1 - \boldsymbol{y}_2$ shows that $\|\boldsymbol{y}_1 - \boldsymbol{y}_2\| = 0$; thus $\boldsymbol{y}_1 = \boldsymbol{y}_2$. Therefore, each $L \in \mathbb{H}^*$ corresponds to a unique $\boldsymbol{y} \in \mathbb{H}$ satisfying $L(\boldsymbol{x}) = \langle \boldsymbol{y}, \boldsymbol{x} \rangle$ for all $\boldsymbol{x} \in \mathbb{H}$. □

§3.4 Dual Spaces and Adjoint Operators

Proof (cont.)

Therefore, the vector $L(\boldsymbol{x})\boldsymbol{z} - L(\boldsymbol{z})\boldsymbol{x} \in N$ for all $\boldsymbol{x} \in \mathbb{H}$; thus for each $\boldsymbol{x} \in \mathbb{H}$,

$$\begin{aligned} 0 = \langle \boldsymbol{z}, L(\boldsymbol{x})\boldsymbol{z} - L(\boldsymbol{z})\boldsymbol{x} \rangle &= L(\boldsymbol{x})\|\boldsymbol{z}\|^2 - L(\boldsymbol{z})\langle \boldsymbol{z}, \boldsymbol{x} \rangle \\ &= L(\boldsymbol{x}) - L(\boldsymbol{z})\langle \boldsymbol{z}, \boldsymbol{x} \rangle \end{aligned}$$

so that letting $\boldsymbol{y} = \overline{L(\boldsymbol{z})}\boldsymbol{z}$, we have

$$L(\boldsymbol{x}) = \langle \boldsymbol{y}, \boldsymbol{x} \rangle \qquad \forall\, \boldsymbol{x} \in \mathbb{H}\,.$$

Suppose that $\boldsymbol{y}_1, \boldsymbol{y}_2 \in \mathbb{H}$ satisfy $L(\boldsymbol{x}) = \langle \boldsymbol{y}_1, \boldsymbol{x} \rangle = \langle \boldsymbol{y}_2, \boldsymbol{x} \rangle$ for all $\boldsymbol{x} \in \mathbb{H}$. Then

$$\langle \boldsymbol{y}_1 - \boldsymbol{y}_2, \boldsymbol{x} \rangle = 0 \qquad \forall\, \boldsymbol{x} \in \mathbb{H}\,.$$

In particular, letting $\boldsymbol{x} = \boldsymbol{y}_1 - \boldsymbol{y}_2$ shows that $\|\boldsymbol{y}_1 - \boldsymbol{y}_2\| = 0$; thus $\boldsymbol{y}_1 = \boldsymbol{y}_2$. Therefore, each $L \in \mathbb{H}^*$ corresponds to a unique $\boldsymbol{y} \in \mathbb{H}$ satisfying $L(\boldsymbol{x}) = \langle \boldsymbol{y}, \boldsymbol{x} \rangle$ for all $\boldsymbol{x} \in \mathbb{H}$. □

§3.4 Dual Spaces and Adjoint Operators

Proof (cont.)

Finally, let $\varphi : \mathbb{H}^* \to \mathbb{H}$ denote the map satisfying

$$L(\boldsymbol{x}) = \langle \varphi(L), \boldsymbol{x} \rangle \qquad \forall\, \boldsymbol{x} \in \mathbb{H}\,.$$

Using the identity that $\|\boldsymbol{y}\| = \sup\limits_{\|\boldsymbol{x}\|=1} \big|\langle \boldsymbol{y}, \boldsymbol{x} \rangle\big|$ for all $\boldsymbol{y} \in \mathbb{H}$,

$$\|\varphi(L)\| = \sup_{\|\boldsymbol{x}\|=1} \big|\langle \varphi(L), \boldsymbol{x} \rangle\big| = \sup_{\|\boldsymbol{x}\|=1} \big|L(\boldsymbol{x})\big| = \|L\|_{\mathscr{B}(\mathbb{H},\mathbb{F})}\,. \qquad \square$$

Remark: Let $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ be a Hilbert space, and φ be the map given in the Riesz Representation Theorem. Define

$$\langle L_1, L_2 \rangle_{\mathbb{H}^*} = \langle \varphi(L_1), \varphi(L_2) \rangle \qquad \forall\, L_1, L_2 \in \mathbb{H}^*\,.$$

Then $(\mathbb{H}^*, \langle\cdot,\cdot\rangle_{\mathbb{H}^*})$ is a Hilbert space, and $\|\cdot\|_{\mathscr{B}(\mathbb{H},\mathbb{F})}$ is the norm induced by the inner product above. The operator norm $\|\cdot\|_{\mathscr{B}(\mathbb{H},\mathbb{F})}$ sometimes is denoted by $\|\cdot\|_{H^*}$.

§3.4 Dual Spaces and Adjoint Operators

Proof (cont.)

Finally, let $\varphi : \mathbb{H}^* \to \mathbb{H}$ denote the map satisfying

$$L(\boldsymbol{x}) = \langle \varphi(L), \boldsymbol{x} \rangle \qquad \forall\, \boldsymbol{x} \in \mathbb{H}\,.$$

Using the identity that $\|\boldsymbol{y}\| = \sup\limits_{\|\boldsymbol{x}\|=1} \big|\langle \boldsymbol{y}, \boldsymbol{x} \rangle\big|$ for all $\boldsymbol{y} \in \mathbb{H}$,

$$\|\varphi(L)\| = \sup_{\|\boldsymbol{x}\|=1} \big|\langle \varphi(L), \boldsymbol{x} \rangle\big| = \sup_{\|\boldsymbol{x}\|=1} \big|L(\boldsymbol{x})\big| = \|L\|_{\mathscr{B}(\mathbb{H},\mathbb{F})}\,. \qquad \square$$

Remark: Let $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ be a Hilbert space, and φ be the map given in the Riesz Representation Theorem. Define

$$\langle L_1, L_2 \rangle_{\mathbb{H}^*} = \big\langle \varphi(L_1), \varphi(L_2) \big\rangle \qquad \forall\, L_1, L_2 \in \mathbb{H}^*\,.$$

Then $(\mathbb{H}^*, \langle\cdot,\cdot\rangle_{\mathbb{H}^*})$ is a Hilbert space, and $\|\cdot\|_{\mathscr{B}(\mathbb{H},\mathbb{F})}$ is the norm induced by the inner product above. The operator norm $\|\cdot\|_{\mathscr{B}(\mathbb{H},\mathbb{F})}$ sometimes is denoted by $\|\cdot\|_{H^*}$.

§3.4 Dual Spaces and Adjoint Operators

Let $(\mathbb{V}, \langle\cdot,\cdot\rangle_{\mathbb{V}})$ and $(\mathbb{W}, \langle\cdot,\cdot\rangle_{\mathbb{W}})$ be Hilbert spaces over a common scalar field $\mathbb{F}$, where $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, and $A \in \mathscr{B}(\mathbb{V}, \mathbb{W})$. Note that the boundedness of A implies that

$$\|A\boldsymbol{v}\|_{\mathbb{W}} \leqslant \|A\|_{\mathscr{B}(\mathbb{V},\mathbb{W})}\|\boldsymbol{v}\|_{\mathbb{V}} < \infty \qquad \forall\, \boldsymbol{v} \in \mathbb{V}\,.$$

For a given $\boldsymbol{w} \in \mathbb{W}$, define $L : \mathbb{V} \to \mathbb{F}$ by

$$L(\boldsymbol{v}) = \langle \boldsymbol{w}, A\boldsymbol{v}\rangle_{\mathbb{W}}\,.$$

Then $L \in \mathbb{V}'$ (the algebraic dual space of $\mathbb{W}$) and the Cauchy-Schwarz inequality implies that

$$\big|L(\boldsymbol{v})\big| \leqslant \|\boldsymbol{w}\|_{\mathbb{W}}\|A\boldsymbol{v}\|_{\mathbb{W}} \leqslant \|\boldsymbol{w}\|_{\mathbb{W}}\|A\|_{\mathscr{B}(\mathbb{V},\mathbb{W})}\|\boldsymbol{v}\|_{\mathbb{V}}$$

so that

$$\sup_{\|\boldsymbol{v}\|_{\mathbb{V}}=1} \big|L(\boldsymbol{v})\big| \leqslant \|A\|_{\mathscr{B}(\mathbb{V},\mathbb{W})}\|\boldsymbol{w}\|_{\mathbb{W}} < \infty\,.$$

Therefore, $L \in \mathbb{V}^*$ (the continuous dual space of $\mathbb{W}$).

§3.4 Dual Spaces and Adjoint Operators

Let $(\mathbb{V}, \langle\cdot,\cdot\rangle_{\mathbb{V}})$ and $(\mathbb{W}, \langle\cdot,\cdot\rangle_{\mathbb{W}})$ be Hilbert spaces over a common scalar field $\mathbb{F}$, where $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, and $A \in \mathcal{B}(\mathbb{V}, \mathbb{W})$. Note that the boundedness of A implies that

$$\|A\boldsymbol{v}\|_{\mathbb{W}} \leqslant \|A\|_{\mathcal{B}(\mathbb{V},\mathbb{W})} \|\boldsymbol{v}\|_{\mathbb{V}} < \infty \qquad \forall\, \boldsymbol{v} \in \mathbb{V}\,.$$

For a given $\boldsymbol{w} \in \mathbb{W}$, define $L : \mathbb{V} \to \mathbb{F}$ by

$$L(\boldsymbol{v}) = \langle \boldsymbol{w}, A\boldsymbol{v}\rangle_{\mathbb{W}}\,.$$

Then $L \in \mathbb{V}'$ (the algebraic dual space of $\mathbb{W}$) and the Cauchy-Schwarz inequality implies that

$$\big|L(\boldsymbol{v})\big| \leqslant \|\boldsymbol{w}\|_{\mathbb{W}} \|A\boldsymbol{v}\|_{\mathbb{W}} \leqslant \|\boldsymbol{w}\|_{\mathbb{W}} \|A\|_{\mathcal{B}(\mathbb{V},\mathbb{W})} \|\boldsymbol{v}\|_{\mathbb{V}}$$

so that

$$\sup_{\|\boldsymbol{v}\|_{\mathbb{V}}=1} \big|L(\boldsymbol{v})\big| \leqslant \|A\|_{\mathcal{B}(\mathbb{V},\mathbb{W})} \|\boldsymbol{w}\|_{\mathbb{W}} < \infty\,.$$

Therefore, $L \in \mathbb{V}^*$ (the continuous dual space of $\mathbb{W}$).

§3.4 Dual Spaces and Adjoint Operators

Let $(\mathbb{V}, \langle\cdot,\cdot\rangle_{\mathbb{V}})$ and $(\mathbb{W}, \langle\cdot,\cdot\rangle_{\mathbb{W}})$ be Hilbert spaces over a common scalar field $\mathbb{F}$, where $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, and $A \in \mathcal{B}(\mathbb{V}, \mathbb{W})$. Note that the boundedness of A implies that

$$\|A\boldsymbol{v}\|_{\mathbb{W}} \leqslant \|A\|_{\mathcal{B}(\mathbb{V},\mathbb{W})}\|\boldsymbol{v}\|_{\mathbb{V}} < \infty \qquad \forall\, \boldsymbol{v} \in \mathbb{V}\,.$$

For a given $\boldsymbol{w} \in \mathbb{W}$, define $L : \mathbb{V} \to \mathbb{F}$ by

$$L(\boldsymbol{v}) = \langle \boldsymbol{w}, A\boldsymbol{v}\rangle_{\mathbb{W}}\,.$$

Then $L \in \mathbb{V}'$ (the algebraic dual space of $\mathbb{W}$) and the Cauchy-Schwarz inequality implies that

$$\big|L(\boldsymbol{v})\big| \leqslant \|\boldsymbol{w}\|_{\mathbb{W}}\|A\boldsymbol{v}\|_{\mathbb{W}} \leqslant \|\boldsymbol{w}\|_{\mathbb{W}}\|A\|_{\mathcal{B}(\mathbb{V},\mathbb{W})}\|\boldsymbol{v}\|_{\mathbb{V}}$$

so that

$$\sup_{\|\boldsymbol{v}\|_{\mathbb{V}}=1}\big|L(\boldsymbol{v})\big| \leqslant \|A\|_{\mathcal{B}(\mathbb{V},\mathbb{W})}\|\boldsymbol{w}\|_{\mathbb{W}} < \infty\,.$$

Therefore, $L \in \mathbb{V}^*$ (the continuous dual space of $\mathbb{W}$).

§3.4 Dual Spaces and Adjoint Operators

By the Riesz representation theorem, there exists a unique vector $\boldsymbol{u} \in \mathbb{V}$ such that

$$L(\boldsymbol{v}) = \langle \boldsymbol{u}, \boldsymbol{v} \rangle_{\mathbb{V}} \qquad \forall\, \boldsymbol{v} \in \mathbb{V}\,.$$

The map $\boldsymbol{w} \mapsto \boldsymbol{u}$ is denoted by A^* (so $A^* : \mathbb{W} \to \mathbb{V}$), and A^* is called the ***adjoint operator*** of A.

We note that A^* satisfies that

$$\langle \boldsymbol{w}, A\boldsymbol{v} \rangle_{\mathbb{W}} = \langle A^*(\boldsymbol{w}), \boldsymbol{v} \rangle_{\mathbb{V}} \qquad \forall\, \boldsymbol{v} \in \mathbb{V}\,, \boldsymbol{w} \in \mathbb{W}$$

so that for all $\boldsymbol{v} \in \mathbb{V}$ and $\boldsymbol{w}_1, \boldsymbol{w}_2$ in $\mathbb{W}$,

$$\begin{aligned}\Big\langle A^*(\lambda \boldsymbol{w}_1 + \mu \boldsymbol{w}_2), \boldsymbol{v} \Big\rangle_{\mathbb{V}} &= \langle \lambda \boldsymbol{w}_1 + \mu \boldsymbol{w}_2, A\boldsymbol{v} \rangle_{\mathbb{W}} = \bar{\lambda} \langle \boldsymbol{w}_1, A\boldsymbol{v} \rangle_{\mathbb{W}} + \bar{\mu} \langle \boldsymbol{w}_2, A\boldsymbol{v} \rangle_{\mathbb{W}} \\ &= \bar{\lambda} \Big\langle A^*(\boldsymbol{w}_1), \boldsymbol{v} \Big\rangle_{\mathbb{V}} + \bar{\mu} \Big\langle A^*(\boldsymbol{w}_2), \boldsymbol{v} \Big\rangle_{\mathbb{V}} \\ &= \Big\langle \lambda A^*(\boldsymbol{w}_1) + \mu A^*(\boldsymbol{w}_2), \boldsymbol{v} \Big\rangle_{\mathbb{V}}\,.\end{aligned}$$

§3.4 Dual Spaces and Adjoint Operators

By the Riesz representation theorem, there exists a unique vector $\boldsymbol{u} \in \mathbb{V}$ such that

$$L(\boldsymbol{v}) = \langle \boldsymbol{u}, \boldsymbol{v} \rangle_{\mathbb{V}} \qquad \forall\, \boldsymbol{v} \in \mathbb{V}\,.$$

The map $\boldsymbol{w} \mapsto \boldsymbol{u}$ is denoted by A^* (so $A^* : \mathbb{W} \to \mathbb{V}$), and A^* is called the ***adjoint operator*** of A.

We note that A^* satisfies that

$$\langle \boldsymbol{w}, A\boldsymbol{v} \rangle_{\mathbb{W}} = \langle A^*(\boldsymbol{w}), \boldsymbol{v} \rangle_{\mathbb{V}} \qquad \forall\, \boldsymbol{v} \in \mathbb{V}\,, \boldsymbol{w} \in \mathbb{W}$$

so that for all $\boldsymbol{v} \in \mathbb{V}$ and $\boldsymbol{w}_1, \boldsymbol{w}_2$ in $\mathbb{W}$,

$$\begin{aligned}
&\Big\langle A^*(\lambda \boldsymbol{w}_1 + \mu \boldsymbol{w}_2), \boldsymbol{v} \Big\rangle_{\mathbb{V}} \\
&\quad = \langle \lambda \boldsymbol{w}_1 + \mu \boldsymbol{w}_2, A\boldsymbol{v} \rangle_{\mathbb{W}} = \bar{\lambda} \langle \boldsymbol{w}_1, A\boldsymbol{v} \rangle_{\mathbb{W}} + \bar{\mu} \langle \boldsymbol{w}_2, A\boldsymbol{v} \rangle_{\mathbb{W}} \\
&\quad = \bar{\lambda} \Big\langle A^*(\boldsymbol{w}_1), \boldsymbol{v} \Big\rangle_{\mathbb{V}} + \bar{\mu} \Big\langle A^*(\boldsymbol{w}_2), \boldsymbol{v} \Big\rangle_{\mathbb{V}} \\
&\quad = \Big\langle \lambda A^*(\boldsymbol{w}_1) + \mu A^*(\boldsymbol{w}_2), \boldsymbol{v} \Big\rangle_{\mathbb{V}}\,.
\end{aligned}$$

§3.4 Dual Spaces and Adjoint Operators

Therefore,

$$A^*(\lambda \boldsymbol{w}_1 + \mu \boldsymbol{w}_2) = \lambda A^*(\boldsymbol{w}_1) + \mu A^*(\boldsymbol{w}_2) \qquad \forall\, \boldsymbol{w}_1, \boldsymbol{w}_2 \in \mathbb{W}\,;$$

thus $A^* \in \mathcal{L}(\mathbb{W}, \mathbb{V})$. Moreover,

$$\begin{aligned} \|A^*\|_{\mathcal{B}(\mathbb{W},\mathbb{V})} &= \sup_{\|\boldsymbol{w}\|_{\mathbb{W}}=1} \sup_{\|\boldsymbol{v}\|_{\mathbb{V}}=1} \big|\langle A^*\boldsymbol{w}, \boldsymbol{v}\rangle_{\mathbb{V}}\big| = \sup_{\|\boldsymbol{w}\|_{\mathbb{W}}=1} \sup_{\|\boldsymbol{v}\|_{\mathbb{V}}=1} \big|\langle \boldsymbol{w}, A\boldsymbol{v}\rangle_{\mathbb{W}}\big| \\ &= \sup_{\|\boldsymbol{v}\|_{\mathbb{V}}=1} \sup_{\|\boldsymbol{w}\|_{\mathbb{W}}=1} \big|\langle \boldsymbol{w}, A\boldsymbol{v}\rangle_{\mathbb{W}}\big| = \|A\|_{\mathcal{B}(\mathbb{V},\mathbb{W})} \end{aligned}$$

thus A^* is indeed bounded.

Definition

Let $(\mathbb{V}, \langle\cdot,\cdot\rangle_{\mathbb{V}})$ and $(\mathbb{W}, \langle\cdot,\cdot\rangle_{\mathbb{W}})$ be Hilbert spaces over field $\mathbb{F}$, where $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, and $A \in \mathcal{B}(\mathbb{V}, \mathbb{W})$. The adjoint operator of A, denoted by A^*, is the unique element in $\mathcal{B}(\mathbb{W}, \mathbb{V})$ satisfying that

$$\langle \boldsymbol{w}, A\boldsymbol{v}\rangle_{\mathbb{W}} = \langle A^*\boldsymbol{w}, \boldsymbol{v}\rangle_{\mathbb{V}} \qquad \forall\, \boldsymbol{v} \in \mathbb{V}\,, \boldsymbol{w} \in \mathbb{W}\,.$$

§3.4 Dual Spaces and Adjoint Operators

Therefore,

$$A^*(\lambda \boldsymbol{w}_1 + \mu \boldsymbol{w}_2) = \lambda A^*(\boldsymbol{w}_1) + \mu A^*(\boldsymbol{w}_2) \qquad \forall\, \boldsymbol{w}_1, \boldsymbol{w}_2 \in \mathbb{W}\,;$$

thus $A^* \in \mathcal{L}(\mathbb{W}, \mathbb{V})$. Moreover,

$$\begin{aligned}\|A^*\|_{\mathcal{B}(\mathbb{W},\mathbb{V})} &= \sup_{\|\boldsymbol{w}\|_{\mathbb{W}}=1} \sup_{\|\boldsymbol{v}\|_{\mathbb{V}}=1} \big|\langle A^*\boldsymbol{w}, \boldsymbol{v}\rangle_{\mathbb{V}}\big| = \sup_{\|\boldsymbol{w}\|_{\mathbb{W}}=1} \sup_{\|\boldsymbol{v}\|_{\mathbb{V}}=1} \big|\langle \boldsymbol{w}, A\boldsymbol{v}\rangle_{\mathbb{W}}\big| \\ &= \sup_{\|\boldsymbol{v}\|_{\mathbb{V}}=1} \sup_{\|\boldsymbol{w}\|_{\mathbb{W}}=1} \big|\langle \boldsymbol{w}, A\boldsymbol{v}\rangle_{\mathbb{W}}\big| = \|A\|_{\mathcal{B}(\mathbb{V},\mathbb{W})}\end{aligned}$$

thus A^* is indeed bounded.

Definition

Let $(\mathbb{V}, \langle\cdot,\cdot\rangle_{\mathbb{V}})$ and $(\mathbb{W}, \langle\cdot,\cdot\rangle_{\mathbb{W}})$ be Hilbert spaces over field $\mathbb{F}$, where $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, and $A \in \mathcal{B}(\mathbb{V}, \mathbb{W})$. The adjoint operator of A, denoted by A^*, is the unique element in $\mathcal{B}(\mathbb{W}, \mathbb{V})$ satisfying that

$$\langle \boldsymbol{w}, A\boldsymbol{v}\rangle_{\mathbb{W}} = \langle A^*\boldsymbol{w}, \boldsymbol{v}\rangle_{\mathbb{V}} \qquad \forall\, \boldsymbol{v} \in \mathbb{V}\,, \boldsymbol{w} \in \mathbb{W}\,.$$

§3.4 Dual Spaces and Adjoint Operators

Therefore,

$$A^*(\lambda \boldsymbol{w}_1 + \mu \boldsymbol{w}_2) = \lambda A^*(\boldsymbol{w}_1) + \mu A^*(\boldsymbol{w}_2) \qquad \forall\, \boldsymbol{w}_1, \boldsymbol{w}_2 \in \mathbb{W}\,;$$

thus $A^* \in \mathscr{L}(\mathbb{W}, \mathbb{V})$. Moreover,

$$\begin{aligned}\|A^*\|_{\mathscr{B}(\mathbb{W},\mathbb{V})} &= \sup_{\|\boldsymbol{w}\|_{\mathbb{W}}=1} \sup_{\|\boldsymbol{v}\|_{\mathbb{V}}=1} \big|\langle A^*\boldsymbol{w}, \boldsymbol{v}\rangle_{\mathbb{V}}\big| = \sup_{\|\boldsymbol{w}\|_{\mathbb{W}}=1} \sup_{\|\boldsymbol{v}\|_{\mathbb{V}}=1} \big|\langle \boldsymbol{w}, A\boldsymbol{v}\rangle_{\mathbb{W}}\big| \\ &= \sup_{\|\boldsymbol{v}\|_{\mathbb{V}}=1} \sup_{\|\boldsymbol{w}\|_{\mathbb{W}}=1} \big|\langle \boldsymbol{w}, A\boldsymbol{v}\rangle_{\mathbb{W}}\big| = \|A\|_{\mathscr{B}(\mathbb{V},\mathbb{W})}\end{aligned}$$

thus A^* is indeed bounded.

Definition

Let $(\mathbb{V}, \langle\cdot,\cdot\rangle_{\mathbb{V}})$ and $(\mathbb{W}, \langle\cdot,\cdot\rangle_{\mathbb{W}})$ be Hilbert spaces over field $\mathbb{F}$, where $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, and $A \in \mathscr{B}(\mathbb{V}, \mathbb{W})$. The adjoint operator of A, denoted by A^*, is the unique element in $\mathscr{B}(\mathbb{W}, \mathbb{V})$ satisfying that

$$\langle \boldsymbol{w}, A\boldsymbol{v}\rangle_{\mathbb{W}} = \langle A^*\boldsymbol{w}, \boldsymbol{v}\rangle_{\mathbb{V}} \qquad \forall\, \boldsymbol{v} \in \mathbb{V}\,, \boldsymbol{w} \in \mathbb{W}\,.$$

§3.4 Dual Spaces and Adjoint Operators

Let $(\mathbb{V}, \langle\cdot,\cdot\rangle_{\mathbb{V}})$ and $(\mathbb{W}, \langle\cdot,\cdot\rangle_{\mathbb{W}})$ be Hilbert spaces over field $\mathbb{F}$, where $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, and $A \in \mathscr{B}(\mathbb{V}, \mathbb{W})$. By the property of inner product,

$$\langle A\boldsymbol{v}, \boldsymbol{w}\rangle_{\mathbb{W}} = \overline{\langle \boldsymbol{w}, A\boldsymbol{v}\rangle_{\mathbb{W}}} = \overline{\langle A^*\boldsymbol{w}, \boldsymbol{v}\rangle_{\mathbb{V}}} = \langle \boldsymbol{v}, A^*\boldsymbol{w}\rangle_{\mathbb{V}} \quad \forall\, \boldsymbol{v} \in \mathbb{V}, \boldsymbol{w} \in \mathbb{W}\,.$$

Therefore, the adjoint operator A^* of A satisfies

$$\langle \boldsymbol{w}, A\boldsymbol{v}\rangle_{\mathbb{W}} = \langle A^*\boldsymbol{w}, \boldsymbol{v}\rangle_{\mathbb{V}}\,, \langle A\boldsymbol{v}, \boldsymbol{w}\rangle_{\mathbb{W}} = \langle \boldsymbol{v}, A^*\boldsymbol{w}\rangle_{\mathbb{V}} \quad \forall\, \boldsymbol{v} \in \mathbb{V}, \boldsymbol{w} \in \mathbb{W}. \quad (2)$$

Proposition

Let $(\mathbb{V}, \langle\cdot,\cdot\rangle_{\mathbb{V}})$ and $(\mathbb{W}, \langle\cdot,\cdot\rangle_{\mathbb{W}})$ be Hilbert spaces over field $\mathbb{F}$, where $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, and $A \in \mathscr{B}(\mathbb{V}, \mathbb{W})$. Then $(A^*)^* = A$.

Proof.

Let $\boldsymbol{v} \in \mathbb{V}$ be given. Then if $\boldsymbol{w} \in \mathbb{W}$, using (2) we find that

$$\langle A\boldsymbol{v}, \boldsymbol{w}\rangle_{\mathbb{W}} = \langle \boldsymbol{v}, A^*\boldsymbol{w}\rangle_{\mathbb{V}} = \langle (A^*)^*\boldsymbol{v}, \boldsymbol{w}\rangle_{\mathbb{W}}\,.$$

Therefore, $A\boldsymbol{v} = (A^*)^*\boldsymbol{v}$ for all $\boldsymbol{v} \in \mathbb{H}$; thus $A = (A^*)^*$. □

§3.4 Dual Spaces and Adjoint Operators

Let $(\mathbb{V}, \langle\cdot,\cdot\rangle_{\mathbb{V}})$ and $(\mathbb{W}, \langle\cdot,\cdot\rangle_{\mathbb{W}})$ be Hilbert spaces over field $\mathbb{F}$, where $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, and $A \in \mathscr{B}(\mathbb{V}, \mathbb{W})$. By the property of inner product,

$$\langle A\boldsymbol{v}, \boldsymbol{w}\rangle_{\mathbb{W}} = \overline{\langle \boldsymbol{w}, A\boldsymbol{v}\rangle_{\mathbb{W}}} = \overline{\langle A^*\boldsymbol{w}, \boldsymbol{v}\rangle_{\mathbb{V}}} = \langle \boldsymbol{v}, A^*\boldsymbol{w}\rangle_{\mathbb{V}} \quad \forall\, \boldsymbol{v} \in \mathbb{V}, \boldsymbol{w} \in \mathbb{W}\,.$$

Therefore, the adjoint operator A^* of A satisfies

$$\langle \boldsymbol{w}, A\boldsymbol{v}\rangle_{\mathbb{W}} = \langle A^*\boldsymbol{w}, \boldsymbol{v}\rangle_{\mathbb{V}}\,, \langle A\boldsymbol{v}, \boldsymbol{w}\rangle_{\mathbb{W}} = \langle \boldsymbol{v}, A^*\boldsymbol{w}\rangle_{\mathbb{V}} \quad \forall\, \boldsymbol{v} \in \mathbb{V}, \boldsymbol{w} \in \mathbb{W}. \tag{2}$$

Proposition

Let $(\mathbb{V}, \langle\cdot,\cdot\rangle_{\mathbb{V}})$ and $(\mathbb{W}, \langle\cdot,\cdot\rangle_{\mathbb{W}})$ be Hilbert spaces over field $\mathbb{F}$, where $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, and $A \in \mathscr{B}(\mathbb{V}, \mathbb{W})$. Then $(A^*)^* = A$.

Proof.

Let $\boldsymbol{v} \in \mathbb{V}$ be given. Then if $\boldsymbol{w} \in \mathbb{W}$, using (2) we find that

$$\langle A\boldsymbol{v}, \boldsymbol{w}\rangle_{\mathbb{W}} = \langle \boldsymbol{v}, A^*\boldsymbol{w}\rangle_{\mathbb{V}} = \langle (A^*)^*\boldsymbol{v}, \boldsymbol{w}\rangle_{\mathbb{W}}\,.$$

Therefore, $A\boldsymbol{v} = (A^*)^*\boldsymbol{v}$ for all $\boldsymbol{v} \in \mathbb{H}$; thus $A = (A^*)^*$. □

§3.4 Dual Spaces and Adjoint Operators

Let $(\mathbb{V}, \langle\cdot,\cdot\rangle_{\mathbb{V}})$ and $(\mathbb{W}, \langle\cdot,\cdot\rangle_{\mathbb{W}})$ be Hilbert spaces over field $\mathbb{F}$, where $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, and $A \in \mathcal{B}(\mathbb{V}, \mathbb{W})$. By the property of inner product,

$$\langle A\boldsymbol{v}, \boldsymbol{w}\rangle_{\mathbb{W}} = \overline{\langle \boldsymbol{w}, A\boldsymbol{v}\rangle_{\mathbb{W}}} = \overline{\langle A^*\boldsymbol{w}, \boldsymbol{v}\rangle_{\mathbb{V}}} = \langle \boldsymbol{v}, A^*\boldsymbol{w}\rangle_{\mathbb{V}} \quad \forall\, \boldsymbol{v} \in \mathbb{V}, \boldsymbol{w} \in \mathbb{W}\,.$$

Therefore, the adjoint operator A^* of A satisfies

$$\langle \boldsymbol{w}, A\boldsymbol{v}\rangle_{\mathbb{W}} = \langle A^*\boldsymbol{w}, \boldsymbol{v}\rangle_{\mathbb{V}}\,, \langle A\boldsymbol{v}, \boldsymbol{w}\rangle_{\mathbb{W}} = \langle \boldsymbol{v}, A^*\boldsymbol{w}\rangle_{\mathbb{V}} \quad \forall\, \boldsymbol{v} \in \mathbb{V}, \boldsymbol{w} \in \mathbb{W}. \quad (2)$$

Proposition

Let $(\mathbb{V}, \langle\cdot,\cdot\rangle_{\mathbb{V}})$ and $(\mathbb{W}, \langle\cdot,\cdot\rangle_{\mathbb{W}})$ be Hilbert spaces over field $\mathbb{F}$, where $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, and $A \in \mathcal{B}(\mathbb{V}, \mathbb{W})$. Then $(A^*)^* = A$.

Proof.

Let $\boldsymbol{v} \in \mathbb{V}$ be given. Then if $\boldsymbol{w} \in \mathbb{W}$, using (2) we find that

$$\langle A\boldsymbol{v}, \boldsymbol{w}\rangle_{\mathbb{W}} = \langle \boldsymbol{v}, A^*\boldsymbol{w}\rangle_{\mathbb{V}} = \langle (A^*)^*\boldsymbol{v}, \boldsymbol{w}\rangle_{\mathbb{W}}\,.$$

Therefore, $A\boldsymbol{v} = (A^*)^*\boldsymbol{v}$ for all $\boldsymbol{v} \in \mathbb{H}$; thus $A = (A^*)^*$. □

§3.4 Dual Spaces and Adjoint Operators

Let $(\mathbb{V}, \langle\cdot,\cdot\rangle_{\mathbb{V}})$ and $(\mathbb{W}, \langle\cdot,\cdot\rangle_{\mathbb{W}})$ be finite dimensional inner product spaces over $\mathbb{C}$, $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ and $\widetilde{\mathcal{B}} = \{\mathbf{w}_1, \cdots, \mathbf{w}_m\}$ be orthonormal basis of $\mathbb{V}$ and $\mathbb{W}$, respectively, and $L \in \mathcal{L}(\mathbb{V};\mathbb{W})$. Then

the (i,j)-entry of $[L^*]_{\mathcal{B},\widetilde{\mathcal{B}}}$

$$= \big\langle [\mathbf{v}_i]_{\mathcal{B}}, [L^*]_{\mathcal{B},\widetilde{\mathcal{B}}}[\mathbf{w}_j]_{\widetilde{\mathcal{B}}} \big\rangle_{\mathbb{C}^n} = \langle \mathbf{v}_i, L^*\mathbf{w}_j \rangle_{\mathbb{V}} = \langle L\mathbf{v}_i, \mathbf{w}_j \rangle_{\mathbb{W}} = \overline{\langle \mathbf{w}_j, L\mathbf{v}_i \rangle_{\mathbb{W}}}$$

$=$ the complex conjugate of the (j,i)-entry of $[L]_{\widetilde{\mathcal{B}},\mathcal{B}}$.

This observation motivates the following

Definition (Conjugate transpose of matrices)

Let $A = [a_{ij}]_{m\times n}$ be an $m \times n$ complex matrix. The conjugate transpose of A, denoted by A^{H}, A^* or $A^\dagger$ (the last one is often used in quantum mechanics), is an $n \times m$ matrix $[b_{ij}]_{n\times m}$ given by $b_{ij} = \overline{a_{ji}}$.

§3.4 Dual Spaces and Adjoint Operators

Let $(\mathbb{V}, \langle\cdot,\cdot\rangle_{\mathbb{V}})$ and $(\mathbb{W}, \langle\cdot,\cdot\rangle_{\mathbb{W}})$ be finite dimensional inner product spaces over $\mathbb{C}$, $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ and $\widetilde{\mathcal{B}} = \{\mathbf{w}_1, \cdots, \mathbf{w}_m\}$ be orthonormal basis of $\mathbb{V}$ and $\mathbb{W}$, respectively, and $L \in \mathcal{L}(\mathbb{V};\mathbb{W})$. Then

the (i,j)-entry of $[L^*]_{\mathcal{B},\widetilde{\mathcal{B}}}$

$$= \Big\langle [\mathbf{v}_i]_{\mathcal{B}}, [L^*]_{\mathcal{B},\widetilde{\mathcal{B}}}[\mathbf{w}_j]_{\widetilde{\mathcal{B}}} \Big\rangle_{\mathbb{C}^n} = \langle \mathbf{v}_i, L^*\mathbf{w}_j\rangle_{\mathbb{V}} = \langle L\mathbf{v}_i, \mathbf{w}_j\rangle_{\mathbb{W}} = \overline{\langle \mathbf{w}_j, L\mathbf{v}_i\rangle_{\mathbb{W}}}$$

$=$ the complex conjugate of the (j,i)-entry of $[L]_{\widetilde{\mathcal{B}},\mathcal{B}}$.

This observation motivates the following

Definition (Conjugate transpose of matrices)

Let $A = [a_{ij}]_{m\times n}$ be an $m \times n$ complex matrix. The conjugate transpose of A, denoted by A^{H}, A^* or $A^\dagger$ (the last one is often used in quantum mechanics), is an $n \times m$ matrix $[b_{ij}]_{n\times m}$ given by $b_{ij} = \overline{a_{ji}}$.

§3.4 Dual Spaces and Adjoint Operators

Let $(\mathbb{V}, \langle\cdot,\cdot\rangle_{\mathbb{V}})$ and $(\mathbb{W}, \langle\cdot,\cdot\rangle_{\mathbb{W}})$ be finite dimensional inner product spaces over $\mathbb{C}$, $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ and $\widetilde{\mathcal{B}} = \{\mathbf{w}_1, \cdots, \mathbf{w}_m\}$ be orthonormal basis of $\mathbb{V}$ and $\mathbb{W}$, respectively, and $L \in \mathcal{L}(\mathbb{V};\mathbb{W})$. Then

$$\begin{aligned} &\text{the } (i,j)\text{-entry of } [L^*]_{\mathcal{B},\widetilde{\mathcal{B}}} \\ &= \Big\langle [\mathbf{v}_i]_{\mathcal{B}}, [L^*]_{\mathcal{B},\widetilde{\mathcal{B}}}[\mathbf{w}_j]_{\widetilde{\mathcal{B}}} \Big\rangle_{\mathbb{C}^n} = \langle \mathbf{v}_i, L^*\mathbf{w}_j\rangle_{\mathbb{V}} = \langle L\mathbf{v}_i, \mathbf{w}_j\rangle_{\mathbb{W}} = \overline{\langle \mathbf{w}_j, L\mathbf{v}_i\rangle_{\mathbb{W}}} \\ &= \text{the complex conjugate of the } (j,i)\text{-entry of } [L]_{\widetilde{\mathcal{B}},\mathcal{B}}. \end{aligned}$$

This observation motivates the following

Definition (Conjugate transpose of matrices)

Let $A = [a_{ij}]_{m\times n}$ be an $m \times n$ complex matrix. The conjugate transpose of A, denoted by A^{H}, A^* or $A^\dagger$ (the last one is often used in quantum mechanics), is an $n \times m$ matrix $[b_{ij}]_{n\times m}$ given by $b_{ij} = \overline{a_{ji}}$.

§3.4 Dual Spaces and Adjoint Operators

Remark: For real matrices, the conjugate transpose is just the transpose.

Theorem

Let $(\mathbb{V}, \langle\cdot,\cdot\rangle_{\mathbb{V}})$ and $(\mathbb{W}, \langle\cdot,\cdot\rangle_{\mathbb{W}})$ be finite dimensional inner product spaces over $\mathbb{C}$, $\mathcal{B}$ and $\widetilde{\mathcal{B}}$ be orthonormal basis of $\mathbb{V}$ and $\mathbb{W}$, respectively. If $L \in \mathscr{L}(\mathbb{V};\mathbb{W})$, then $[L^*]_{\mathcal{B},\widetilde{\mathcal{B}}} = [L]^{\dagger}_{\widetilde{\mathcal{B}},\mathcal{B}}$.

Definition

Let $A = [a_{ij}]$ be a square matrix.

1. A is said to be ***Hermitian*** if $A = A^\dagger$.
2. A is said to be ***skew Hermitian*** if $A^\dagger = -A$.
3. A is said to be ***normal*** if $AA^\dagger = A^\dagger A$.
4. A is said to be ***unitary*** if $A^{-1} = A^\dagger$ (explained in §3.5).

§3.4 Dual Spaces and Adjoint Operators

Remark: For real matrices, the conjugate transpose is just the transpose.

Theorem

Let $(\mathbb{V}, \langle\cdot,\cdot\rangle_{\mathbb{V}})$ *and* $(\mathbb{W}, \langle\cdot,\cdot\rangle_{\mathbb{W}})$ *be finite dimensional inner product spaces over* $\mathbb{C}$, $\mathcal{B}$ *and* $\widetilde{\mathcal{B}}$ *be orthonormal basis of* $\mathbb{V}$ *and* $\mathbb{W}$, *respectively. If* $L \in \mathscr{L}(\mathbb{V};\mathbb{W})$, *then* $[L^*]_{\mathcal{B},\widetilde{\mathcal{B}}} = [L]^{\dagger}_{\widetilde{\mathcal{B}},\mathcal{B}}$.

Definition

Let $A = [a_{ij}]$ be a square matrix.

1. A is said to be ***Hermitian*** if $A = A^\dagger$.
2. A is said to be ***skew Hermitian*** if $A^\dagger = -A$.
3. A is said to be ***normal*** if $AA^\dagger = A^\dagger A$.
4. A is said to be ***unitary*** if $A^{-1} = A^\dagger$ (explained in §3.5).

§3.4 Dual Spaces and Adjoint Operators

Remark: For real matrices, the conjugate transpose is just the transpose.

Theorem

Let $(\mathbb{V}, \langle\cdot,\cdot\rangle_{\mathbb{V}})$ *and* $(\mathbb{W}, \langle\cdot,\cdot\rangle_{\mathbb{W}})$ *be finite dimensional inner product spaces over* $\mathbb{C}$, $\mathcal{B}$ *and* $\widetilde{\mathcal{B}}$ *be orthonormal basis of* $\mathbb{V}$ *and* $\mathbb{W}$, *respectively. If* $L \in \mathcal{L}(\mathbb{V};\mathbb{W})$, *then* $[L^*]_{\mathcal{B},\widetilde{\mathcal{B}}} = [L]^{\dagger}_{\widetilde{\mathcal{B}},\mathcal{B}}$.

Definition

Let $A = [a_{ij}]$ be a square matrix.

1. A is said to be ***Hermitian*** if $A = A^{\dagger}$.
2. A is said to be ***skew Hermitian*** if $A^{\dagger} = -A$.
3. A is said to be ***normal*** if $AA^{\dagger} = A^{\dagger}A$.
4. A is said to be ***unitary*** if $A^{-1} = A^{\dagger}$ (explained in §3.5).

§3.5 Unitary Operators and Unitary Matrices

§3.5.1 Unitary operators

Definition

Let $(\mathbb{H}, \langle \cdot, \cdot \rangle)$ be a Hilbert space, and $U \in \mathcal{B}(\mathbb{H})$.

1. U is said to be **self-adjoint** if $U^* = U$.
2. U is said to be **unitary** if $UU^* = U^*U = \mathrm{Id}$, where Id denotes the identity map on $\mathbb{H}$.

The collection of self-adjoint operators on $\mathbb{H}$ is denoted by $\mathcal{B}_{\mathrm{sa}}(\mathbb{H})$, and the collection of unitary operators on $\mathbb{H}$ is denoted by $\mathrm{U}(\mathbb{H})$.

Remark: Let $(\mathbb{H}, \langle \cdot, \cdot \rangle)$ be a Hilbert space over $\mathbb{F}$, and $U \in \mathcal{B}(\mathbb{H})$.

1. If $\mathbb{F} = \mathbb{R}$, then U satisfying $UU^* = U^*U = \mathrm{Id}$ is often called **orthogonal** instead of unitary. Therefore, when the term "unitary" is used, we often assume that $\mathbb{F} = \mathbb{C}$.
2. If U is unitary, then U^* is also unitary.

§3.5 Unitary Operators and Unitary Matrices

§3.5.1 Unitary operators

Definition

Let $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ be a Hilbert space, and $U \in \mathscr{B}(\mathbb{H})$.

1. U is said to be **self-adjoint** if $U^* = U$.
2. U is said to be **unitary** if $UU^* = U^*U = \mathrm{Id}$, where Id denotes the identity map on $\mathbb{H}$.

The collection of self-adjoint operators on $\mathbb{H}$ is denoted by $\mathscr{B}_{\mathsf{sa}}(\mathbb{H})$, and the collection of unitary operators on $\mathbb{H}$ is denoted by $\mathrm{U}(\mathbb{H})$.

Remark: Let $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ be a Hilbert space over $\mathbb{F}$, and $U \in \mathscr{B}(\mathbb{H})$.

1. If $\mathbb{F} = \mathbb{R}$, then U satisfying $UU^* = U^*U = \mathrm{Id}$ is often called **orthogonal** instead of unitary. Therefore, when the term "unitary" is used, we often assume that $\mathbb{F} = \mathbb{C}$.
2. If U is unitary, then U^* is also unitary.

§3.5 Unitary Operators and Unitary Matrices

§3.5.1 Unitary operators

Definition

Let $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ be a Hilbert space, and $U \in \mathcal{B}(\mathbb{H})$.

1. U is said to be **self-adjoint** if $U^* = U$.
2. U is said to be **unitary** if $UU^* = U^*U = \mathrm{Id}$, where Id denotes the identity map on $\mathbb{H}$.

The collection of self-adjoint operators on $\mathbb{H}$ is denoted by $\mathcal{B}_{\mathsf{sa}}(\mathbb{H})$, and the collection of unitary operators on $\mathbb{H}$ is denoted by $\mathrm{U}(\mathbb{H})$.

Remark: Let $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ be a Hilbert space over $\mathbb{F}$, and $U \in \mathcal{B}(\mathbb{H})$.

1. If $\mathbb{F} = \mathbb{R}$, then U satisfying $UU^* = U^*U = \mathrm{Id}$ is often called **orthogonal** instead of unitary. Therefore, when the term "unitary" is used, we often assume that $\mathbb{F} = \mathbb{C}$.
2. If U is unitary, then U^* is also unitary.

§3.5 Unitary Operators and Unitary Matrices

Theorem

Let $(\mathbb{H},\langle\cdot,\cdot\rangle)$ *be a Hilbert space over field* $\mathbb{C}$*, and* $U\in\mathcal{B}(\mathbb{H})$*. The following three statements are equivalent.*

1. *U is unitary.*
2. *U is surjective and* $\|U\boldsymbol{x}\|=\|\boldsymbol{x}\|$ *for all* $\boldsymbol{x}\in\mathbb{H}$*.*
3. *U is surjective and* $\langle U\boldsymbol{x},U\boldsymbol{y}\rangle=\langle\boldsymbol{x},\boldsymbol{y}\rangle$ *for all* $\boldsymbol{x},\boldsymbol{y}\in\mathbb{H}$*.*

Proof.

① ⇒ ②: Let $\boldsymbol{z}\in\mathbb{H}$ be given. Then $\boldsymbol{y}=U^*\boldsymbol{z}\in\mathbb{H}$ satisfies $U\boldsymbol{y}=\boldsymbol{z}$. This implies that U is surjective. Moreover, if $\boldsymbol{x}\in\mathbb{H}$ is given, then

$$\|\boldsymbol{x}\|^2=\langle\boldsymbol{x},\boldsymbol{x}\rangle=\langle\boldsymbol{x},U^*U\boldsymbol{x}\rangle=\langle U\boldsymbol{x},U\boldsymbol{x}\rangle=\|U\boldsymbol{x}\|^2;$$

thus $\|U\boldsymbol{x}\|=\|\boldsymbol{x}\|$ for all $\boldsymbol{x}\in\mathbb{H}$. □

§3.5 Unitary Operators and Unitary Matrices

Theorem

Let $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ be a Hilbert space over field $\mathbb{C}$, and $U \in \mathscr{B}(\mathbb{H})$. The following three statements are equivalent.

1. *U is unitary.*
2. *U is surjective and $\|U\boldsymbol{x}\| = \|\boldsymbol{x}\|$ for all $\boldsymbol{x} \in \mathbb{H}$.*
3. *U is surjective and $\langle U\boldsymbol{x}, U\boldsymbol{y}\rangle = \langle \boldsymbol{x}, \boldsymbol{y}\rangle$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{H}$.*

Proof.

① ⇒ ②: Let $\boldsymbol{z} \in \mathbb{H}$ be given. Then $\boldsymbol{y} = U^*\boldsymbol{z} \in \mathbb{H}$ satisfies $U\boldsymbol{y} = \boldsymbol{z}$. This implies that U is surjective. Moreover, if $\boldsymbol{x} \in \mathbb{H}$ is given, then

$$\|\boldsymbol{x}\|^2 = \langle \boldsymbol{x}, \boldsymbol{x}\rangle = \langle \boldsymbol{x}, U^*U\boldsymbol{x}\rangle = \langle U\boldsymbol{x}, U\boldsymbol{x}\rangle = \|U\boldsymbol{x}\|^2 ;$$

thus $\|U\boldsymbol{x}\| = \|\boldsymbol{x}\|$ for all $\boldsymbol{x} \in \mathbb{H}$. □

§3.5 Unitary Operators and Unitary Matrices

Proof (cont.)

② ⇒ ③: Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{H}$. Then

$$\begin{aligned}
\|U(\boldsymbol{x}+\boldsymbol{y})\|^2 &= \langle U(\boldsymbol{x}+\boldsymbol{y}), U(\boldsymbol{x}+\boldsymbol{y})\rangle \\
&= \|U\boldsymbol{x}\|^2 + \langle U\boldsymbol{x}, U\boldsymbol{y}\rangle + \langle U\boldsymbol{y}, U\boldsymbol{x}\rangle + \|U\boldsymbol{y}\|^2 \\
&= \|U\boldsymbol{x}\|^2 + 2\mathrm{Re}(\langle U\boldsymbol{x}, U\boldsymbol{y}\rangle) + \|U\boldsymbol{y}\|^2 , \\
\|\boldsymbol{x}+\boldsymbol{y}\|^2 &= \|\boldsymbol{x}\|^2 + \langle \boldsymbol{x}, \boldsymbol{y}\rangle + \langle \boldsymbol{y}, \boldsymbol{x}\rangle + \|\boldsymbol{y}\|^2 \\
&= \|\boldsymbol{x}\|^2 + 2\mathrm{Re}(\langle \boldsymbol{x}, \boldsymbol{y}\rangle) + \|\boldsymbol{y}\|^2 ,
\end{aligned}$$

and

$$\begin{aligned}
\|U(\boldsymbol{x}+i\boldsymbol{y})\|^2 &= \langle U(\boldsymbol{x}+i\boldsymbol{y}), U(\boldsymbol{x}+i\boldsymbol{y})\rangle \\
&= \|U\boldsymbol{x}\|^2 + i\langle U\boldsymbol{x}, U\boldsymbol{y}\rangle - i\langle U\boldsymbol{y}, U\boldsymbol{x}\rangle + \|U\boldsymbol{y}\|^2 \\
&= \|U\boldsymbol{x}\|^2 - 2\mathrm{Im}(\langle U\boldsymbol{x}, U\boldsymbol{y}\rangle) + \|U\boldsymbol{y}\|^2 , \\
\|\boldsymbol{x}+i\boldsymbol{y}\|^2 &= \|\boldsymbol{x}\|^2 + i\langle \boldsymbol{x}, \boldsymbol{y}\rangle - i\langle \boldsymbol{y}, \boldsymbol{x}\rangle + \|\boldsymbol{y}\|^2 \\
&= \|\boldsymbol{x}\|^2 - 2\mathrm{Im}(\langle \boldsymbol{x}, \boldsymbol{y}\rangle) + \|\boldsymbol{y}\|^2 .
\end{aligned}$$

□

§3.5 Unitary Operators and Unitary Matrices

Proof (cont.)

② ⇒ ③: Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{H}$. Then

$$\begin{aligned}
\|U(\boldsymbol{x}+\boldsymbol{y})\|^2 &= \langle U(\boldsymbol{x}+\boldsymbol{y}), U(\boldsymbol{x}+\boldsymbol{y})\rangle \\
&= \|U\boldsymbol{x}\|^2 + \langle U\boldsymbol{x}, U\boldsymbol{y}\rangle + \langle U\boldsymbol{y}, U\boldsymbol{x}\rangle + \|U\boldsymbol{y}\|^2 \\
&= \|U\boldsymbol{x}\|^2 + 2\mathrm{Re}(\langle U\boldsymbol{x}, U\boldsymbol{y}\rangle) + \|U\boldsymbol{y}\|^2 , \\
\|\boldsymbol{x}+\boldsymbol{y}\|^2 &= \|\boldsymbol{x}\|^2 + \langle \boldsymbol{x}, \boldsymbol{y}\rangle + \langle \boldsymbol{y}, \boldsymbol{x}\rangle + \|\boldsymbol{y}\|^2 \\
&= \|\boldsymbol{x}\|^2 + 2\mathrm{Re}(\langle \boldsymbol{x}, \boldsymbol{y}\rangle) + \|\boldsymbol{y}\|^2 ,
\end{aligned}$$

and

$$\begin{aligned}
\|U(\boldsymbol{x}+i\boldsymbol{y})\|^2 &= \langle U(\boldsymbol{x}+i\boldsymbol{y}), U(\boldsymbol{x}+i\boldsymbol{y})\rangle \\
&= \|U\boldsymbol{x}\|^2 + i\langle U\boldsymbol{x}, U\boldsymbol{y}\rangle - i\langle U\boldsymbol{y}, U\boldsymbol{x}\rangle + \|U\boldsymbol{y}\|^2 \\
&= \|U\boldsymbol{x}\|^2 - 2\mathrm{Im}(\langle U\boldsymbol{x}, U\boldsymbol{y}\rangle) + \|U\boldsymbol{y}\|^2 , \\
\|\boldsymbol{x}+i\boldsymbol{y}\|^2 &= \|\boldsymbol{x}\|^2 + i\langle \boldsymbol{x}, \boldsymbol{y}\rangle - i\langle \boldsymbol{y}, \boldsymbol{x}\rangle + \|\boldsymbol{y}\|^2 \\
&= \|\boldsymbol{x}\|^2 - 2\mathrm{Im}(\langle \boldsymbol{x}, \boldsymbol{y}\rangle) + \|\boldsymbol{y}\|^2 .
\end{aligned}$$

□

§3.5 Unitary Operators and Unitary Matrices

Proof (cont.)

Since $\|U\boldsymbol{x}\| = \|\boldsymbol{x}\|$ for all $\boldsymbol{x} \in \mathbb{H}$, we have

$$\begin{aligned}\mathrm{Re}(\langle U\boldsymbol{x}, U\boldsymbol{y}\rangle) &= \mathrm{Re}(\langle \boldsymbol{x}, \boldsymbol{y}\rangle)\,,\\ \mathrm{Im}(\langle U\boldsymbol{x}, U\boldsymbol{y}\rangle) &= \mathrm{Im}(\langle \boldsymbol{x}, \boldsymbol{y}\rangle)\,.\end{aligned}$$

Therefore, $\langle U\boldsymbol{x}, U\boldsymbol{y}\rangle = \langle \boldsymbol{x}, \boldsymbol{y}\rangle$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{H}$.

③ ⇒ ①: Let $\boldsymbol{x} \in \mathbb{H}$ be given. Then

$$\langle U^*U\boldsymbol{x}, \boldsymbol{y}\rangle = \langle U\boldsymbol{x}, U\boldsymbol{y}\rangle = \langle \boldsymbol{x}, \boldsymbol{y}\rangle \qquad \forall\, \boldsymbol{y} \in \mathbb{H}\,;$$

thus $U^*U\boldsymbol{x} = \boldsymbol{x}$. This implies that $U^*U = \mathrm{Id}$ on $\mathbb{H}$.

On the other hand, since U is surjective, for each $\boldsymbol{y} \in \mathbb{H}$ there exists $\boldsymbol{x} \in \mathbb{H}$ such that $U\boldsymbol{x} = \boldsymbol{y}$. Using $U^*U = \mathrm{Id}$, this $\boldsymbol{x}$ must be $U^*\boldsymbol{y}$; thus $UU^*\boldsymbol{y} = \boldsymbol{y}$ for all $\boldsymbol{y} \in \mathbb{H}$. This shows that $UU^* = \mathrm{Id}$ on $\mathbb{H}$; thus U is unitary. □

§3.5 Unitary Operators and Unitary Matrices

Proof (cont.)

Since $\|U\boldsymbol{x}\| = \|\boldsymbol{x}\|$ for all $\boldsymbol{x} \in \mathbb{H}$, we have

$$\begin{aligned}\mathrm{Re}(\langle U\boldsymbol{x}, U\boldsymbol{y}\rangle) &= \mathrm{Re}(\langle \boldsymbol{x}, \boldsymbol{y}\rangle)\,,\\ \mathrm{Im}(\langle U\boldsymbol{x}, U\boldsymbol{y}\rangle) &= \mathrm{Im}(\langle \boldsymbol{x}, \boldsymbol{y}\rangle)\,.\end{aligned}$$

Therefore, $\langle U\boldsymbol{x}, U\boldsymbol{y}\rangle = \langle \boldsymbol{x}, \boldsymbol{y}\rangle$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{H}$.

③ $\Rightarrow$ ①: Let $\boldsymbol{x} \in \mathbb{H}$ be given. Then

$$\langle U^*U\boldsymbol{x}, \boldsymbol{y}\rangle = \langle U\boldsymbol{x}, U\boldsymbol{y}\rangle = \langle \boldsymbol{x}, \boldsymbol{y}\rangle \qquad \forall\, \boldsymbol{y} \in \mathbb{H}\,;$$

thus $U^*U\boldsymbol{x} = \boldsymbol{x}$. This implies that $U^*U = \mathrm{Id}$ on $\mathbb{H}$.

On the other hand, since U is surjective, for each $\boldsymbol{y} \in \mathbb{H}$ there exists $\boldsymbol{x} \in \mathbb{H}$ such that $U\boldsymbol{x} = \boldsymbol{y}$. Using $U^*U = \mathrm{Id}$, this $\boldsymbol{x}$ must be $U^*\boldsymbol{y}$; thus $UU^*\boldsymbol{y} = \boldsymbol{y}$ for all $\boldsymbol{y} \in \mathbb{H}$. This shows that $UU^* = \mathrm{Id}$ on $\mathbb{H}$; thus U is unitary. □

§3.5 Unitary Operators and Unitary Matrices

Proof (cont.)

Since $\|U\boldsymbol{x}\| = \|\boldsymbol{x}\|$ for all $\boldsymbol{x} \in \mathbb{H}$, we have

$$\begin{aligned}\mathrm{Re}(\langle U\boldsymbol{x}, U\boldsymbol{y}\rangle) &= \mathrm{Re}(\langle \boldsymbol{x}, \boldsymbol{y}\rangle)\,,\\ \mathrm{Im}(\langle U\boldsymbol{x}, U\boldsymbol{y}\rangle) &= \mathrm{Im}(\langle \boldsymbol{x}, \boldsymbol{y}\rangle)\,.\end{aligned}$$

Therefore, $\langle U\boldsymbol{x}, U\boldsymbol{y}\rangle = \langle \boldsymbol{x}, \boldsymbol{y}\rangle$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{H}$.

③ ⇒ ①: Let $\boldsymbol{x} \in \mathbb{H}$ be given. Then

$$\langle U^*U\boldsymbol{x}, \boldsymbol{y}\rangle = \langle U\boldsymbol{x}, U\boldsymbol{y}\rangle = \langle \boldsymbol{x}, \boldsymbol{y}\rangle \qquad \forall\, \boldsymbol{y} \in \mathbb{H}\,;$$

thus $U^*U\boldsymbol{x} = \boldsymbol{x}$. This implies that $U^*U = \mathrm{Id}$ on $\mathbb{H}$.

On the other hand, since U is surjective, for each $\boldsymbol{y} \in \mathbb{H}$ there exists $\boldsymbol{x} \in \mathbb{H}$ such that $U\boldsymbol{x} = \boldsymbol{y}$. Using $U^*U = \mathrm{Id}$, this $\boldsymbol{x}$ must be $U^*\boldsymbol{y}$; thus $UU^*\boldsymbol{y} = \boldsymbol{y}$ for all $\boldsymbol{y} \in \mathbb{H}$. This shows that $UU^* = \mathrm{Id}$ on $\mathbb{H}$; thus U is unitary. □

§3.5 Unitary Operators and Unitary Matrices

Corollary

Let $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ *be a Hilbert space over field* $\mathbb{C}$. *If* $U \in \mathrm{U}(\mathbb{H})$, *then* $\|U\| = 1$.

Definition

Let $(X, \|\cdot\|)$ be a Banach space, and $T \in \mathcal{B}(X)$. The spectrum of T, denoted by $\sigma(T)$, is the collection of all $\lambda \in \mathbb{C}$ for which the operator $T - \lambda\mathrm{Id}$ is not invertible. In other words,

$$\sigma(T) = \left\{\lambda \in \mathbb{C} \,\middle|\, (T - \lambda\mathrm{Id}) \text{ is not bijective}\right\}.$$

A number $\lambda \in \mathbb{C}$ is called an eigenvalue of T if $T - \lambda\mathrm{Id}$ is not one-to-one. The collection of all eigenvalues of T is called the point spectrum of T and is denoted by $\sigma_p(T)$.

§3.5 Unitary Operators and Unitary Matrices

Corollary

Let $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ *be a Hilbert space over field* $\mathbb{C}$. *If* $U \in \mathrm{U}(\mathbb{H})$, *then* $\|U\| = 1$.

Definition

Let $(X, \|\cdot\|)$ be a Banach space, and $T \in \mathcal{B}(X)$. The spectrum of T, denoted by $\sigma(T)$, is the collection of all $\lambda \in \mathbb{C}$ for which the operator $T - \lambda\mathrm{Id}$ is not invertible. In other words,

$$\sigma(T) = \left\{\lambda \in \mathbb{C} \,\middle|\, (T - \lambda\mathrm{Id}) \text{ is not bijective}\right\}.$$

A number $\lambda \in \mathbb{C}$ is called an eigenvalue of T if $T - \lambda\mathrm{Id}$ is not one-to-one. The collection of all eigenvalues of T is called the point spectrum of T and is denoted by $\sigma_p(T)$.

§3.5 Unitary Operators and Unitary Matrices

Corollary

Let $(\mathbb{H}, \langle\cdot,\cdot\rangle)$ *be a Hilbert space over field* $\mathbb{C}$*. If* $U \in \mathrm{U}(\mathbb{H})$*, then* $\|U\| = 1$*.*

Definition

Let $(X, \|\cdot\|)$ be a Banach space, and $T \in \mathcal{B}(X)$. The spectrum of T, denoted by $\sigma(T)$, is the collection of all $\lambda \in \mathbb{C}$ for which the operator $T - \lambda\mathrm{Id}$ is not invertible. In other words,

$$\sigma(T) = \left\{\lambda \in \mathbb{C} \,\middle|\, (T - \lambda\mathrm{Id}) \text{ is not bijective}\right\}.$$

A number $\lambda \in \mathbb{C}$ is called an eigenvalue of T if $T - \lambda\mathrm{Id}$ is not one-to-one. The collection of all eigenvalues of T is called the point spectrum of T and is denoted by $\sigma_p(T)$.

§3.5 Unitary Operators and Unitary Matrices

§3.5.2 Unitary matrices

Definition

A unitary matrix A is the matrix representation of some unitary map $U : \mathbb{H} \to \mathbb{H}$, where $\mathbb{H}$ is a finite dimensional inner product space over $\mathbb{C}$, relative to an orthonormal basis of $\mathbb{H}$.

By the definition of unitary maps, the facts that

$$[TS]_{\mathcal{B}_3,\mathcal{B}_1} = [T]_{\mathcal{B}_3,\mathcal{B}_2}[S]_{\mathcal{B}_2,\mathcal{B}_1} \quad \text{and} \quad [L^*]_{\mathcal{B},\widetilde{\mathcal{B}}} = [L]^{\dagger}_{\widetilde{\mathcal{B}},\mathcal{B}}$$

provide an alternative definition of unitary matrices:

Definition (Alternative Definition of Unitary Matrices)

A square matrix A is said to be unitary if $AA^{\dagger} = A^{\dagger}A = \mathrm{I}$. The collection of all $n \times n$ unitary matrices is denoted by $\mathrm{U}(n)$.

§3.5 Unitary Operators and Unitary Matrices

§3.5.2 Unitary matrices

Definition

A unitary matrix A is the matrix representation of some unitary map $U : \mathbb{H} \to \mathbb{H}$, where $\mathbb{H}$ is a finite dimensional inner product space over $\mathbb{C}$, relative to an orthonormal basis of $\mathbb{H}$.

By the definition of unitary maps, the facts that

$$[TS]_{\mathcal{B}_3,\mathcal{B}_1} = [T]_{\mathcal{B}_3,\mathcal{B}_2}[S]_{\mathcal{B}_2,\mathcal{B}_1} \quad \text{and} \quad [L^*]_{\mathcal{B},\widetilde{\mathcal{B}}} = [L]^{\dagger}_{\widetilde{\mathcal{B}},\mathcal{B}}$$

provide an alternative definition of unitary matrices:

Definition (Alternative Definition of Unitary Matrices)

A square matrix A is said to be unitary if $AA^{\dagger} = A^{\dagger}A = \mathrm{I}$. The collection of all $n \times n$ unitary matrices is denoted by $\mathrm{U}(n)$.

§3.5 Unitary Operators and Unitary Matrices

Corollary

If $A \in \mathrm{U}(n)$*, then* $A^{-1} = A^{\dagger}$ *and* $|\det(A)| = 1$.

Definition

The special unitary group of degree n, denoted by $\mathrm{SU}(n)$, is the collection of $n \times n$ unitary matrices with determinant 1. An element in $\mathrm{SU}(n)$ is called a special unitary matrix.

Definition (Walsh-Hadamard Matrix)

For $m \in \mathbb{N} \cup \{0\}$, the Walsh-Hadamard matrix H_m is a $2^m \times 2^m$ matrix defined recursively by

1. $\mathrm{H}_0 = 1$;
2. $\mathrm{H}_m = \dfrac{1}{\sqrt{2}} \begin{bmatrix} \mathrm{H}_{m-1} & \mathrm{H}_{m-1} \\ \mathrm{H}_{m-1} & -\mathrm{H}_{m-1} \end{bmatrix}$ for all $m \in \mathbb{N}$.

§3.5 Unitary Operators and Unitary Matrices

Corollary

If $A \in \mathrm{U}(n)$, *then* $A^{-1} = A^{\dagger}$ *and* $|\det(A)| = 1$.

Definition

The special unitary group of degree n, denoted by $\mathrm{SU}(n)$, is the collection of $n \times n$ unitary matrices with determinant 1. An element in $\mathrm{SU}(n)$ is called a special unitary matrix.

Definition (Walsh-Hadamard Matrix)

For $m \in \mathbb{N} \cup \{0\}$, the Walsh-Hadamard matrix H_m is a $2^m \times 2^m$ matrix defined recursively by

1. $\mathrm{H}_0 = 1$;
2. $\mathrm{H}_m = \dfrac{1}{\sqrt{2}} \begin{bmatrix} \mathrm{H}_{m-1} & \mathrm{H}_{m-1} \\ \mathrm{H}_{m-1} & -\mathrm{H}_{m-1} \end{bmatrix}$ for all $m \in \mathbb{N}$.

§3.5 Unitary Operators and Unitary Matrices

Corollary

If $A \in \mathrm{U}(n)$, then $A^{-1} = A^{\dagger}$ and $|\det(A)| = 1$.

Definition

The special unitary group of degree n, denoted by $\mathrm{SU}(n)$, is the collection of $n \times n$ unitary matrices with determinant 1. An element in $\mathrm{SU}(n)$ is called a special unitary matrix.

Definition (Walsh-Hadamard Matrix)

For $m \in \mathbb{N} \cup \{0\}$, the Walsh-Hadamard matrix H_m is a $2^m \times 2^m$ matrix defined recursively by

1. $\mathrm{H}_0 = 1$;
2. $\mathrm{H}_m = \dfrac{1}{\sqrt{2}} \begin{bmatrix} \mathrm{H}_{m-1} & \mathrm{H}_{m-1} \\ \mathrm{H}_{m-1} & -\mathrm{H}_{m-1} \end{bmatrix}$ for all $m \in \mathbb{N}$.

§3.5 Unitary Operators and Unitary Matrices

Remark:

1. We note that H_m is symmetric and orthogonal/unitary for all $m \in \mathbb{N}$ (which can be proved by induction).

2. The original definition of the Hadamard matrix (of order 2^m), again denoted by H_m, is a $2^m \times 2^m$ matrix defined recursively by

   ⓐ $\mathrm{H}_0 = 1$; ⓑ $\mathrm{H}_m = \begin{bmatrix} \mathrm{H}_{m-1} & \mathrm{H}_{m-1} \\ \mathrm{H}_{m-1} & -\mathrm{H}_{m-1} \end{bmatrix}$ for all $m \in \mathbb{N}$.

   However, in quantum computing we usually only consider unitary matrices, so the factor $1/\sqrt{2}$ is to normalized the original Hadamard matrices so that the norm of each colum (and also each row) all become 1. Therefore, the Hadamard matrices given in the definition last page is sometimes called the normalized Hadamard matrices.

§3.5 Unitary Operators and Unitary Matrices

Remark:

1. We note that H_m is symmetric and orthogonal/unitary for all $m \in \mathbb{N}$ (which can be proved by induction).
2. The original definition of the Hadamard matrix (of order 2^m), again denoted by H_m, is a $2^m \times 2^m$ matrix defined recursively by

   ⓐ $\mathrm{H}_0 = 1$; ⓑ $\mathrm{H}_m = \begin{bmatrix} \mathrm{H}_{m-1} & \mathrm{H}_{m-1} \\ \mathrm{H}_{m-1} & -\mathrm{H}_{m-1} \end{bmatrix}$ for all $m \in \mathbb{N}$.

   However, in quantum computing we usually only consider unitary matrices, so the factor $1/\sqrt{2}$ is to normalized the original Hadamard matrices so that the norm of each colum (and also each row) all become 1. Therefore, the Hadamard matrices given in the definition last page is sometimes called the normalized Hadamard matrices.

§3.5 Unitary Operators and Unitary Matrices

3. Let the (k,ℓ)-entry of H_m be denoted by $h_{k\ell}$; that is, $\mathrm{H}_m = [h_{k\ell}]_{1\leqslant k,\ell\leqslant 2^m}$. Then

$$h_{k\ell} = 2^{-\frac{m}{2}}(-1)^{(k-1)\bullet(\ell-1)},$$

where the bitwise dot product $\bullet$ of two numbers k and ℓ is given by

$$k\bullet\ell = \sum_{j=1}^{m} k_j\ell_j = k_1\ell_1 + k_2\ell_2 + \cdots + k_m\ell_m$$

if $k = (k_1k_2\cdots k_m)_2$ and $\ell = (\ell_1\ell_2\cdots\ell_m)_2$.

In matlab®, the bitwise dot product of x and y can be computed by

$$x\bullet y = \mathbf{de2bi}(x,n) * \mathbf{de2bi}(y,n)'$$

if both x and y can be expressed as n bits binary numbers.

§3.5 Unitary Operators and Unitary Matrices

Exercise: For matrices $A = [a_{k\ell}]$ and $B = [b_{k\ell}]$ of the same size $m \times n$, define the Hadamard product of A and B, denoted by $A \odot B$, as an $m \times n$ matrix whose (k, ℓ)-entry is give by $a_{k\ell} b_{k\ell}$; that is,

$$C = A \odot B\,, \quad C = [c_{k\ell}]\,, \quad c_{k\ell} = a_{k\ell} b_{k\ell}\,.$$

In matlab®, the Hadamard product of A and B can be computed by $A \odot B = A \mathbin{.*} B$. **In the following, we will always use $.*$ to denote the Hadamard product.**

§3.5 Unitary Operators and Unitary Matrices

Let H_n be the **unnormalized** Hadamard matrix whose (k,ℓ)-entry is given by $(-1)^{(k-1)\bullet(\ell-1)}$, and $\boldsymbol{r}_j$ be the $(j+1)$-th row of H_n. Define $\varphi : \{0,1\}^n \to \{\boldsymbol{r}_0, \boldsymbol{r}_1, \cdots, \boldsymbol{r}_{2^n-1}\}$ by

$$\varphi(j_1, j_2, \cdots, j_n) = \boldsymbol{r}_j \quad \text{if} \quad j = (j_1 j_2 \cdots j_n)_2 \,.$$

Show that $\varphi : (\{0,1\}^n, \oplus) \to \big(\{\boldsymbol{r}_0, \boldsymbol{r}_1, \cdots, \boldsymbol{r}_{2^n-1}\}, .*\big)$ is a **group isomorphism**, where $\oplus$ is the element-wise addition in $\mathbb{Z}_2$; that is,

$$(x_1, \cdots, x_n) \oplus (y_1, \cdots, y_n) = (x_1 \oplus y_1, \cdots, x_n \oplus y_n) \,.$$

In other words, show that $\varphi : \{0,1\}^n \to \{\boldsymbol{r}_0, \boldsymbol{r}_1, \cdots, \boldsymbol{r}_{2^n-1}\}$ defined above is a bijection and

$$\varphi\big((k_1, \cdots, k_n) \oplus (\ell_1, \cdots, \ell_n)\big) = \boldsymbol{r}_k .* \boldsymbol{r}_\ell$$

for all $k = (k_1 k_2 \cdots k_n)_2$ and $\ell = (\ell_1 \ell_2 \cdots \ell_n)_2$.

§3.5 Unitary Operators and Unitary Matrices

Let H_n be the **unnormalized** Hadamard matrix whose (k,ℓ)-entry is given by $(-1)^{(k-1)\bullet(\ell-1)}$, and $\boldsymbol{r}_j$ be the $(j+1)$-th row of H_n. Define $\varphi:\{0,1\}^n\to\{\boldsymbol{r}_0,\boldsymbol{r}_1,\cdots,\boldsymbol{r}_{2^n-1}\}$ by

$$\varphi(j_1,j_2,\cdots,j_n)=\boldsymbol{r}_j \quad \text{if} \quad j=(j_1j_2\cdots j_n)_2\,.$$

Show that $\varphi:(\{0,1\}^n,\oplus)\to\big(\{\boldsymbol{r}_0,\boldsymbol{r}_1,\cdots,\boldsymbol{r}_{2^n-1}\},\ .*\big)$ is a **group isomorphism**, where $\oplus$ is the element-wise addition in $\mathbb{Z}_2$; that is,

$$(x_1,\cdots,x_n)\oplus(y_1,\cdots,y_n)=(x_1\oplus y_1,\cdots,x_n\oplus y_n)\,.$$

In other words, show that $\varphi:\{0,1\}^n\to\{\boldsymbol{r}_0,\boldsymbol{r}_1,\cdots,\boldsymbol{r}_{2^n-1}\}$ defined above is a bijection and

$$\varphi\big((k_1,\cdots,k_n)\oplus(\ell_1,\cdots,\ell_n)\big)=\boldsymbol{r}_k\ .*\ \boldsymbol{r}_\ell$$

for all $k=(k_1k_2\cdots k_n)_2$ and $\ell=(\ell_1\ell_2\cdots\ell_n)_2$.

§3.6 Tensor Products

Recall that for $\alpha_0, \alpha_1, \beta_0, \beta_1$ satisfying $|\alpha_0|^2+|\alpha_1|^2 = |\beta_0|^2+|\beta_1|^2 = 1$, the quantum states $|\psi_1\rangle = \alpha_0|0\rangle+\alpha_1|1\rangle$ and $|\psi_2\rangle = \beta_0|0\rangle+\beta_1|1\rangle$ are the short-hand notation for the two wave functions

$$\psi_1(x,t) = \begin{cases} \alpha_0 & \text{if } x = |0\rangle\,, \\ \alpha_1 & \text{if } x = |1\rangle\,, \end{cases} \qquad \psi_2(y,t) = \begin{cases} \beta_0 & \text{if } y = |0\rangle\,, \\ \beta_1 & \text{if } y = |1\rangle\,, \end{cases}$$

If two qubits with probability amplitude ψ_1 and ψ_2 are manipulated **independently**, then the probability amplitude of the combined system is $\psi(x,y) = \psi_1(x)\psi_2(y)$ given by

$$\psi(x,y) = \begin{cases} \alpha_0\beta_0 & \text{if } x = y = |0\rangle\,, \\ \alpha_0\beta_1 & \text{if } x = |0\rangle \text{ and } y = |1\rangle\,, \\ \alpha_1\beta_0 & \text{if } x = |1\rangle \text{ and } y = |0\rangle\,, \\ \alpha_1\beta_1 & \text{if } x = y = |1\rangle\,, \end{cases}$$

and the short-hand notation for the wave function above is

$$|\psi\rangle = \alpha_0\beta_0|00\rangle + \alpha_0\beta_1|01\rangle + \alpha_1\beta_0|10\rangle + \alpha_1\beta_1|11\rangle\,.$$

§3.6 Tensor Products

Recall that for $\alpha_0, \alpha_1, \beta_0, \beta_1$ satisfying $|\alpha_0|^2+|\alpha_1|^2 = |\beta_0|^2+|\beta_1|^2 = 1$, the quantum states $|\psi_1\rangle = \alpha_0|0\rangle+\alpha_1|1\rangle$ and $|\psi_2\rangle = \beta_0|0\rangle+\beta_1|1\rangle$ are the short-hand notation for the two wave functions

$$\psi_1(x,t) = \begin{cases} \alpha_0 & \text{if } x = |0\rangle, \\ \alpha_1 & \text{if } x = |1\rangle, \end{cases} \qquad \psi_2(y,t) = \begin{cases} \beta_0 & \text{if } y = |0\rangle, \\ \beta_1 & \text{if } y = |1\rangle, \end{cases}$$

If two qubits with probability amplitude ψ_1 and ψ_2 are manipulated **independently**, then the probability amplitude of the combined system is $\psi(x,y) = \psi_1(x)\psi_2(y)$ given by

$$\psi(x,y) = \begin{cases} \alpha_0\beta_0 & \text{if } x = y = |0\rangle, \\ \alpha_0\beta_1 & \text{if } x = |0\rangle \text{ and } y = |1\rangle, \\ \alpha_1\beta_0 & \text{if } x = |1\rangle \text{ and } y = |0\rangle, \\ \alpha_1\beta_1 & \text{if } x = y = |1\rangle, \end{cases}$$

and the short-hand notation for the wave function above is

$$|\psi\rangle = \alpha_0\beta_0|00\rangle + \alpha_0\beta_1|01\rangle + \alpha_1\beta_0|10\rangle + \alpha_1\beta_1|11\rangle .$$

§3.6 Tensor Products

In this section, we would like design a proper idea of "product" so that the product of $|\psi_1\rangle$ and $|\psi_2\rangle$ is $|\psi\rangle$. This product, called **the tensor product** and denoted by $\otimes$, satisfies $|\psi\rangle = |\psi_1\rangle \otimes |\psi_2\rangle$.

Caution: The definition of the tensor product given in the following is purely mathematics. You do **not** need to understand this section fully in order to learn quantum computing; however, we encourage you to go through this once for it will explain a lot of things that normal textbooks for quantum computing will not talk about.

Idea: Recall that we talk about "**multi-linearity**" of the computations of the tensor product of quantum states in Chapter 2. To make sense of the tensor product of vectors, we need to treat these vectors as "**linear functionals**" so that we can treat $\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_n$ as an element in $\mathcal{L}(\mathbb{V}_1, \cdots, \mathbb{V}_n; \mathbb{F})$.

§3.6 Tensor Products

In this section, we would like design a proper idea of "product" so that the product of $|\psi_1\rangle$ and $|\psi_2\rangle$ is $|\psi\rangle$. This product, called **the tensor product** and denoted by $\otimes$, satisfies $|\psi\rangle = |\psi_1\rangle \otimes |\psi_2\rangle$.

Caution: The definition of the tensor product given in the following is purely mathematics. You do **not** need to understand this section fully in order to learn quantum computing; however, we encourage you to go through this once for it will explain a lot of things that normal textbooks for quantum computing will not talk about.

Idea: Recall that we talk about "**multi-linearity**" of the computations of the tensor product of quantum states in Chapter 2. To make sense of the tensor product of vectors, we need to treat these vectors as "**linear functionals**" so that we can treat $\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_n$ as an element in $\mathscr{L}(\mathbb{V}_1, \cdots, \mathbb{V}_n; \mathbb{F})$.

§3.6 Tensor Products

In this section, we would like design a proper idea of "product" so that the product of $|\psi_1\rangle$ and $|\psi_2\rangle$ is $|\psi\rangle$. This product, called **the tensor product** and denoted by $\otimes$, satisfies $|\psi\rangle = |\psi_1\rangle \otimes |\psi_2\rangle$.

Caution: The definition of the tensor product given in the following is purely mathematics. You do **not** need to understand this section fully in order to learn quantum computing; however, we encourage you to go through this once for it will explain a lot of things that normal textbooks for quantum computing will not talk about.

Idea: Recall that we talk about "**multi-linearity**" of the computations of the tensor product of quantum states in Chapter 2. To make sense of the tensor product of vectors, we need to treat these vectors as "**linear functionals**" so that we can treat $\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_n$ as an element in $\mathscr{L}(\mathbb{V}_1, \cdots, \mathbb{V}_n; \mathbb{F})$.

§3.6 Tensor Products

In this section, we would like design a proper idea of "product" so that the product of $|\psi_1\rangle$ and $|\psi_2\rangle$ is $|\psi\rangle$. This product, called **the tensor product** and denoted by $\otimes$, satisfies $|\psi\rangle = |\psi_1\rangle \otimes |\psi_2\rangle$.

Caution: The definition of the tensor product given in the following is purely mathematics. You do **not** need to understand this section fully in order to learn quantum computing; however, we encourage you to go through this once for it will explain a lot of things that normal textbooks for quantum computing will not talk about.

Idea: Recall that we talk about "**multi-linearity**" of the computations of the tensor product of quantum states in Chapter 2. To make sense of the tensor product of vectors, we need to treat these vectors as "**linear functionals**" so that we can treat $\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_n$ as an element in $\mathcal{L}(\mathbb{V}_1, \cdots, \mathbb{V}_n; \mathbb{F})$.

§3.6 Tensor Product of Vector Spaces

§3.6.1 Tensor product

Let $\mathbb{V}$ be a vector space over a scalar field $\mathbb{F}$. Then $\mathbb{V}'$ is also a vector space over $\mathbb{F}$. This enables us to consider $(\mathbb{V}')'$, the algebraic dual space of $\mathbb{V}'$. In general, $(\mathbb{V}')' = \mathbb{V}$ is not true, but there is an injection $\iota : \mathbb{V} \to (\mathbb{V}')'$ in the sense that

$$\iota(v)(f) \equiv f(v) \qquad \forall\, f \in \mathbb{V}' . \tag{3}$$

The linear embedding $\iota : \mathbb{V} \to (\mathbb{V}')'$ is a natural vector space isomorphism provided, again, $\dim(\mathbb{V})$ is finite, the proof being evident as the embedding is a linear and injective map between spaces with equal finite dimension.

The embedding (3) permits us to define a vector space $\mathbb{V}_1 \otimes \mathbb{V}_2 \otimes \cdots \otimes \mathbb{V}_n$ called the tensor product of vector spaces $\mathbb{V}_1, \mathbb{V}_2, \cdots, \mathbb{V}_n$ with a common field of scalars $\mathbb{F}$.

§3.6 Tensor Product of Vector Spaces

§3.6.1 Tensor product

Let $\mathbb{V}$ be a vector space over a scalar field $\mathbb{F}$. Then $\mathbb{V}'$ is also a vector space over $\mathbb{F}$. This enables us to consider $(\mathbb{V}')'$, the algebraic dual space of $\mathbb{V}'$. In general, $(\mathbb{V}')' = \mathbb{V}$ is not true, but there is an injection $\iota : \mathbb{V} \to (\mathbb{V}')'$ in the sense that

$$\iota(v)(f) \equiv f(v) \qquad \forall\, f \in \mathbb{V}'. \tag{3}$$

The linear embedding $\iota : \mathbb{V} \to (\mathbb{V}')'$ is a natural vector space isomorphism provided, again, $\dim(\mathbb{V})$ is finite, the proof being evident as the embedding is a linear and injective map between spaces with equal finite dimension.

The embedding (3) permits us to define a vector space $\mathbb{V}_1 \otimes \mathbb{V}_2 \otimes \cdots \otimes \mathbb{V}_n$ called the tensor product of vector spaces $\mathbb{V}_1, \mathbb{V}_2, \cdots, \mathbb{V}_n$ with a common field of scalars $\mathbb{F}$.

§3.6 Tensor Product of Vector Spaces

§3.6.1 Tensor product

Let $\mathbb{V}$ be a vector space over a scalar field $\mathbb{F}$. Then $\mathbb{V}'$ is also a vector space over $\mathbb{F}$. This enables us to consider $(\mathbb{V}')'$, the algebraic dual space of $\mathbb{V}'$. In general, $(\mathbb{V}')' = \mathbb{V}$ is not true, but there is an injection $\iota : \mathbb{V} \to (\mathbb{V}')'$ in the sense that

$$\iota(v)(f) \equiv f(v) \qquad \forall\, f \in \mathbb{V}'\,. \tag{3}$$

The linear embedding $\iota : \mathbb{V} \to (\mathbb{V}')'$ is a natural vector space isomorphism provided, again, $\dim(\mathbb{V})$ is finite, the proof being evident as the embedding is a linear and injective map between spaces with equal finite dimension.

The embedding (3) permits us to define a vector space $\mathbb{V}_1 \otimes \mathbb{V}_2 \otimes \cdots \otimes \mathbb{V}_n$ called the tensor product of vector spaces $\mathbb{V}_1, \mathbb{V}_2, \cdots, \mathbb{V}_n$ with a common field of scalars $\mathbb{F}$.

§3.6 Tensor Product of Vector Spaces

§3.6.1 Tensor product

Let $\mathbb{V}$ be a vector space over a scalar field $\mathbb{F}$. Then $\mathbb{V}'$ is also a vector space over $\mathbb{F}$. This enables us to consider $(\mathbb{V}')'$, the algebraic dual space of $\mathbb{V}'$. In general, $(\mathbb{V}')' = \mathbb{V}$ is not true, but there is an injection $\iota : \mathbb{V} \to (\mathbb{V}')'$ in the sense that

$$\iota(v)(f) \equiv f(v) \qquad \forall\, f \in \mathbb{V}'. \tag{3}$$

The linear embedding $\iota : \mathbb{V} \to (\mathbb{V}')'$ is a natural vector space isomorphism provided, again, $\dim(\mathbb{V})$ is finite, the proof being evident as the embedding is a linear and injective map between spaces with equal finite dimension.

The embedding (3) permits us to define a vector space $\mathbb{V}_1 \otimes \mathbb{V}_2 \otimes \cdots \otimes \mathbb{V}_n$ called the tensor product of vector spaces $\mathbb{V}_1, \mathbb{V}_2, \cdots, \mathbb{V}_n$ with a common field of scalars $\mathbb{F}$.

§3.6 Tensor Product of Vector Spaces

§3.6.1 Tensor product

Let $\mathbb{V}$ be a vector space over a scalar field $\mathbb{F}$. Then $\mathbb{V}'$ is also a vector space over $\mathbb{F}$. This enables us to consider $(\mathbb{V}')'$, the algebraic dual space of $\mathbb{V}'$. In general, $(\mathbb{V}')' = \mathbb{V}$ is not true, but there is an injection $\iota : \mathbb{V} \to (\mathbb{V}')'$ in the sense that

$$\iota(v)(f) \equiv f(v) \qquad \forall\, f \in \mathbb{V}' . \tag{3}$$

The linear embedding $\iota : \mathbb{V} \to (\mathbb{V}')'$ is a natural vector space isomorphism provided, again, $\dim(\mathbb{V})$ is finite, the proof being evident as the embedding is a linear and injective map between spaces with equal finite dimension.

The embedding (3) permits us to define a vector space $\mathbb{V}_1 \otimes \mathbb{V}_2 \otimes \cdots \otimes \mathbb{V}_n$ called the tensor product of vector spaces $\mathbb{V}_1, \mathbb{V}_2, \cdots, \mathbb{V}_n$ with a common field of scalars $\mathbb{F}$.

§3.6 Tensor Product of Vector Spaces

Before proceeding to the definition of the tensor product of vector spaces, we first look at the tensor product of vectors.

Definition

Let $\mathbb{V}_1, \cdots, \mathbb{V}_n$ be vector spaces over a common scalar field $\mathbb{F}$, and $\boldsymbol{v}_j \in \mathbb{V}_j$ be given for $1 \leqslant j \leqslant n$. The tensor product $\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_n$ is a function from $\mathbb{V}_1' \oplus \cdots \oplus \mathbb{V}_n'$ to $\mathbb{F}$ defined by

$$(\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_n)(f_1, ..., f_n) = \prod_{j=1}^{n} f_j(\boldsymbol{v}_j) \equiv f_1(\boldsymbol{v}_1) \cdot \cdots \cdot f_n(\boldsymbol{v}_n)\,.$$

Warning: The tensor product space $\mathbb{V}_1 \otimes \cdots \otimes \mathbb{V}_n$ is **NOT** the collection of all vectors of the form $\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_n$ with $v_j \in \mathbb{V}_j$.

§3.6 Tensor Product of Vector Spaces

Before proceeding to the definition of the tensor product of vector spaces, we first look at the tensor product of vectors.

Definition

Let $\mathbb{V}_1, \cdots, \mathbb{V}_n$ be vector spaces over a common scalar field $\mathbb{F}$, and $\boldsymbol{v}_j \in \mathbb{V}_j$ be given for $1 \leqslant j \leqslant n$. The tensor product $\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_n$ is a function from $\mathbb{V}_1' \oplus \cdots \oplus \mathbb{V}_n'$ to $\mathbb{F}$ defined by

$$(\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_n)(f_1, \dots, f_n) = \prod_{j=1}^{n} f_j(\boldsymbol{v}_j) \equiv f_1(\boldsymbol{v}_1) \cdot \cdots \cdot f_n(\boldsymbol{v}_n) .$$

Warning: The tensor product space $\mathbb{V}_1 \otimes \cdots \otimes \mathbb{V}_n$ is **NOT** the collection of all vectors of the form $\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_n$ with $v_j \in \mathbb{V}_j$.

§3.6 Tensor Product of Vector Spaces

Proposition

Let $\mathbb{U}, \mathbb{V}, \mathbb{W}$ *be vector spaces over a common scalar field* $\mathbb{F}$*, and* $\boldsymbol{u} \in \mathbb{U}$, $\boldsymbol{v} \in \mathbb{V}$ *and* $\boldsymbol{w} \in \mathbb{W}$. *Then*

$$\boldsymbol{u} \otimes \boldsymbol{v} \otimes \boldsymbol{w} = (\boldsymbol{u} \otimes \boldsymbol{v}) \otimes \boldsymbol{w} = \boldsymbol{u} \otimes (\boldsymbol{v} \otimes \boldsymbol{w}) .$$

Proof.

Let $f \in \mathbb{U}'$, $g \in \mathbb{V}'$ and $h \in \mathbb{W}'$. Then

$$\begin{aligned}
&(\boldsymbol{u} \otimes \boldsymbol{v} \otimes \boldsymbol{w})(f, g, h) \\
&\quad = f(\boldsymbol{u}) \cdot g(\boldsymbol{v}) \cdot h(\boldsymbol{w}) = \big[f(\boldsymbol{u}) \cdot g(\boldsymbol{v})\big] \cdot h(\boldsymbol{w}) = (\boldsymbol{u} \otimes \boldsymbol{v})(f, g) \cdot h(\boldsymbol{w}) \\
&\quad = \big[(\boldsymbol{u} \otimes \boldsymbol{v}) \otimes \boldsymbol{w}\big]\big((f, g), h\big) = \big[(\boldsymbol{u} \otimes \boldsymbol{v}) \otimes \boldsymbol{w}\big](f, g, h)
\end{aligned}$$

so that $\boldsymbol{u}\otimes\boldsymbol{v}\otimes\boldsymbol{w} = (\boldsymbol{u}\otimes\boldsymbol{v})\otimes\boldsymbol{w}$. The identity $\boldsymbol{u}\otimes\boldsymbol{v}\otimes\boldsymbol{w} = \boldsymbol{u}\otimes(\boldsymbol{v}\otimes\boldsymbol{w})$ can be proved in the similar fashion. □

§3.6 Tensor Product of Vector Spaces

Proposition

Let $\mathbb{U}, \mathbb{V}, \mathbb{W}$ be vector spaces over a common scalar field $\mathbb{F}$, and $\boldsymbol{u} \in \mathbb{U}$, $\boldsymbol{v} \in \mathbb{V}$ and $\boldsymbol{w} \in \mathbb{W}$. Then

$$\boldsymbol{u} \otimes \boldsymbol{v} \otimes \boldsymbol{w} = (\boldsymbol{u} \otimes \boldsymbol{v}) \otimes \boldsymbol{w} = \boldsymbol{u} \otimes (\boldsymbol{v} \otimes \boldsymbol{w}) .$$

Proof.

Let $f \in \mathbb{U}'$, $g \in \mathbb{V}'$ and $h \in \mathbb{W}'$. Then

$$\begin{aligned}
&(\boldsymbol{u} \otimes \boldsymbol{v} \otimes \boldsymbol{w})(f, g, h) \\
&\quad = f(\boldsymbol{u}) \cdot g(\boldsymbol{v}) \cdot h(\boldsymbol{w}) = \big[f(\boldsymbol{u}) \cdot g(\boldsymbol{v})\big] \cdot h(\boldsymbol{w}) = (\boldsymbol{u} \otimes \boldsymbol{v})(f, g) \cdot h(\boldsymbol{w}) \\
&\quad = \big[(\boldsymbol{u} \otimes \boldsymbol{v}) \otimes \boldsymbol{w}\big]\big((f, g), h\big) = \big[(\boldsymbol{u} \otimes \boldsymbol{v}) \otimes \boldsymbol{w}\big](f, g, h)
\end{aligned}$$

so that $\boldsymbol{u} \otimes \boldsymbol{v} \otimes \boldsymbol{w} = (\boldsymbol{u} \otimes \boldsymbol{v}) \otimes \boldsymbol{w}$. The identity $\boldsymbol{u} \otimes \boldsymbol{v} \otimes \boldsymbol{w} = \boldsymbol{u} \otimes (\boldsymbol{v} \otimes \boldsymbol{w})$ can be proved in the similar fashion. □

§3.6 Tensor Product of Vector Spaces

Proposition

Let $\mathbb{V}_1, \cdots, \mathbb{V}_n$ be vector spaces over a common scalar field $\mathbb{F}$. For $1 \leqslant j \leqslant n$, let $\boldsymbol{v}_\ell \in \mathbb{V}_\ell$ for $\ell \neq j$, $\boldsymbol{u}_j, \boldsymbol{w}_j \in \mathbb{V}_j$, and $c \in \mathbb{F}$. Then

$$\begin{aligned}&\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_{j-1} \otimes (c\boldsymbol{u}_j + \boldsymbol{w}_j) \otimes \boldsymbol{v}_{j+1} \otimes \cdots \otimes \boldsymbol{v}_n \\ &= c(\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_{j-1} \otimes \boldsymbol{u}_j \otimes \boldsymbol{v}_{j+1} \otimes \cdots \otimes \boldsymbol{v}_n) \\ &\quad + (\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_{j-1} \otimes \boldsymbol{w}_j \otimes \boldsymbol{v}_{j+1} \otimes \cdots \otimes \boldsymbol{v}_n) .\end{aligned}$$

Proof.

Let $f_\ell \in \mathbb{V}_\ell'$ for $1 \leqslant \ell \leqslant n$ be given. Then

$$\begin{aligned}&(\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_{j-1} \otimes (c\boldsymbol{u}_j + \boldsymbol{w}_j) \otimes \boldsymbol{v}_{j+1} \otimes \cdots \otimes \boldsymbol{v}_n)(f_1, \cdots, f_n) \\ &= f_1(\boldsymbol{v}_1) \cdots f_{j-1}(\boldsymbol{v}_{j-1}) \cdot f_j(c\boldsymbol{u}_j + \boldsymbol{w}_j) \cdot f_{j+1}(\boldsymbol{v}_j) \cdots f_n(\boldsymbol{v}_n) \\ &= c f_1(\boldsymbol{v}_1) \cdots f_{j-1}(\boldsymbol{v}_{j-1}) \cdot f_j(\boldsymbol{u}_j) \cdot f_{j+1}(\boldsymbol{v}_j) \cdots f_n(\boldsymbol{v}_n) \\ &\quad + f_1(\boldsymbol{v}_1) \cdots f_{j-1}(\boldsymbol{v}_{j-1}) \cdot f_j(\boldsymbol{w}_j) \cdot f_{j+1}(\boldsymbol{v}_j) \cdots f_n(\boldsymbol{v}_n)\end{aligned}$$

which establishes the proposition. □

§3.6 Tensor Product of Vector Spaces

Proposition

Let $\mathbb{V}_1, \cdots, \mathbb{V}_n$ be vector spaces over a common scalar field $\mathbb{F}$. For $1 \leqslant j \leqslant n$, let $\boldsymbol{v}_\ell \in \mathbb{V}_\ell$ for $\ell \neq j$, $\boldsymbol{u}_j, \boldsymbol{w}_j \in \mathbb{V}_j$, and $c \in \mathbb{F}$. Then

$$\begin{aligned}
&\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_{j-1} \otimes (c\boldsymbol{u}_j + \boldsymbol{w}_j) \otimes \boldsymbol{v}_{j+1} \otimes \cdots \otimes \boldsymbol{v}_n \\
&= c\,(\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_{j-1} \otimes \boldsymbol{u}_j \otimes \boldsymbol{v}_{j+1} \otimes \cdots \otimes \boldsymbol{v}_n) \\
&\quad + (\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_{j-1} \otimes \boldsymbol{w}_j \otimes \boldsymbol{v}_{j+1} \otimes \cdots \otimes \boldsymbol{v}_n)\,.
\end{aligned}$$

Proof.

Let $f_\ell \in \mathbb{V}_\ell'$ for $1 \leqslant \ell \leqslant n$ be given. Then

$$\begin{aligned}
&(\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_{j-1} \otimes (c\boldsymbol{u}_j + \boldsymbol{w}_j) \otimes \boldsymbol{v}_{j+1} \otimes \cdots \otimes \boldsymbol{v}_n)(f_1, \cdots, f_n) \\
&= f_1(\boldsymbol{v}_1) \cdots f_{j-1}(\boldsymbol{v}_{j-1}) \cdot f_j(c\boldsymbol{u}_j + \boldsymbol{w}_j) \cdot f_{j+1}(\boldsymbol{v}_j) \cdots f_n(\boldsymbol{v}_n) \\
&= c\, f_1(\boldsymbol{v}_1) \cdots f_{j-1}(\boldsymbol{v}_{j-1}) \cdot f_j(\boldsymbol{u}_j) \cdot f_{j+1}(\boldsymbol{v}_j) \cdots f_n(\boldsymbol{v}_n) \\
&\quad + f_1(\boldsymbol{v}_1) \cdots f_{j-1}(\boldsymbol{v}_{j-1}) \cdot f_j(\boldsymbol{w}_j) \cdot f_{j+1}(\boldsymbol{v}_j) \cdots f_n(\boldsymbol{v}_n)
\end{aligned}$$

which establishes the proposition. □

§3.6 Tensor Product of Vector Spaces

Proposition

Let $\mathbb{V}_1, \cdots, \mathbb{V}_n$ be vector spaces over a common scalar field $\mathbb{F}$. For $1 \leqslant j \leqslant n$, let $\boldsymbol{v}_\ell \in \mathbb{V}_\ell$ for $\ell \neq j$, $\boldsymbol{u}_j, \boldsymbol{w}_j \in \mathbb{V}_j$, and $c \in \mathbb{F}$. Then

$$\begin{aligned}
&\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_{j-1} \otimes (c\boldsymbol{u}_j + \boldsymbol{w}_j) \otimes \boldsymbol{v}_{j+1} \otimes \cdots \otimes \boldsymbol{v}_n \\
&= c\,(\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_{j-1} \otimes \boldsymbol{u}_j \otimes \boldsymbol{v}_{j+1} \otimes \cdots \otimes \boldsymbol{v}_n) \\
&\quad + (\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_{j-1} \otimes \boldsymbol{w}_j \otimes \boldsymbol{v}_{j+1} \otimes \cdots \otimes \boldsymbol{v}_n)\,.
\end{aligned}$$

Proof.

Let $f_\ell \in \mathbb{V}_\ell'$ for $1 \leqslant \ell \leqslant n$ be given. Then

$$\begin{aligned}
&\big(\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_{j-1} \otimes (c\boldsymbol{u}_j + \boldsymbol{w}_j) \otimes \boldsymbol{v}_{j+1} \otimes \cdots \otimes \boldsymbol{v}_n\big)(f_1, \cdots, f_n) \\
&= f_1(\boldsymbol{v}_1) \cdots f_{j-1}(\boldsymbol{v}_{j-1}) \cdot f_j(c\boldsymbol{u}_j + \boldsymbol{w}_j) \cdot f_{j+1}(\boldsymbol{v}_j) \cdots f_n(\boldsymbol{v}_n) \\
&= c\,(\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_{j-1} \otimes \boldsymbol{u}_j \otimes \boldsymbol{v}_{j+1} \otimes \cdots \otimes \boldsymbol{v}_n)(f_1, \cdots, f_n) \\
&\quad + (\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_{j-1} \otimes \boldsymbol{w}_j \otimes \boldsymbol{v}_{j+1} \otimes \cdots \otimes \boldsymbol{v}_n)(f_1, \cdots, f_n)
\end{aligned}$$

which establishes the proposition. □

§3.6 Tensor Product of Vector Spaces

Proposition

Let $\mathbb{V}_1, \cdots, \mathbb{V}_n$ be vector spaces over a common scalar field $\mathbb{F}$. For $1 \leqslant j \leqslant n$, let $\boldsymbol{v}_\ell \in \mathbb{V}_\ell$ for $\ell \neq j$, $\boldsymbol{u}_j, \boldsymbol{w}_j \in \mathbb{V}_j$, and $c \in \mathbb{F}$. Then

$$\begin{aligned}&\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_{j-1} \otimes (c\boldsymbol{u}_j + \boldsymbol{w}_j) \otimes \boldsymbol{v}_{j+1} \otimes \cdots \otimes \boldsymbol{v}_n \\ &= c\,(\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_{j-1} \otimes \boldsymbol{u}_j \otimes \boldsymbol{v}_{j+1} \otimes \cdots \otimes \boldsymbol{v}_n) \\ &\quad + (\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_{j-1} \otimes \boldsymbol{w}_j \otimes \boldsymbol{v}_{j+1} \otimes \cdots \otimes \boldsymbol{v}_n)\,.\end{aligned}$$

Proof.

Let $f_\ell \in \mathbb{V}_\ell'$ for $1 \leqslant \ell \leqslant n$ be given. Then

$$\begin{aligned}&\big(\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_{j-1} \otimes (c\boldsymbol{u}_j + \boldsymbol{w}_j) \otimes \boldsymbol{v}_{j+1} \otimes \cdots \otimes \boldsymbol{v}_n\big)(f_1, \cdots, f_n) \\ &= f_1(\boldsymbol{v}_1) \cdots f_{j-1}(\boldsymbol{v}_{j-1}) \cdot f_j(c\boldsymbol{u}_j + \boldsymbol{w}_j) \cdot f_{j+1}(\boldsymbol{v}_j) \cdots f_n(\boldsymbol{v}_n) \\ &= c\,(\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_{j-1} \otimes \boldsymbol{u}_j \otimes \boldsymbol{v}_{j+1} \otimes \cdots \otimes \boldsymbol{v}_n)(f_1, \cdots, f_n) \\ &\quad + (\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_{j-1} \otimes \boldsymbol{w}_j \otimes \boldsymbol{v}_{j+1} \otimes \cdots \otimes \boldsymbol{v}_n)(f_1, \cdots, f_n)\end{aligned}$$

which establishes the proposition. □

§3.6 Tensor Product of Vector Spaces

Proposition

Let $\mathbb{V}_1, \cdots, \mathbb{V}_n$ be vector spaces over a common scalar field $\mathbb{F}$, and $\boldsymbol{v}_j \in \mathbb{V}_j$ be given for $1 \leqslant j \leqslant n$. Then $\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_n \in \mathcal{L}(\mathbb{V}_1', \cdots, \mathbb{V}_n'; \mathbb{F})$.

Proof.

Let $1 \leqslant j \leqslant n$, $f_\ell \in \mathbb{V}_\ell'$ for $\ell \neq j$, g_j, $h_j \in \mathbb{V}_j'$ and $c \in \mathbb{F}$. Then

$$\begin{aligned}
&(\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_n)(f_1, \cdots, f_{j-1}, cg_j + h_j, f_{j+1}, \cdots, f_n) \\
&\quad = f_1(\boldsymbol{v}_1) \cdots f_{j-1}(\boldsymbol{v}_{j-1}) \cdot \big[cg_j(\boldsymbol{v}_j) + h_j(\boldsymbol{v}_j)\big] \cdot f_{j+1}(\boldsymbol{v}_{j+1}) \cdots f_n(\boldsymbol{v}_n) \\
&\quad = c\, f_1(\boldsymbol{v}_1) \cdots f_{j-1}(\boldsymbol{v}_{j-1}) \cdot g_j(\boldsymbol{v}_j) \cdot f_{j+1}(\boldsymbol{v}_{j+1}) \cdots f_n(\boldsymbol{v}_n) \\
&\qquad + f_1(\boldsymbol{v}_1) \cdots f_{j-1}(\boldsymbol{v}_{j-1}) \cdot h_j(\boldsymbol{v}_j) \cdot f_{j+1}(\boldsymbol{v}_{j+1}) \cdots f_n(\boldsymbol{v}_n) \\
&\quad = c(\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_n)(f_1, \cdots, f_{j-1}, g_j, f_{j+1}, \cdots, f_n) \\
&\qquad + (\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_n)(f_1, \cdots, f_{j-1}, h_j, f_{j+1}, \cdots, f_n)
\end{aligned}$$

which shows that $\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_n$ satisfies the multi-linearity. □

§3.6 Tensor Product of Vector Spaces

PROPOSITION

Let $\mathbb{V}_1, \cdots, \mathbb{V}_n$ be vector spaces over a common scalar field $\mathbb{F}$, and $\boldsymbol{v}_j \in \mathbb{V}_j$ be given for $1 \leqslant j \leqslant n$. Then $\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_n \in \mathcal{L}(\mathbb{V}_1', \cdots, \mathbb{V}_n'; \mathbb{F})$.

Proof.

Let $1 \leqslant j \leqslant n$, $f_\ell \in \mathbb{V}_\ell'$ for $\ell \neq j$, g_j, $h_j \in \mathbb{V}_j'$ and $c \in \mathbb{F}$. Then

$$
\begin{aligned}
&(\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_n)(f_1, \cdots, f_{j-1}, cg_j + h_j, f_{j+1}, \cdots, f_n) \\
&\quad = f_1(\boldsymbol{v}_1) \cdots f_{j-1}(\boldsymbol{v}_{j-1}) \cdot \big[cg_j(\boldsymbol{v}_j) + h_j(\boldsymbol{v}_j)\big] \cdot f_{j+1}(\boldsymbol{v}_{j+1}) \cdots f_n(\boldsymbol{v}_n) \\
&\quad = c\, f_1(\boldsymbol{v}_1) \cdots f_{j-1}(\boldsymbol{v}_{j-1}) \cdot g_j(\boldsymbol{v}_j) \cdot f_{j+1}(\boldsymbol{v}_{j+1}) \cdots f_n(\boldsymbol{v}_n) \\
&\qquad + f_1(\boldsymbol{v}_1) \cdots f_{j-1}(\boldsymbol{v}_{j-1}) \cdot h_j(\boldsymbol{v}_j) \cdot f_{j+1}(\boldsymbol{v}_{j+1}) \cdots f_n(\boldsymbol{v}_n) \\
&\quad = c\,(\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_n)(f_1, \cdots, f_{j-1}, g_j, f_{j+1}, \cdots, f_n) \\
&\qquad + (\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_n)(f_1, \cdots, f_{j-1}, h_j, f_{j+1}, \cdots, f_n)
\end{aligned}
$$

which shows that $\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_n$ satisfies the multi-linearity. □

§3.6 Tensor Product of Vector Spaces

The fact that $\mathcal{L}(\mathbb{V}_1', \cdots, \mathbb{V}_n'; \mathbb{F})$ is a vector space over $\mathbb{F}$ motivates the definition of the tensor product of vector spaces.

Definition

Let $\mathbb{V}_1$, $\cdots$, $\mathbb{V}_n$ be vector spaces over a common scalar field $\mathbb{F}$. The tensor product space $\mathbb{V}_1 \otimes \cdots \otimes \mathbb{V}_n$ is the subspace of $\mathcal{L}(\mathbb{V}_1', \cdots, \mathbb{V}_n'; \mathbb{F})$ spanned by all finite linear combinations of tensor products $\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_n$, where $\boldsymbol{v}_j \in \mathbb{V}_j$ for $1 \leqslant j \leqslant n$ and $\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_n \in \mathcal{L}(\mathbb{V}_1', \cdots, \mathbb{V}_n'; \mathbb{F})$ is given by

$$\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_n : (f_1, ..., f_n) \mapsto \prod_{j=1}^{n} f_j(\boldsymbol{v}_j) \equiv f_1(\boldsymbol{v}_1) \cdot \cdots \cdot f_n(\boldsymbol{v}_n) .$$

§3.6 Tensor Product of Vector Spaces

Proposition

Let $\mathbb{U}, \mathbb{V}, \mathbb{W}$ *be vector spaces over a common scalar field* $\mathbb{F}$*. Then*

$$\mathbb{U} \otimes \mathbb{V} \otimes \mathbb{W} = (\mathbb{U} \otimes \mathbb{V}) \otimes \mathbb{W} = \mathbb{U} \otimes (\mathbb{V} \otimes \mathbb{W}) .$$

Proof.

The proposition follows from the previous proposition that

$$\boldsymbol{u} \otimes \boldsymbol{v} \otimes \boldsymbol{w} = (\boldsymbol{u} \otimes \boldsymbol{v}) \otimes \boldsymbol{w} = \boldsymbol{u} \otimes (\boldsymbol{v} \otimes \boldsymbol{w}) \quad \forall\, \boldsymbol{u} \in \mathbb{U}, \boldsymbol{v} \in \mathbb{V}, \boldsymbol{w} \in \mathbb{W}$$

and the definition of tensor product spaces. □

Proposition

Let $\mathbb{V}_1, \cdots, \mathbb{V}_n$ *be finite-dimensional vector spaces over a common scalar field* $\mathbb{F}$*. Then* $\mathbb{V}_1 \otimes \cdots \otimes \mathbb{V}_n$ *is finite-dimensional and*

$$\dim(\mathbb{V}_1 \otimes \cdots \otimes \mathbb{V}_n) = \prod_{j=1}^{n} \dim(\mathbb{V}_j) = \dim(\mathbb{V}_1) \cdot \cdots \cdot \dim(\mathbb{V}_n) .$$

§3.6 Tensor Product of Vector Spaces

Proposition

Let $\mathbb{U}, \mathbb{V}, \mathbb{W}$ *be vector spaces over a common scalar field* $\mathbb{F}$. *Then*

$$\mathbb{U} \otimes \mathbb{V} \otimes \mathbb{W} = (\mathbb{U} \otimes \mathbb{V}) \otimes \mathbb{W} = \mathbb{U} \otimes (\mathbb{V} \otimes \mathbb{W}) .$$

Proof.

The proposition follows from the previous proposition that

$$\boldsymbol{u} \otimes \boldsymbol{v} \otimes \boldsymbol{w} = (\boldsymbol{u} \otimes \boldsymbol{v}) \otimes \boldsymbol{w} = \boldsymbol{u} \otimes (\boldsymbol{v} \otimes \boldsymbol{w}) \quad \forall\, \boldsymbol{u} \in \mathbb{U}, \boldsymbol{v} \in \mathbb{V}, \boldsymbol{w} \in \mathbb{W}$$

and the definition of tensor product spaces. □

Proposition

Let $\mathbb{V}_1, \cdots, \mathbb{V}_n$ *be finite-dimensional vector spaces over a common scalar field* $\mathbb{F}$. *Then* $\mathbb{V}_1 \otimes \cdots \otimes \mathbb{V}_n$ *is finite-dimensional and*

$$\dim(\mathbb{V}_1 \otimes \cdots \otimes \mathbb{V}_n) = \prod_{j=1}^{n} \dim(\mathbb{V}_j) = \dim(\mathbb{V}_1) \cdot \cdots \cdot \dim(\mathbb{V}_n) .$$

§3.6 Tensor Product of Vector Spaces

Proof.

By previous proposition it suffices to show the case $n = 2$.

Let $\{\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n\}$ and $\{\tilde{\mathbf{e}}_1, \tilde{\mathbf{e}}_2, \cdots, \tilde{\mathbf{e}}_m\}$ be basis of $\mathbb{V}_1$ and $\mathbb{V}_2$, respectively. For $\boldsymbol{x} \in \mathbb{V}_1$ and $\boldsymbol{y} \in \mathbb{V}_2$, write $\boldsymbol{x} = \sum_{k=1}^{n} x_k \mathbf{e}_k$ and $\boldsymbol{y} = \sum_{\ell=1}^{m} y_\ell \tilde{\mathbf{e}}_\ell$. Then

$$\boldsymbol{x} \otimes \boldsymbol{y} = \Big(\sum_{k=1}^{n} x_k \mathbf{e}_k\Big) \otimes \Big(\sum_{\ell=1}^{m} y_\ell \tilde{\mathbf{e}}_\ell\Big) = \sum_{k=1}^{n} \sum_{\ell=1}^{m} x_k y_\ell (\mathbf{e}_k \otimes \tilde{\mathbf{e}}_\ell) .$$

Since vectors in $\mathbb{V}_1 \otimes \mathbb{V}_2$ can be expressed as a linear combination of vectors of the form $\boldsymbol{x} \otimes \boldsymbol{y}$, we find that every vectors in $\mathbb{V}_1 \otimes \mathbb{V}_2$ can be expressed as a linear combination of vectors from the set $\mathcal{B} \equiv \{\mathbf{e}_k \otimes \tilde{\mathbf{e}}_\ell \,|\, 1 \leqslant k \leqslant n, 1 \leqslant \ell \leqslant m\}$. Since $\#\mathcal{B} = nm$, it suffices to show that $\mathcal{B}$ is a linearly independent set. □

§3.6 Tensor Product of Vector Spaces

Proof.

By previous proposition it suffices to show the case $n = 2$.

Let $\{\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n\}$ and $\{\tilde{\mathbf{e}}_1, \tilde{\mathbf{e}}_2, \cdots, \tilde{\mathbf{e}}_m\}$ be basis of $\mathbb{V}_1$ and $\mathbb{V}_2$, respectively. For $\boldsymbol{x} \in \mathbb{V}_1$ and $\boldsymbol{y} \in \mathbb{V}_2$, write $\boldsymbol{x} = \sum_{k=1}^{n} x_k \mathbf{e}_k$ and $\boldsymbol{y} = \sum_{\ell=1}^{m} y_\ell \tilde{\mathbf{e}}_\ell$. Then

$$\boldsymbol{x} \otimes \boldsymbol{y} = \Big(\sum_{k=1}^{n} x_k \mathbf{e}_k\Big) \otimes \Big(\sum_{\ell=1}^{m} y_\ell \tilde{\mathbf{e}}_\ell\Big) = \sum_{k=1}^{n} \sum_{\ell=1}^{m} x_k y_\ell (\mathbf{e}_k \otimes \tilde{\mathbf{e}}_\ell) .$$

Since vectors in $\mathbb{V}_1 \otimes \mathbb{V}_2$ can be expressed as a linear combination of vectors of the form $\boldsymbol{x} \otimes \boldsymbol{y}$, we find that every vectors in $\mathbb{V}_1 \otimes \mathbb{V}_2$ can be expressed as a linear combination of vectors from the set $\mathcal{B} \equiv \big\{\mathbf{e}_k \otimes \tilde{\mathbf{e}}_\ell \,\big|\, 1 \leqslant k \leqslant n, 1 \leqslant \ell \leqslant m\big\}$. Since $\#\mathcal{B} = nm$, it suffices to show that $\mathcal{B}$ is a linearly independent set. □

§3.6 Tensor Product of Vector Spaces

Proof.

By previous proposition it suffices to show the case $n = 2$.
Let $\{\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n\}$ and $\{\widetilde{\mathbf{e}}_1, \widetilde{\mathbf{e}}_2, \cdots, \widetilde{\mathbf{e}}_m\}$ be basis of $\mathbb{V}_1$ and $\mathbb{V}_2$, respectively. For $\boldsymbol{x} \in \mathbb{V}_1$ and $\boldsymbol{y} \in \mathbb{V}_2$, write $\boldsymbol{x} = \sum_{k=1}^{n} x_k \mathbf{e}_k$ and $\boldsymbol{y} = \sum_{\ell=1}^{m} y_\ell \widetilde{\mathbf{e}}_\ell$. Then

$$\boldsymbol{x} \otimes \boldsymbol{y} = \Big(\sum_{k=1}^{n} x_k \mathbf{e}_k\Big) \otimes \Big(\sum_{\ell=1}^{m} y_\ell \widetilde{\mathbf{e}}_\ell\Big) = \sum_{k=1}^{n} \sum_{\ell=1}^{m} x_k y_\ell (\mathbf{e}_k \otimes \widetilde{\mathbf{e}}_\ell) .$$

Since vectors in $\mathbb{V}_1 \otimes \mathbb{V}_2$ can be expressed as a linear combination of vectors of the form $\boldsymbol{x} \otimes \boldsymbol{y}$, we find that every vectors in $\mathbb{V}_1 \otimes \mathbb{V}_2$ can be expressed as a linear combination of vectors from the set $\mathcal{B} \equiv \big\{\mathbf{e}_k \otimes \widetilde{\mathbf{e}}_\ell \,\big|\, 1 \leqslant k \leqslant n, 1 \leqslant \ell \leqslant m\big\}$. Since $\#\mathcal{B} = nm$, it suffices to show that $\mathcal{B}$ is a linearly independent set. □

§3.6 Tensor Product of Vector Spaces

Proof (cont.)

Let $\{c_{k\ell}\}_{1\leqslant k\leqslant n, 1\leqslant \ell\leqslant m}$ be a collection of scalars in $\mathbb{F}$ such that

$$\sum_{k=1}^{n}\sum_{\ell=1}^{m} c_{k\ell}\,\mathbf{e}_k \otimes \widetilde{\mathbf{e}}_\ell = 0 \quad \text{(the zero vector in } \mathbb{V}_1 \otimes \mathbb{V}_2\text{)}.$$

Let $f_i \in \mathbb{V}_1'$ and $g_j \in \mathbb{V}_2'$ satisfy

$$f_i(\mathbf{e}_k) = \delta_{ik} \qquad \text{and} \qquad g_j(\widetilde{\mathbf{e}}_\ell) = \delta_{j\ell}\,,$$

where $\delta_{\cdot\cdot}$ are the Kronecker delta. Then for each $1 \leqslant i \leqslant n$ and $1 \leqslant j \leqslant m$,

$$0 = \Big(\sum_{k=1}^{n}\sum_{\ell=1}^{m} c_{k\ell}\mathbf{e}_k \otimes \widetilde{\mathbf{e}}_\ell\Big)(f_i, g_j) = \sum_{k=1}^{n}\sum_{\ell=1}^{m} c_{k\ell}\delta_{ik}\delta_{j\ell} = c_{ij}\,.$$

This implies that $\mathcal{B}$ is a linearly independent set; thus $\dim(\mathbb{V}_1 \otimes \mathbb{V}_2) = \#\mathcal{B} = nm$. □

§3.6 Tensor Product of Vector Spaces

Proof (cont.)

Let $\{c_{k\ell}\}_{1\leqslant k\leqslant n, 1\leqslant \ell\leqslant m}$ be a collection of scalars in $\mathbb{F}$ such that

$$\sum_{k=1}^{n}\sum_{\ell=1}^{m} c_{k\ell}\,\mathbf{e}_k\otimes\widetilde{\mathbf{e}}_\ell = 0 \quad \text{(the zero vector in } \mathbb{V}_1\otimes\mathbb{V}_2\text{)}.$$

Let $f_i\in\mathbb{V}_1'$ and $g_j\in\mathbb{V}_2'$ satisfy

$$f_i(\mathbf{e}_k)=\delta_{ik} \qquad \text{and} \qquad g_j(\widetilde{\mathbf{e}}_\ell)=\delta_{j\ell}\,,$$

where $\delta_{\cdot\cdot}$ are the Kronecker delta. Then for each $1\leqslant i\leqslant n$ and $1\leqslant j\leqslant m$,

$$0=\Big(\sum_{k=1}^{n}\sum_{\ell=1}^{m} c_{k\ell}\mathbf{e}_k\otimes\widetilde{\mathbf{e}}_\ell\Big)(f_i,g_j)=\sum_{k=1}^{n}\sum_{\ell=1}^{m} c_{k\ell}\delta_{ik}\delta_{j\ell}=c_{ij}\,.$$

This implies that $\mathcal{B}$ is a linearly independent set; thus $\dim(\mathbb{V}_1\otimes\mathbb{V}_2)=\#\mathcal{B}=nm$. □

§3.6 Tensor Product of Vector Spaces

Next we consider the (matrix/coordinate) representation of tensor product of vectors. We start with the following

Example

Let $\mathbb{U}$ and $\mathbb{V}$ be vector spaces over a common scalar field $\mathbb{F}$, and $\mathcal{B}_{\mathbb{U}} = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$ and $\mathcal{B}_{\mathbb{V}} = \{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ be basis of $\mathbb{U}$ and $\mathbb{V}$. For $\boldsymbol{x} \in \mathbb{U}$ and $\boldsymbol{y} \in \mathbb{V}$, there exist unique $x_1, x_2, x_3, y_1, y_2 \in \mathbb{F}$ such that

$$\boldsymbol{x} = x_1\boldsymbol{u}_1 + x_2\boldsymbol{u}_2 + x_3\boldsymbol{u}_3 \quad \text{and} \quad \boldsymbol{y} = y_1\boldsymbol{v}_1 + y_2\boldsymbol{v}_2, .$$

By the properties of tensor product of vectors,

$$\boldsymbol{x} \otimes \boldsymbol{y} = \sum_{i=1}^{3}\sum_{j=1}^{2} x_i y_j (\boldsymbol{u}_i \otimes \boldsymbol{v}_j)$$

so that the coordinate of $\boldsymbol{x} \otimes \boldsymbol{y}$ w.r.t. the ordered basis $\mathcal{B} = \{\boldsymbol{u}_1 \otimes \boldsymbol{v}_1, \boldsymbol{u}_1 \otimes \boldsymbol{v}_2, \boldsymbol{u}_2 \otimes \boldsymbol{v}_1, \boldsymbol{u}_2 \otimes \boldsymbol{v}_2, \boldsymbol{u}_3 \otimes \boldsymbol{v}_1, \boldsymbol{u}_3 \otimes \boldsymbol{v}_2\}$ of $\mathbb{U} \otimes \mathbb{V}$ is given by $(x_1y_1, x_1y_2, x_2y_1, x_2y_2, x_3y_1, x_3y_2)$.

§3.6 Tensor Product of Vector Spaces

Next we consider the (matrix/coordinate) representation of tensor product of vectors. We start with the following

Example

Let $\mathbb{U}$ and $\mathbb{V}$ be vector spaces over a common scalar field $\mathbb{F}$, and $\mathcal{B}_{\mathbb{U}} = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$ and $\mathcal{B}_{\mathbb{V}} = \{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ be basis of $\mathbb{U}$ and $\mathbb{V}$. For $\boldsymbol{x} \in \mathbb{U}$ and $\boldsymbol{y} \in \mathbb{V}$, there exist unique $x_1, x_2, x_3, y_1, y_2 \in \mathbb{F}$ such that

$$\boldsymbol{x} = x_1\boldsymbol{u}_1 + x_2\boldsymbol{u}_2 + x_3\boldsymbol{u}_3 \quad \text{and} \quad \boldsymbol{y} = y_1\boldsymbol{v}_1 + y_2\boldsymbol{v}_2, .$$

By the properties of tensor product of vectors,

$$\boldsymbol{x} \otimes \boldsymbol{y} = \sum_{i=1}^{3}\sum_{j=1}^{2} x_i y_j (\boldsymbol{u}_i \otimes \boldsymbol{v}_j)$$

so that the coordinate of $\boldsymbol{x} \otimes \boldsymbol{y}$ w.r.t. the ordered basis $\mathcal{B} = \{\boldsymbol{u}_1 \otimes \boldsymbol{v}_1, \boldsymbol{u}_1 \otimes \boldsymbol{v}_2, \boldsymbol{u}_2 \otimes \boldsymbol{v}_1, \boldsymbol{u}_2 \otimes \boldsymbol{v}_2, \boldsymbol{u}_3 \otimes \boldsymbol{v}_1, \boldsymbol{u}_3 \otimes \boldsymbol{v}_2\}$ of $\mathbb{U} \otimes \mathbb{V}$ is given by $(x_1y_1, x_1y_2, x_2y_1, x_2y_2, x_3y_1, x_3y_2)$.

§3.6 Tensor Product of Vector Spaces

Next we consider the (matrix/coordinate) representation of tensor product of vectors. We start with the following

Example

Let $\mathbb{U}$ and $\mathbb{V}$ be vector spaces over a common scalar field $\mathbb{F}$, and $\mathcal{B}_{\mathbb{U}} = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$ and $\mathcal{B}_{\mathbb{V}} = \{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ be basis of $\mathbb{U}$ and $\mathbb{V}$. For $\boldsymbol{x} \in \mathbb{U}$ and $\boldsymbol{y} \in \mathbb{V}$, there exist unique $x_1, x_2, x_3, y_1, y_2 \in \mathbb{F}$ such that

$$\boldsymbol{x} = x_1\boldsymbol{u}_1 + x_2\boldsymbol{u}_2 + x_3\boldsymbol{u}_3 \quad \text{and} \quad \boldsymbol{y} = y_1\boldsymbol{v}_1 + y_2\boldsymbol{v}_2, .$$

By the properties of tensor product of vectors,

$$\boldsymbol{x} \otimes \boldsymbol{y} = \sum_{i=1}^{3}\sum_{j=1}^{2} x_i y_j (\boldsymbol{u}_i \otimes \boldsymbol{v}_j)$$

so that the coordinate of $\boldsymbol{x} \otimes \boldsymbol{y}$ w.r.t. the ordered basis $\mathcal{B} = \{\boldsymbol{u}_1 \otimes \boldsymbol{v}_1, \boldsymbol{u}_1 \otimes \boldsymbol{v}_2, \boldsymbol{u}_2 \otimes \boldsymbol{v}_1, \boldsymbol{u}_2 \otimes \boldsymbol{v}_2, \boldsymbol{u}_3 \otimes \boldsymbol{v}_1, \boldsymbol{u}_3 \otimes \boldsymbol{v}_2\}$ of $\mathbb{U} \otimes \mathbb{V}$ is given by $(x_1y_1, x_1y_2, x_2y_1, x_2y_2, x_3y_1, x_3y_2)$.

§3.6 Tensor Product of Vector Spaces

Example (cont.)

Writing the coordinate in terms of a column vector, we have

$$[\mathbf{x}\otimes\mathbf{y}]_{\mathcal{B}}=\begin{bmatrix}x_1y_1\\x_1y_2\\x_2y_1\\x_2y_2\\x_3y_1\\x_3y_2\end{bmatrix}=\begin{bmatrix}x_1\begin{bmatrix}y_1\\y_2\end{bmatrix}\\x_2\begin{bmatrix}y_1\\y_2\end{bmatrix}\\x_3\begin{bmatrix}y_1\\y_2\end{bmatrix}\end{bmatrix}\equiv\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}\otimes\begin{bmatrix}y_1\\y_2\end{bmatrix}=[\mathbf{x}]_{\mathcal{B}_{\mathbb{U}}}\otimes[\mathbf{y}]_{\mathcal{B}_{\mathbb{V}}}.$$

The ordered basis $\mathcal{B}$ is called the **induced basis** of the ordered basis $\mathcal{B}_{\mathbb{U}}$ and $\mathcal{B}_{\mathbb{V}}$.

Remark: For given basis of vector spaces, there are two induced basis, one for direct sum of spaces and one for tensor product of spaces. We will abuse the use of the word “induced” but keep in mind that it refers to one particular type.

§3.6 Tensor Product of Vector Spaces

Example (cont.)

Writing the coordinate in terms of a column vector, we have

$$[\mathbf{x}\otimes\mathbf{y}]_{\mathcal{B}}=\begin{bmatrix}x_1y_1\\x_1y_2\\x_2y_1\\x_2y_2\\x_3y_1\\x_3y_2\end{bmatrix}=\begin{bmatrix}x_1\begin{bmatrix}y_1\\y_2\end{bmatrix}\\x_2\begin{bmatrix}y_1\\y_2\end{bmatrix}\\x_3\begin{bmatrix}y_1\\y_2\end{bmatrix}\end{bmatrix}\equiv\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}\otimes\begin{bmatrix}y_1\\y_2\end{bmatrix}=[\mathbf{x}]_{\mathcal{B}_{\mathbb{U}}}\otimes[\mathbf{y}]_{\mathcal{B}_{\mathbb{V}}}.$$

The ordered basis $\mathcal{B}$ is called the **induced basis** of the ordered basis $\mathcal{B}_{\mathbb{U}}$ and $\mathcal{B}_{\mathbb{V}}$.

Remark: For given basis of vector spaces, there are two induced basis, one for direct sum of spaces and one for tensor product of spaces. We will abuse the use of the word “induced” but keep in mind that it refers to one particular type.

§3.6 Tensor Product of Vector Spaces

Example (cont.)

Writing the coordinate in terms of a column vector, we have

$$[\mathbf{x}\otimes\mathbf{y}]_{\mathcal{B}}=\begin{bmatrix}x_1y_1\\x_1y_2\\x_2y_1\\x_2y_2\\x_3y_1\\x_3y_2\end{bmatrix}=\begin{bmatrix}x_1\begin{bmatrix}y_1\\y_2\end{bmatrix}\\x_2\begin{bmatrix}y_1\\y_2\end{bmatrix}\\x_3\begin{bmatrix}y_1\\y_2\end{bmatrix}\end{bmatrix}\equiv\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}\otimes\begin{bmatrix}y_1\\y_2\end{bmatrix}=[\mathbf{x}]_{\mathcal{B}_{\mathbb{U}}}\otimes[\mathbf{y}]_{\mathcal{B}_{\mathbb{V}}}.$$

The ordered basis $\mathcal{B}$ is called the **induced basis** of the ordered basis $\mathcal{B}_{\mathbb{U}}$ and $\mathcal{B}_{\mathbb{V}}$.

Remark: For given basis of vector spaces, there are two induced basis, one for direct sum of spaces and one for tensor product of spaces. We will abuse the use of the word “induced” but keep in mind that it refers to one particular type.

§3.6 Tensor Product of Vector Spaces

Example (cont.)

Writing the coordinate in terms of a column vector, we have

$$[\boldsymbol{x}\otimes\boldsymbol{y}]_{\mathcal{B}}=\begin{bmatrix}x_1y_1\\x_1y_2\\x_2y_1\\x_2y_2\\x_3y_1\\x_3y_2\end{bmatrix}=\begin{bmatrix}x_1\begin{bmatrix}y_1\\y_2\end{bmatrix}\\x_2\begin{bmatrix}y_1\\y_2\end{bmatrix}\\x_3\begin{bmatrix}y_1\\y_2\end{bmatrix}\end{bmatrix}\equiv\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}\otimes\begin{bmatrix}y_1\\y_2\end{bmatrix}=[\boldsymbol{x}]_{\mathcal{B}_{\mathbb{U}}}\otimes[\boldsymbol{y}]_{\mathcal{B}_{\mathbb{V}}}.$$

The ordered basis $\mathcal{B}$ is called the **induced basis** of the ordered basis $\mathcal{B}_{\mathbb{U}}$ and $\mathcal{B}_{\mathbb{V}}$.

Remark: For given basis of vector spaces, there are two induced basis, one for direct sum of spaces and one for tensor product of spaces. We will abuse the use of the word “induced” but keep in mind that it refers to one particular type.

§3.6 Tensor Product of Vector Spaces

Example (cont.)

Writing the coordinate in terms of a column vector, we have

$$[\boldsymbol{x} \otimes \boldsymbol{y}]_{\mathcal{B}} = \begin{bmatrix} x_1 y_1 \\ x_1 y_2 \\ x_2 y_1 \\ x_2 y_2 \\ x_3 y_1 \\ x_3 y_2 \end{bmatrix} = \begin{bmatrix} x_1 \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \\ x_2 \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \\ x_3 \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \end{bmatrix} \equiv \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \otimes \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = [\boldsymbol{x}]_{\mathcal{B}_{\mathbb{U}}} \otimes [\boldsymbol{y}]_{\mathcal{B}_{\mathbb{V}}} .$$

The ordered basis $\mathcal{B}$ is called the **induced basis** of the ordered basis $\mathcal{B}_{\mathbb{U}}$ and $\mathcal{B}_{\mathbb{V}}$.

Remark: For given basis of vector spaces, there are two induced basis, one for direct sum of spaces and one for tensor product of spaces. We will abuse the use of the word “induced” but keep in mind that it refers to one particular type.

§3.6 Tensor Product of Vector Spaces

Example (cont.)

Writing the coordinate in terms of a column vector, we have

$$[\boldsymbol{x}\otimes\boldsymbol{y}]_{\mathcal{B}}=\begin{bmatrix}x_1y_1\\x_1y_2\\x_2y_1\\x_2y_2\\x_3y_1\\x_3y_2\end{bmatrix}=\begin{bmatrix}x_1\begin{bmatrix}y_1\\y_2\end{bmatrix}\\x_2\begin{bmatrix}y_1\\y_2\end{bmatrix}\\x_3\begin{bmatrix}y_1\\y_2\end{bmatrix}\end{bmatrix}\equiv\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}\otimes\begin{bmatrix}y_1\\y_2\end{bmatrix}=[\boldsymbol{x}]_{\mathcal{B}_{\mathbb{U}}}\otimes[\boldsymbol{y}]_{\mathcal{B}_{\mathbb{V}}}.$$

The ordered basis $\mathcal{B}$ is called the **induced basis** of the ordered basis $\mathcal{B}_{\mathbb{U}}$ and $\mathcal{B}_{\mathbb{V}}$.

Remark: For given basis of vector spaces, there are two induced basis, one for direct sum of spaces and one for tensor product of spaces. We will abuse the use of the word “induced” but keep in mind that it refers to one particular type.

§3.6 Tensor Product of Vector Spaces

The example above motivates the formal/computational definition of the tensor product of vectors in $\mathbb{C}^m$ and $\mathbb{C}^n$ as follows. Let $\boldsymbol{x} = [x_1, \cdots, x_m]^{\mathrm{T}} \in \mathbb{C}^m$ and $\boldsymbol{y} = [y_1, \cdots, y_n]^{\mathrm{T}} \in \mathbb{C}^n$. The "tensor product" of $\boldsymbol{x}$ and $\boldsymbol{y}$, denoted by $[\boldsymbol{x} \otimes \boldsymbol{y}]$, is a vector in $\mathbb{C}^{mn}$ given by

$$[\boldsymbol{x} \otimes \boldsymbol{y}] = \begin{bmatrix} x_1 y_1 \\ \vdots \\ x_1 y_n \\ \vdots \\ x_m y_1 \\ \vdots \\ x_m y_n \end{bmatrix} = \begin{bmatrix} x_1 \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \\ \vdots \\ x_m \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \end{bmatrix} \equiv \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} \otimes \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = [\boldsymbol{x}] \otimes [\boldsymbol{y}].$$

In fact, $[\boldsymbol{x} \otimes \boldsymbol{y}] \in \mathbb{C}^{mn}$ is indeed the coordinate of $\boldsymbol{x} \otimes \boldsymbol{y}$ w.r.t. the induced ordered basis $\{\mathrm{e}_1 \otimes \tilde{\mathrm{e}}_1, \mathrm{e}_1 \otimes \tilde{\mathrm{e}}_2, \cdots, \mathrm{e}_1 \otimes \tilde{\mathrm{e}}_n, \mathrm{e}_2 \otimes \tilde{\mathrm{e}}_1, \mathrm{e}_2 \otimes \tilde{\mathrm{e}}_2, \cdots, \mathrm{e}_2 \otimes \tilde{\mathrm{e}}_n, \cdots, \mathrm{e}_m \otimes \tilde{\mathrm{e}}_1, \mathrm{e}_m \otimes \tilde{\mathrm{e}}_2, \cdots, \mathrm{e}_m \otimes \tilde{\mathrm{e}}_n\}$ of $\mathbb{C}^m \otimes \mathbb{C}^n$.

§3.6 Tensor Product of Vector Spaces

The example above motivates the formal/computational definition of the tensor product of vectors in $\mathbb{C}^m$ and $\mathbb{C}^n$ as follows. Let $\boldsymbol{x} = [x_1, \cdots, x_m]^{\mathrm{T}} \in \mathbb{C}^m$ and $\boldsymbol{y} = [y_1, \cdots, y_n]^{\mathrm{T}} \in \mathbb{C}^n$. The "tensor product" of $\boldsymbol{x}$ and $\boldsymbol{y}$, denoted by $[\boldsymbol{x} \otimes \boldsymbol{y}]$, is a vector in $\mathbb{C}^{mn}$ given by

$$[\boldsymbol{x} \otimes \boldsymbol{y}] = \begin{bmatrix} x_1 y_1 \\ \vdots \\ x_1 y_n \\ \vdots \\ x_m y_1 \\ \vdots \\ x_m y_n \end{bmatrix} = \begin{bmatrix} x_1 \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \\ \vdots \\ x_m \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \end{bmatrix} \equiv \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} \otimes \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = [\boldsymbol{x}] \otimes [\boldsymbol{y}].$$

In fact, $[\boldsymbol{x} \otimes \boldsymbol{y}] \in \mathbb{C}^{mn}$ is indeed the coordinate of $\boldsymbol{x} \otimes \boldsymbol{y}$ w.r.t. the induced ordered basis $\{\mathbf{e}_1 \otimes \tilde{\mathbf{e}}_1, \mathbf{e}_1 \otimes \tilde{\mathbf{e}}_2, \cdots, \mathbf{e}_1 \otimes \tilde{\mathbf{e}}_n, \mathbf{e}_2 \otimes \tilde{\mathbf{e}}_1, \mathbf{e}_2 \otimes \tilde{\mathbf{e}}_2, \cdots, \mathbf{e}_2 \otimes \tilde{\mathbf{e}}_n, \cdots, \mathbf{e}_m \otimes \tilde{\mathbf{e}}_1, \mathbf{e}_m \otimes \tilde{\mathbf{e}}_2, \cdots, \mathbf{e}_m \otimes \tilde{\mathbf{e}}_n\}$ of $\mathbb{C}^m \otimes \mathbb{C}^n$.

§3.6 Tensor Product of Vector Spaces

Suppose that X, Y, Z, W are vector spaces over a common scalar field $\mathbb{F}$. Then $\mathcal{L}(X,Z)$ and $\mathcal{L}(Y,W)$ are also vector spaces so that $\mathcal{L}(X,Z)\otimes\mathcal{L}(Y,W)$ is a well-defined concept. In the following, we talk about an **alternative definition** of $A\otimes B$ if $A\in\mathcal{L}(X,Z)$ and $B\in\mathcal{L}(Y,W)$.

Definition

Let X, Y, Z, W be vector spaces over a common scalar field $\mathbb{F}$, and $A\in\mathcal{L}(X,Z)$, $B\in\mathcal{L}(Y,W)$. The tensor product of A and B, denoted by $A\otimes B$, is an element in $\mathcal{L}(X\otimes Y, Z\otimes W)$ satisfying

$$(A\otimes B)(\boldsymbol{x}\otimes\boldsymbol{y}) = (A\boldsymbol{x})\otimes(B\boldsymbol{y}) \qquad \forall\, \boldsymbol{x}\in X \text{ and } \boldsymbol{y}\in Y.$$

Remark: To avoid confusion, instead of treating $A\otimes B$ as the tensor product of A and B one can also treat $A\otimes B$ as a "**new operation**" between A and B (but with the same notation).

§3.6 Tensor Product of Vector Spaces

Suppose that X, Y, Z, W are vector spaces over a common scalar field $\mathbb{F}$. Then $\mathscr{L}(X, Z)$ and $\mathscr{L}(Y, W)$ are also vector spaces so that $\mathscr{L}(X, Z) \otimes \mathscr{L}(Y, W)$ is a well-defined concept. In the following, we talk about an **alternative definition** of $A \otimes B$ if $A \in \mathscr{L}(X, Z)$ and $B \in \mathscr{L}(Y, W)$.

Definition

Let X, Y, Z, W be vector spaces over a common scalar field $\mathbb{F}$, and $A \in \mathscr{L}(X, Z)$, $B \in \mathscr{L}(Y, W)$. The tensor product of A and B, denoted by $A \otimes B$, is an element in $\mathscr{L}(X \otimes Y, Z \otimes W)$ satisfying

$$(A \otimes B)(\boldsymbol{x} \otimes \boldsymbol{y}) = (A\boldsymbol{x}) \otimes (B\boldsymbol{y}) \qquad \forall\, \boldsymbol{x} \in X \text{ and } \boldsymbol{y} \in Y .$$

Remark: To avoid confusion, instead of treating $A \otimes B$ as the tensor product of A and B one can also treat $A \otimes B$ as a "**new operation**" between A and B (but with the same notation).

§3.6 Tensor Product of Vector Spaces

Suppose that X, Y, Z, W are vector spaces over a common scalar field $\mathbb{F}$. Then $\mathcal{L}(X,Z)$ and $\mathcal{L}(Y,W)$ are also vector spaces so that $\mathcal{L}(X,Z)\otimes\mathcal{L}(Y,W)$ is a well-defined concept. In the following, we talk about an **alternative definition** of $A\otimes B$ if $A\in\mathcal{L}(X,Z)$ and $B\in\mathcal{L}(Y,W)$.

Definition

Let X, Y, Z, W be vector spaces over a common scalar field $\mathbb{F}$, and $A\in\mathcal{L}(X,Z)$, $B\in\mathcal{L}(Y,W)$. The tensor product of A and B, denoted by $A\otimes B$, is an element in $\mathcal{L}(X\otimes Y, Z\otimes W)$ satisfying

$$(A\otimes B)(\boldsymbol{x}\otimes\boldsymbol{y}) = (A\boldsymbol{x})\otimes(B\boldsymbol{y}) \qquad \forall\, \boldsymbol{x}\in X \text{ and } \boldsymbol{y}\in Y.$$

Remark: To avoid confusion, instead of treating $A\otimes B$ as the tensor product of A and B one can also treat $A\otimes B$ as a "**new operation**" between A and B (but with the same notation).

§3.6 Tensor Product of Vector Spaces

Proposition

Let A, B, C be linear maps on vector spaces X, Y, Z (over a common scalar field $\mathbb{F}$). Then $(A \otimes B) \otimes C = A \otimes (B \otimes C)$.

- **Matrix representation of tensor product of linear maps**

Suppose that X, Y, Z, W are finite dimensional vector spaces over a common scalar field $\mathbb{F}$, and $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$, $\{\boldsymbol{y}_1, \cdots, \boldsymbol{y}_\ell\}$, $\{\boldsymbol{z}_1, \cdots, \boldsymbol{z}_m\}$ and $\{\boldsymbol{w}_1, \cdots, \boldsymbol{w}_k\}$ are basis of X, Y, Z, W, respectively. Let $A \in \mathcal{L}(X, Z)$, $B \in \mathcal{L}(Y, W)$, and the matrix representations of A and B be $[A] = [a_{ij}]_{1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n}$ and $[B] = [b_{ij}]_{1 \leqslant i \leqslant k, 1 \leqslant j \leqslant \ell}$, respectively, so that

$$A(c_1\boldsymbol{x}_1 + \cdots + c_n\boldsymbol{x}_n) = \sum_{i=1}^{m}\Big(\sum_{r=1}^{n} a_{ir}c_r\Big)\boldsymbol{z}_i$$

and

$$B(d_1\boldsymbol{y}_1 + \cdots + d_\ell\boldsymbol{y}_\ell) = \sum_{j=1}^{k}\Big(\sum_{s=1}^{\ell} b_{js}d_s\Big)\boldsymbol{w}_j\,.$$

§3.6 Tensor Product of Vector Spaces

Proposition

Let A, B, C be linear maps on vector spaces X, Y, Z (over a common scalar field $\mathbb{F}$). Then $(A \otimes B) \otimes C = A \otimes (B \otimes C)$.

- **Matrix representation of tensor product of linear maps**

Suppose that X, Y, Z, W are finite dimensional vector spaces over a common scalar field $\mathbb{F}$, and $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$, $\{\boldsymbol{y}_1, \cdots, \boldsymbol{y}_\ell\}$, $\{\boldsymbol{z}_1, \cdots, \boldsymbol{z}_m\}$ and $\{\boldsymbol{w}_1, \cdots, \boldsymbol{w}_k\}$ are basis of X, Y, Z, W, respectively. Let $A \in \mathcal{L}(X, Z)$, $B \in \mathcal{L}(Y, W)$, and the matrix representations of A and B be $[A] = [a_{ij}]_{1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n}$ and $[B] = [b_{ij}]_{1 \leqslant i \leqslant k, 1 \leqslant j \leqslant \ell}$, respectively, so that

$$A(c_1 \boldsymbol{x}_1 + \cdots + c_n \boldsymbol{x}_n) = \sum_{i=1}^{m} \Big(\sum_{r=1}^{n} a_{ir} c_r \Big) \boldsymbol{z}_i$$

and

$$B(d_1 \boldsymbol{y}_1 + \cdots + d_\ell \boldsymbol{y}_\ell) = \sum_{j=1}^{k} \Big(\sum_{s=1}^{\ell} b_{js} d_s \Big) \boldsymbol{w}_j .$$

§3.6 Tensor Product of Vector Spaces

Proposition

Let A, B, C be linear maps on vector spaces X, Y, Z (over a common scalar field $\mathbb{F}$). Then $(A \otimes B) \otimes C = A \otimes (B \otimes C)$.

- **Matrix representation of tensor product of linear maps**

Suppose that X, Y, Z, W are finite dimensional vector spaces over a common scalar field $\mathbb{F}$, and $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$, $\{\boldsymbol{y}_1, \cdots, \boldsymbol{y}_\ell\}$, $\{\boldsymbol{z}_1, \cdots, \boldsymbol{z}_m\}$ and $\{\boldsymbol{w}_1, \cdots, \boldsymbol{w}_k\}$ are basis of X, Y, Z, W, respectively. Let $A \in \mathcal{L}(X, Z)$, $B \in \mathcal{L}(Y, W)$, and the matrix representations of A and B be $[A] = [a_{ij}]_{1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n}$ and $[B] = [b_{ij}]_{1 \leqslant i \leqslant k, 1 \leqslant j \leqslant \ell}$, respectively, so that

$$A(c_1\boldsymbol{x}_1 + \cdots + c_n\boldsymbol{x}_n) = \sum_{i=1}^{m}\Big(\sum_{r=1}^{n} a_{ir}c_r\Big)\boldsymbol{z}_i$$

and

$$B(d_1\boldsymbol{y}_1 + \cdots + d_\ell\boldsymbol{y}_\ell) = \sum_{j=1}^{k}\Big(\sum_{s=1}^{\ell} b_{js}d_s\Big)\boldsymbol{w}_j\,.$$

§3.6 Tensor Product of Vector Spaces

Then

$$\begin{aligned}
(A \otimes B)\big((c_1 \mathbf{x}_1 + \cdots + c_n \mathbf{x}_n) \otimes (d_1 \mathbf{y}_1 + \cdots + d_\ell \mathbf{y}_\ell)\big) \\
&\equiv \big[A(c_1 \mathbf{x}_1 + \cdots + c_n \mathbf{x}_n)\big] \otimes \big[B(d_1 \mathbf{y}_1 + \cdots + d_\ell \mathbf{y}_\ell)\big] \\
&= \Big[\sum_{i=1}^{m} \Big(\sum_{r=1}^{n} a_{ir} c_r\Big) \mathbf{z}_i\Big] \otimes \Big[\sum_{j=1}^{k} \Big(\sum_{s=1}^{\ell} b_{js} d_s\Big) \mathbf{w}_j\Big] \\
&= \sum_{i=1}^{m} \sum_{j=1}^{k} \Big(\sum_{r=1}^{n} \sum_{s=1}^{\ell} a_{ir} b_{js} c_r d_s\Big) \mathbf{z}_i \otimes \mathbf{w}_j \,.
\end{aligned}$$

Since the induced ordered basis of $X \otimes Y$ and $Z \otimes W$ are given respectively by

$$\begin{aligned}
\mathcal{B}_{X \otimes Y} &= \big\{\mathbf{x}_1 \otimes \mathbf{y}_1, \cdots, \mathbf{x}_1 \otimes \mathbf{y}_\ell, \mathbf{x}_2 \otimes \mathbf{y}_1, \cdots, \mathbf{x}_2 \otimes \mathbf{y}_\ell, \cdots, \\
&\qquad \mathbf{x}_n \otimes \mathbf{y}_1, \cdots, \mathbf{x}_n \otimes \mathbf{y}_\ell\big\}, \\
\mathcal{B}_{Z \otimes W} &= \big\{\mathbf{z}_1 \otimes \mathbf{w}_1, \cdots, \mathbf{z}_1 \otimes \mathbf{w}_k, \mathbf{z}_2 \otimes \mathbf{w}_1, \cdots, \mathbf{z}_2 \otimes \mathbf{w}_k, \cdots, \\
&\qquad \mathbf{z}_m \otimes \mathbf{w}_1, \cdots, \mathbf{z}_m \otimes \mathbf{w}_k\big\},
\end{aligned}$$

§3.6 Tensor Product of Vector Spaces

the matrix representation of $A \otimes B$ satisfies

$$[A \otimes B] \begin{bmatrix} c_1 d_1 \\ \vdots \\ c_1 d_\ell \\ c_2 d_1 \\ \vdots \\ c_2 d_\ell \\ \vdots \\ c_n d_1 \\ \vdots \\ c_n d_\ell \end{bmatrix}_{n\ell \times 1} = \begin{bmatrix} \sum_{r=1}^{n} \sum_{s=1}^{\ell} a_{1r} b_{1s} c_r d_s \\ \vdots \\ \sum_{r=1}^{n} \sum_{s=1}^{\ell} a_{1r} b_{ks} c_r d_s \\ \sum_{r=1}^{n} \sum_{s=1}^{\ell} a_{2r} b_{1s} c_r d_s \\ \vdots \\ \sum_{r=1}^{n} \sum_{s=1}^{\ell} a_{2r} b_{ks} c_r d_s \\ \vdots \\ \sum_{r=1}^{n} \sum_{s=1}^{\ell} a_{mr} b_{1s} c_r d_s \\ \vdots \\ \sum_{r=1}^{n} \sum_{s=1}^{\ell} a_{mr} b_{ks} c_r d_s \end{bmatrix}_{mk \times 1}$$

for all $c_1, \cdots, c_n$ and $d_1, \cdots, d_\ell$ in $\mathbb{F}$.

§3.6 Tensor Product of Vector Spaces

Let $c_1 = d_j = 1$, where $1 \leqslant j \leqslant \ell$, and $c_r = d_s = 0$ for other r, s, we find that the j-th column of $[A \otimes B]$ is given by

$$[A \otimes B](:, j) = \begin{bmatrix} a_{11} b_{1j} \\ \vdots \\ a_{11} b_{kj} \\ a_{21} b_{1j} \\ \vdots \\ a_{21} b_{kj} \\ \vdots \\ a_{m1} b_{1j} \\ \vdots \\ a_{m1} b_{kj} \end{bmatrix} = \begin{bmatrix} a_{11} \begin{bmatrix} b_{1j} \\ \vdots \\ b_{kj} \end{bmatrix} \\ a_{21} \begin{bmatrix} b_{1j} \\ \vdots \\ b_{kj} \end{bmatrix} \\ \vdots \\ a_{m1} \begin{bmatrix} b_{1j} \\ \vdots \\ b_{kj} \end{bmatrix} \end{bmatrix} = \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} \otimes \begin{bmatrix} b_{1j} \\ \vdots \\ b_{kj} \end{bmatrix}$$

§3.6 Tensor Product of Vector Spaces

so that the first ℓ columns of $[A \otimes B]$ are given by

$$[A \otimes B](:, 1 : \ell) = \begin{bmatrix} a_{11}[B] \\ a_{21}[B] \\ \vdots \\ a_{m1}[B] \end{bmatrix} \equiv \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} \otimes [B].$$

§3.6 Tensor Product of Vector Spaces

Let $c_2 = d_j = 1$, where $1 \leqslant j \leqslant \ell$, and $c_r = d_s = 0$ for other r, s, we find that the $(\ell + j)$-th column of $[A \otimes B]$ is given by

$$[A \otimes B](:, \ell + j) = \begin{bmatrix} a_{12} b_{1j} \\ \vdots \\ a_{12} b_{kj} \\ a_{22} b_{1j} \\ \vdots \\ a_{22} b_{kj} \\ \vdots \\ a_{m2} b_{1j} \\ \vdots \\ a_{m2} b_{kj} \end{bmatrix} = \begin{bmatrix} a_{12} \begin{bmatrix} b_{1j} \\ \vdots \\ b_{kj} \end{bmatrix} \\ a_{22} \begin{bmatrix} b_{1j} \\ \vdots \\ b_{kj} \end{bmatrix} \\ \vdots \\ a_{m2} \begin{bmatrix} b_{1j} \\ \vdots \\ b_{kj} \end{bmatrix} \end{bmatrix} = \begin{bmatrix} a_{12} \\ \vdots \\ a_{m2} \end{bmatrix} \otimes \begin{bmatrix} b_{1j} \\ \vdots \\ b_{kj} \end{bmatrix}$$

§3.6 Tensor Product of Vector Spaces

so that the $(\ell+1)$-th to 2ℓ-th columns of $[A\otimes B]$ are given by

$$[A\otimes B](:,\ell+1:2\ell)=\begin{bmatrix} a_{12}[B] \\ a_{22}[B] \\ \vdots \\ a_{m2}[B] \end{bmatrix} \equiv \begin{bmatrix} a_{12} \\ \vdots \\ a_{m2} \end{bmatrix} \otimes [B].$$

In general, we can find that the $(p-1)\ell+j$ column of $[A\otimes B]$ by letting $c_p=d_j=1$ and $c_r=d_s=0$ for other r,s and obtain that the matrix representation of $A\otimes B$ is then given by

$$[A\otimes B]=\begin{bmatrix} a_{11}[B] & a_{12}[B] & \cdots & a_{1n}[B] \\ a_{21}[B] & a_{22}[B] & \cdots & a_{2n}[B] \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}[B] & a_{m2}[B] & \cdots & a_{mn}[B] \end{bmatrix} \equiv [A]\otimes[B].$$

§3.6 Tensor Product of Vector Spaces

so that the $(\ell+1)$-th to 2ℓ-th columns of $[A \otimes B]$ are given by

$$[A \otimes B](:, \ell+1 : 2\ell) = \begin{bmatrix} a_{12}[B] \\ a_{22}[B] \\ \vdots \\ a_{m2}[B] \end{bmatrix} \equiv \begin{bmatrix} a_{12} \\ \vdots \\ a_{m2} \end{bmatrix} \otimes [B].$$

In general, we can find that the $(p-1)\ell + j$ column of $[A \otimes B]$ by letting $c_p = d_j = 1$ and $c_r = d_s = 0$ for other r, s and obtain that the matrix representation of $A \otimes B$ is then given by

$$[A \otimes B] = \begin{bmatrix} a_{11}[B] & a_{12}[B] & \cdots & a_{1n}[B] \\ a_{21}[B] & a_{22}[B] & \cdots & a_{2n}[B] \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}[B] & a_{m2}[B] & \cdots & a_{mn}[B] \end{bmatrix} \equiv [A] \otimes [B] .$$

§3.6 Tensor Product of Vector Spaces

Definition

Let $A \in \mathcal{M}(m, n)$ and $B \in \mathcal{M}(k, \ell)$ The tensor product of A and B, denoted by $A \otimes B$, is an $(mk) \times (n\ell)$ matrix given by

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{bmatrix}.$$

Remark: In matlab®, the tensor product $A \otimes B$ of two matrices A and B is given by

$$A \otimes B = \mathbf{kron}(A, B)\,.$$

§3.6 Tensor Product of Vector Spaces

- **Tensor product of Hilbert spaces**

Consider a finite number of (complex) Hilbert spaces $\mathbb{H}_1, \cdots, \mathbb{H}_n$ with respective Hermitian scalar products $\langle \cdot, \cdot \rangle_1, \cdots, \langle \cdot, \cdot \rangle_n$. Relying upon the above definition, we can first define their algebraic tensor product $\mathbb{H}_1 \otimes \cdots \otimes \mathbb{H}_n$. This is not a Hilbert space yet. However it is possible (not so easy) to prove that $\mathbb{H}_1 \otimes \cdots \otimes \mathbb{H}_n$ admits an Hermitian scalar product induced by the ones of each $\mathbb{H}_j$. This scalar product $\langle \cdot, \cdot \rangle$ is the unique right-linear and left-antilinear extension of

$$\langle \boldsymbol{u}_1 \otimes \cdots \otimes \boldsymbol{u}_n, \boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_n \rangle = \prod_{j=1}^{n} \langle \boldsymbol{u}_j, \boldsymbol{v}_j \rangle_j \quad \text{if} \quad \boldsymbol{u}_j, \boldsymbol{v}_j \in \mathbb{H}_j. \tag{4}$$

The (anti)linear extension is necessary because $\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_n$ is not the generic element of $\mathbb{H}_1 \otimes \cdots \otimes \mathbb{H}_n$, the generic element is a finite linear combination of these elements!

§3.6 Tensor Product of Vector Spaces

- **Tensor product of Hilbert spaces**

Consider a finite number of (complex) Hilbert spaces $\mathbb{H}_1, \cdots, \mathbb{H}_n$ with respective Hermitian scalar products $\langle\cdot,\cdot\rangle_1, \cdots, \langle\cdot,\cdot\rangle_n$. Relying upon the above definition, we can first define their algebraic tensor product $\mathbb{H}_1 \otimes \cdots \otimes \mathbb{H}_n$. This is not a Hilbert space yet. However it is possible (not so easy) to prove that $\mathbb{H}_1 \otimes \cdots \otimes \mathbb{H}_n$ admits an Hermitian scalar product induced by the ones of each $\mathbb{H}_j$. This scalar product $\langle\cdot,\cdot\rangle$ is the unique right-linear and left-antilinear extension of

$$\langle \boldsymbol{u}_1 \otimes \cdots \otimes \boldsymbol{u}_n, \boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_n\rangle = \prod_{j=1}^{n} \langle \boldsymbol{u}_j, \boldsymbol{v}_j\rangle_j \quad \text{if} \quad \boldsymbol{u}_j, \boldsymbol{v}_j \in \mathbb{H}_j. \qquad (4)$$

The (anti)linear extension is necessary because $\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_n$ is not the generic element of $\mathbb{H}_1 \otimes \cdots \otimes \mathbb{H}_n$, the generic element is a finite linear combination of these elements!

§3.6 Tensor Product of Vector Spaces

- **Tensor product of Hilbert spaces**

Consider a finite number of (complex) Hilbert spaces $\mathbb{H}_1, \cdots, \mathbb{H}_n$ with respective Hermitian scalar products $\langle\cdot,\cdot\rangle_1, \cdots, \langle\cdot,\cdot\rangle_n$. Relying upon the above definition, we can first define their algebraic tensor product $\mathbb{H}_1 \otimes \cdots \otimes \mathbb{H}_n$. This is not a Hilbert space yet. However it is possible (not so easy) to prove that $\mathbb{H}_1 \otimes \cdots \otimes \mathbb{H}_n$ admits an Hermitian scalar product induced by the ones of each $\mathbb{H}_j$. This scalar product $\langle\cdot,\cdot\rangle$ is the unique right-linear and left-antilinear extension of

$$\langle \boldsymbol{u}_1 \otimes \cdots \otimes \boldsymbol{u}_n, \boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_n \rangle \equiv \prod_{j=1}^{n} \langle \boldsymbol{u}_j, \boldsymbol{v}_j \rangle_j \quad \text{if} \quad \boldsymbol{u}_j, \boldsymbol{v}_j \in \mathbb{H}_j. \qquad (4)$$

The (anti)linear extension is necessary because $\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_n$ is not the generic element of $\mathbb{H}_1 \otimes \cdots \otimes \mathbb{H}_n$, the generic element is a finite linear combination of these elements!

§3.6 Tensor Product of Vector Spaces

Definition

The Hilbertian tensor product of (complex) Hilbert spaces $(\mathbb{H}_1, \langle\cdot,\cdot\rangle_1)$, $\cdots$, $(\mathbb{H}_n, \langle\cdot,\cdot\rangle_n)$ is the (complex) Hilbert space $\mathbb{H}_1 \otimes \cdots \otimes \mathbb{H}_n$ given as the **completion** of the algebraic tensor product $\mathbb{H}_1 \otimes \cdots \otimes \mathbb{H}_n$ with respect to the Hermitian scalar product $\langle\cdot,\cdot\rangle$ which uniquely (anti)linearly extends

$$\langle \boldsymbol{u}_1 \otimes \cdots \otimes \boldsymbol{u}_n, \boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_n \rangle \equiv \prod_{j=1}^{n} \langle \boldsymbol{u}_j, \boldsymbol{v}_j \rangle_j \quad \text{if} \quad \boldsymbol{u}_j, \boldsymbol{v}_j \in \mathbb{H}_j. \tag{4}$$

Remark: The Hilbert spaces $\mathbb{H}_j$ in quantum computing are $\mathbb{C}^2$ for all $1 \leqslant j \leqslant n$, so the tensor product spaces $\mathbb{H}_1 \otimes \cdots \otimes \mathbb{H}_n$ along with the norm induced by the inner product defined by (4) is again a Hilbert space.

§3.6 Tensor Product of Vector Spaces

Definition

The Hilbertian tensor product of (complex) Hilbert spaces $(\mathbb{H}_1, \langle\cdot,\cdot\rangle_1)$, $\cdots$, $(\mathbb{H}_n, \langle\cdot,\cdot\rangle_n)$ is the (complex) Hilbert space $\mathbb{H}_1 \otimes \cdots \otimes \mathbb{H}_n$ given as the **completion** of the algebraic tensor product $\mathbb{H}_1 \otimes \cdots \otimes \mathbb{H}_n$ with respect to the Hermitian scalar product $\langle\cdot,\cdot\rangle$ which uniquely (anti)linearly extends

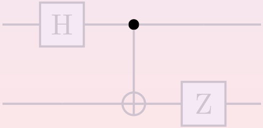
$$\langle \boldsymbol{u}_1 \otimes \cdots \otimes \boldsymbol{u}_n, \boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_n \rangle \equiv \prod_{j=1}^{n} \langle \boldsymbol{u}_j, \boldsymbol{v}_j \rangle_j \quad \text{if} \quad \boldsymbol{u}_j, \boldsymbol{v}_j \in \mathbb{H}_j. \qquad (4)$$

Remark: The Hilbert spaces $\mathbb{H}_j$ in quantum computing are $\mathbb{C}^2$ for all $1 \leqslant j \leqslant n$, so the tensor product spaces $\mathbb{H}_1 \otimes \cdots \otimes \mathbb{H}_n$ along with the norm induced by the inner product defined by (4) is again a Hilbert space.

§3.6 Tensor Product of Vector Spaces

§3.6.2 Correspondence b/w tensor product & quantum circuits

The tensor product of quantum gates represents a unitary transformation when these quantum gates are applied in parallel (at the same time), while the ordinary product of quantum gates represents a unitary transformation when these quantum gates are applied sequentially. Using the matrix representation of quantum gates, there is a way to understand the overall effect of all quantum gates applied in a system. For example, the overall unitary transformation given by the quantum circuit

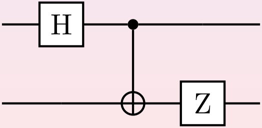


is $(\mathrm{I}_2 \otimes \mathrm{Z})\textbf{CNOT}(\mathrm{H} \otimes \mathrm{I}_2)$.

§3.6 Tensor Product of Vector Spaces

§3.6.2 Correspondence b/w tensor product & quantum circuits

The tensor product of quantum gates represents a unitary transformation when these quantum gates are applied in parallel (at the same time), while the ordinary product of quantum gates represents a unitary transformation when these quantum gates are applied sequentially. Using the matrix representation of quantum gates, there is a way to understand the overall effect of all quantum gates applied in a system. For example, the overall unitary transformation given by the quantum circuit



is $(\mathrm{I}_2 \otimes \mathrm{Z})\textbf{CNOT}(\mathrm{H} \otimes \mathrm{I}_2)$.

§3.6 Tensor Product of Vector Spaces

Using the matrix representation

$$[\mathrm{H}\otimes \mathrm{I}_2] = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \otimes \mathrm{I}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} \mathrm{I}_2 & \mathrm{I}_2 \\ \mathrm{I}_2 & -\mathrm{I}_2 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \end{bmatrix},$$

$$[\mathrm{I}_2\otimes \mathrm{Z}] = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \otimes \mathrm{Z} = \begin{bmatrix} \mathrm{Z} & 0 \\ 0 & \mathrm{Z} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

as well as the matrix representation of **CNOT**, we find that

§3.6 Tensor Product of Vector Spaces

the matrix representation of the overall unitary transformation given by the quantum circuit can be computed by

$$\begin{bmatrix}1&0&0&0\\0&-1&0&0\\0&0&1&0\\0&0&0&-1\end{bmatrix}\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&0&1\\0&0&1&0\end{bmatrix}\begin{bmatrix}1/\sqrt{2}&0&1/\sqrt{2}&0\\0&1/\sqrt{2}&0&1/\sqrt{2}\\1/\sqrt{2}&0&-1/\sqrt{2}&0\\0&1/\sqrt{2}&0&-1/\sqrt{2}\end{bmatrix}$$

$$=\begin{bmatrix}1&0&0&0\\0&-1&0&0\\0&0&1&0\\0&0&0&-1\end{bmatrix}\begin{bmatrix}1/\sqrt{2}&0&1/\sqrt{2}&0\\0&1/\sqrt{2}&0&1/\sqrt{2}\\0&1/\sqrt{2}&0&-1/\sqrt{2}\\1/\sqrt{2}&0&-1/\sqrt{2}&0\end{bmatrix}$$

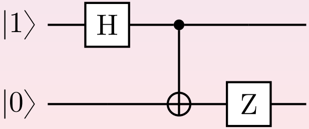
$$=\begin{bmatrix}1/\sqrt{2}&0&1/\sqrt{2}&0\\0&-1/\sqrt{2}&0&-1/\sqrt{2}\\0&1/\sqrt{2}&0&-1/\sqrt{2}\\-1/\sqrt{2}&0&1/\sqrt{2}&0\end{bmatrix}.$$

§3.6 Tensor Product of Vector Spaces

This matrix representation of the overall unitary transform immediately tells us how to produce the EPR pair $\frac{1}{\sqrt{2}}\big(|00\rangle + |11\rangle\big)$: simply apply this circuit to the state $|10\rangle$ since $[|10\rangle] = \begin{bmatrix}0 & 0 & 1 & 0\end{bmatrix}^{\mathrm{T}}$ and

$$\begin{bmatrix} 1/\sqrt{2} & 0 & 1/\sqrt{2} & 0 \\ 0 & -1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 0 & 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ -1/\sqrt{2} & 0 & 1/\sqrt{2} & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 0 \\ 1/\sqrt{2} \end{bmatrix}$$

which corresponds to the EPR pair. Therefore,



= the EPR pair

Figure 1: A way to construct the EPR pair

§3.6 Tensor Product of Vector Spaces

§3.6.3 More examples

In the following examples, for an n-qubit system we always use the ordered basis $\mathcal{B}_n \equiv \{|0\rangle, |1\rangle, |2\rangle, \cdots, |2^n - 1\rangle\}$, where, by writting k in terms of binary number $(k_1 k_2 \cdots k_n)_2$; that is,

$$k = 2^{n-1}k_1 + 2^{n-2}k_2 + \cdots + 2^1 k_{n-1} + 2^0 k_n\,,$$

the k-th basis vector in $\mathcal{B}_n$ is $|k-1\rangle$.

Example (Matrix representation of swap operation)

In an n-qubit system, we use $\textbf{SWAP}_{i,j}$ (here we assume $i < j$ since $\textbf{SWAP}_{i,j} \equiv \textbf{SWAP}_{j,i}$ if $i > j$) to denote the swap operator that swaps the value of the i-th and the j-th qubit; that is

$$\begin{aligned}&\textbf{SWAP}_{i,j}|x_1\rangle \otimes \cdots \otimes |x_n\rangle\\ &\quad = |x_1\rangle \otimes \cdots \otimes |x_{i-1}\rangle \otimes |x_j\rangle \otimes |x_{i+1}\rangle \otimes \cdots\cdots\\ &\qquad \cdots\cdots \otimes |x_{j-1}\rangle \otimes |x_i\rangle \otimes |x_{j+1}\rangle \otimes \cdots \otimes |x_n\rangle\,.\end{aligned}$$

§3.6 Tensor Product of Vector Spaces

§3.6.3 More examples

In the following examples, for an n-qubit system we always use the ordered basis $\mathcal{B}_n \equiv \big\{|0\rangle, |1\rangle, |2\rangle, \cdots, |2^n - 1\rangle\big\}$, where, by writting k in terms of binary number $(k_1 k_2 \cdots k_n)_2$; that is,

$$k = 2^{n-1}k_1 + 2^{n-2}k_2 + \cdots + 2^1 k_{n-1} + 2^0 k_n\,,$$

the k-th basis vector in $\mathcal{B}_n$ is $|k-1\rangle$.

Example (Matrix representation of swap operation)

In an n-qubit system, we use $\textbf{SWAP}_{i,j}$ (here we assume $i < j$ since $\textbf{SWAP}_{i,j} \equiv \textbf{SWAP}_{j,i}$ if $i > j$) to denote the swap operator that swaps the value of the i-th and the j-th qubit; that is

$$\begin{aligned}&\textbf{SWAP}_{i,j}|x_1\rangle\otimes\cdots\otimes|x_n\rangle\\ &\quad= |x_1\rangle\otimes\cdots\otimes|x_{i-1}\rangle\otimes|x_j\rangle\otimes|x_{i+1}\rangle\otimes\cdots\cdots\\ &\qquad\cdots\cdots\otimes|x_{j-1}\rangle\otimes|x_i\rangle\otimes|x_{j+1}\rangle\otimes\cdots\otimes|x_n\rangle\,.\end{aligned}$$

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of swap operation (cont.))

We note that **SWAP**$_{i,j}$ is perfectly defined operator as long as $i \neq j$ and $i, j \leqslant n$. On the other hand, the matrix representation for **SWAP**$_{i,j}$ is a $2^n \times 2^n$ matrix which essentially depends on the number of qubits in a qubit system that **SWAP** gate acts on. Therefore, to denote the matrix representation of **SWAP**$_{i,j}$ one should use something like $[\mathbf{SWAP}_{i,j}]_n$ to indicate the number n of qubits in the system. In the following, for simplicity instead of $\big[\mathbf{SWAP}_{i,j}\big]_n$ we still use **SWAP**$_{i,j}$ to denote the matrix representation of **SWAP**$_{i,j}$ without explicitly indicating (but knowing) the number n of qubits in the system under consideration.

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of swap operation (cont.))

We note that $\textbf{SWAP}_{i,j}$ is perfectly defined operator as long as $i \neq j$ and $i, j \leqslant n$. On the other hand, the matrix representation for $\textbf{SWAP}_{i,j}$ is a $2^n \times 2^n$ matrix which essentially depends on the number of qubits in a qubit system that **SWAP** gate acts on. Therefore, to denote the matrix representation of $\textbf{SWAP}_{i,j}$ one should use something like $[\textbf{SWAP}_{i,j}]_n$ to indicate the number n of qubits in the system. In the following, for simplicity instead of $\left[\textbf{SWAP}_{i,j}\right]_n$ we still use $\textbf{SWAP}_{i,j}$ to denote the matrix representation of $\textbf{SWAP}_{i,j}$ without explicitly indicating (but knowing) the number n of qubits in the system under consideration.

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of swap operation (cont.))

We first consider the swap operator on a 2-qubit system, denoted by **SWAP** and defined by

$$\mathbf{SWAP}|x\rangle \otimes |y\rangle = |y\rangle \otimes |x\rangle .$$

Write $|x\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle$ and $|y\rangle = \beta_0|0\rangle + \beta_1|1\rangle$. Then

$$\begin{aligned}|x\rangle \otimes |y\rangle &= \big(\alpha_0|0\rangle + \alpha_1|1\rangle\big) \otimes \big(\beta_0|0\rangle + \beta_1|1\rangle\big)\\ &= \alpha_0\beta_0|0\rangle\otimes|0\rangle + \alpha_0\beta_1|0\rangle\otimes|1\rangle + \alpha_1\beta_0|1\rangle\otimes|0\rangle + \alpha_1\beta_1|1\rangle\otimes|1\rangle\\ &= \alpha_0\beta_0|0\rangle + \alpha_0\beta_1|1\rangle + \alpha_1\beta_0|2\rangle + \alpha_1\beta_1|3\rangle\end{aligned}$$

and

$$\begin{aligned}|y\rangle \otimes |x\rangle &= \big(\beta_0|0\rangle + \beta_1|1\rangle\big) \otimes \big(\alpha_0|0\rangle + \alpha_1|1\rangle\big)\\ &= \alpha_0\beta_0|0\rangle\otimes|0\rangle + \alpha_1\beta_0|0\rangle\otimes|1\rangle + \alpha_0\beta_1|1\rangle\otimes|0\rangle + \alpha_1\beta_1|1\rangle\otimes|1\rangle\\ &= \alpha_0\beta_0|0\rangle + \alpha_1\beta_0|1\rangle + \alpha_1\beta_0|2\rangle + \alpha_1\beta_1|3\rangle\end{aligned}$$

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of swap operation (cont.))

We first consider the swap operator on a 2-qubit system, denoted by **SWAP** and defined by

$$\textbf{SWAP}|x\rangle \otimes |y\rangle = |y\rangle \otimes |x\rangle .$$

Write $|x\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle$ and $|y\rangle = \beta_0|0\rangle + \beta_1|1\rangle$. Then

$$\begin{aligned}|x\rangle \otimes |y\rangle &= \big(\alpha_0|0\rangle + \alpha_1|1\rangle\big) \otimes \big(\beta_0|0\rangle + \beta_1|1\rangle\big)\\ &= \alpha_0\beta_0|0\rangle \otimes |0\rangle + \alpha_0\beta_1|0\rangle \otimes |1\rangle + \alpha_1\beta_0|1\rangle \otimes |0\rangle + \alpha_1\beta_1|1\rangle \otimes |1\rangle\\ &= \alpha_0\beta_0|0\rangle + \alpha_0\beta_1|1\rangle + \alpha_1\beta_0|2\rangle + \alpha_1\beta_1|3\rangle\end{aligned}$$

and

$$\begin{aligned}|y\rangle \otimes |x\rangle &= \big(\beta_0|0\rangle + \beta_1|1\rangle\big) \otimes \big(\alpha_0|0\rangle + \alpha_1|1\rangle\big)\\ &= \alpha_0\beta_0|0\rangle \otimes |0\rangle + \alpha_1\beta_0|0\rangle \otimes |1\rangle + \alpha_0\beta_1|1\rangle \otimes |0\rangle + \alpha_1\beta_1|1\rangle \otimes |1\rangle\\ &= \alpha_0\beta_0|0\rangle + \alpha_1\beta_0|1\rangle + \alpha_1\beta_0|2\rangle + \alpha_1\beta_1|3\rangle\end{aligned}$$

§3.6 Tensor Product of Vector Spaces

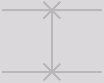
Example (Matrix representation of swap operation (cont.))

so that

$$\textbf{SWAP}\begin{bmatrix}\alpha_0\beta_0\\ \alpha_0\beta_1\\ \alpha_1\beta_0\\ \alpha_1\beta_1\end{bmatrix} = \begin{bmatrix}\alpha_0\beta_0\\ \alpha_1\beta_0\\ \alpha_0\beta_1\\ \alpha_1\beta_1\end{bmatrix} \quad \forall\, |\alpha_0|^2+|\alpha_1|^2=|\beta_0|^2+|\beta_1|^2=1.$$

Therefore, the matrix representation of **SWAP** (relative to the ordered basis $\mathcal{B}_2$) is a 4×4 matrix given by

$$\textbf{SWAP} = \begin{bmatrix}1&0&0&0\\0&0&1&0\\0&1&0&0\\0&0&0&1\end{bmatrix}.$$

The quantum circuit symbol for **SWAP** is 

§3.6 Tensor Product of Vector Spaces

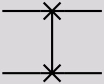
Example (Matrix representation of swap operation (cont.))

so that

$$\mathbf{SWAP}\begin{bmatrix}\alpha_0\beta_0\\ \alpha_0\beta_1\\ \alpha_1\beta_0\\ \alpha_1\beta_1\end{bmatrix} = \begin{bmatrix}\alpha_0\beta_0\\ \alpha_1\beta_0\\ \alpha_0\beta_1\\ \alpha_1\beta_1\end{bmatrix} \quad \forall\, |\alpha_0|^2 + |\alpha_1|^2 = |\beta_0|^2 + |\beta_1|^2 = 1.$$

Therefore, the matrix representation of **SWAP** (relative to the ordered basis $\mathcal{B}_2$) is a 4×4 matrix given by

$$\mathbf{SWAP} = \begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 1\end{bmatrix}.$$

The quantum circuit symbol for **SWAP** is 

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of swap operation (cont.))

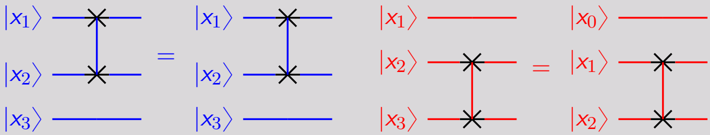
Similarly, on a 3-qubit system, there are three swap operators:

$$\begin{aligned}\textbf{SWAP}_{1,2}|x\rangle\otimes|y\rangle\otimes|z\rangle &= |y\rangle\otimes|x\rangle\otimes|z\rangle\,,\\ \textbf{SWAP}_{2,3}|x\rangle\otimes|y\rangle\otimes|z\rangle &= |x\rangle\otimes|z\rangle\otimes|y\rangle\,,\\ \textbf{SWAP}_{1,3}|x\rangle\otimes|y\rangle\otimes|z\rangle &= |z\rangle\otimes|y\rangle\otimes|x\rangle\,.\end{aligned}$$

Note that

$$\textbf{SWAP}_{1,2} = \textbf{SWAP}\otimes \mathrm{I}_2 \quad\text{and}\quad \textbf{SWAP}_{2,3} = \mathrm{I}_2\otimes\textbf{SWAP}$$

whose validity can be verifies by the quantum circuits:



§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of swap operation (cont.))

By the result of tensor product of linear maps, the matrix representations of **SWAP**$_{1,2}$ and **SWAP**$_{2,3}$ (relative to the ordered basis $\mathcal{B}_3$) can be computed by the two identities

$$\textbf{SWAP}_{1,2} = \textbf{SWAP} \otimes \mathrm{I}_2 = \begin{bmatrix} \mathrm{I}_2 & & & \\ & & \mathrm{I}_2 & \\ & \mathrm{I}_2 & & \\ & & & \mathrm{I}_2 \end{bmatrix} = \begin{bmatrix} 1 & & & & & & & \\ & 1 & & & & & & \\ & & & & 1 & & & \\ & & & & & 1 & & \\ & & 1 & & & & & \\ & & & 1 & & & & \\ & & & & & & 1 & \\ & & & & & & & 1 \end{bmatrix}$$

and

$$\textbf{SWAP}_{2,3} = \mathrm{I}_2 \otimes \textbf{SWAP} = \begin{bmatrix} \textbf{SWAP} & \\ & \textbf{SWAP} \end{bmatrix} = \begin{bmatrix} 1 & & & & & & & \\ & & 1 & & & & & \\ & 1 & & & & & & \\ & & & 1 & & & & \\ & & & & 1 & & & \\ & & & & & & 1 & \\ & & & & & 1 & & \\ & & & & & & & 1 \end{bmatrix}.$$

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of swap operation (cont.))

To compute the matrix representation of **SWAP**$_{1,3}$, writing $|x\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle$, $|y\rangle = \beta_0|0\rangle + \beta_1|1\rangle$ and $|z\rangle = \gamma_0|0\rangle + \gamma_1|1\rangle$ we find that

$$\begin{aligned}\mathbf{SWAP}_{1,3}|x\rangle \otimes |y\rangle \otimes |z\rangle &= |z\rangle \otimes |y\rangle \otimes |x\rangle \\ &= \alpha_0\beta_0\gamma_0|0\rangle + \alpha_1\beta_0\gamma_0|1\rangle + \alpha_0\beta_1\gamma_0|2\rangle + \alpha_1\beta_1\gamma_0|3\rangle \\ &\quad +\alpha_0\beta_0\gamma_1|4\rangle + \alpha_1\beta_0\gamma_1|5\rangle + \alpha_0\beta_1\gamma_1|6\rangle + \alpha_1\beta_1\gamma_1|7\rangle\end{aligned}$$

so that

$$\mathbf{SWAP}_{1,3}\begin{bmatrix}\alpha_0\beta_0\gamma_0\\ \alpha_0\beta_0\gamma_1\\ \alpha_0\beta_1\gamma_0\\ \alpha_0\beta_1\gamma_1\\ \alpha_1\beta_0\gamma_0\\ \alpha_1\beta_0\gamma_1\\ \alpha_1\beta_1\gamma_0\\ \alpha_1\beta_1\gamma_1\end{bmatrix} = \begin{bmatrix}\alpha_0\beta_0\gamma_0\\ \alpha_1\beta_0\gamma_0\\ \alpha_0\beta_1\gamma_0\\ \alpha_1\beta_1\gamma_0\\ \alpha_0\beta_0\gamma_1\\ \alpha_1\beta_0\gamma_1\\ \alpha_0\beta_1\gamma_1\\ \alpha_1\beta_1\gamma_1\end{bmatrix}.$$

§3.6 Tensor Product of Vector Spaces

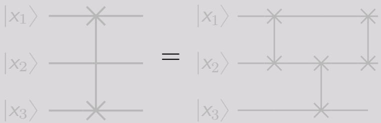
Example (Matrix representation of swap operation (cont.))

Therefore, the matrix representation of $\textbf{SWAP}_{1,3}$ (relative to the ordered basis $\mathcal{B}_3$) is given by

$$\textbf{SWAP}_{1,3} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

$\textbf{SWAP}_{1,3}$ can also be computed using the identities

$$\textbf{SWAP}_{1,3} = \textbf{SWAP}_{1,2} \cdot \textbf{SWAP}_{2,3} \cdot \textbf{SWAP}_{1,2} .$$



§3.6 Tensor Product of Vector Spaces

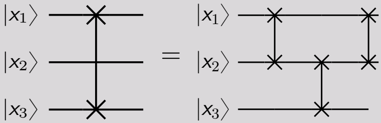
Example (Matrix representation of swap operation (cont.))

Therefore, the matrix representation of $\mathbf{SWAP}_{1,3}$ (relative to the ordered basis $\mathcal{B}_3$) is given by

$$\mathbf{SWAP}_{1,3} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

$\mathbf{SWAP}_{1,3}$ can also be computed using the identities

$$\mathbf{SWAP}_{1,3} = \mathbf{SWAP}_{1,2} \cdot \mathbf{SWAP}_{2,3} \cdot \mathbf{SWAP}_{1,2} .$$



§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of **CNOT**)

The controlled-not gate is a 2-qubit gate defined by

$$\textbf{CNOT}: |x\rangle \otimes |y\rangle \mapsto \begin{cases} |x\rangle \otimes |y\rangle & \text{if } |x\rangle = |0\rangle\,, \\ |x\rangle \otimes (\mathrm{X}|y\rangle) & \text{if } |x\rangle = |1\rangle\,, \end{cases}$$

where X is the NOT gate. Write $|x\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle$ and $|y\rangle = \beta_0|0\rangle + \beta_1|1\rangle$. Then **CNOT** maps

$$\begin{aligned} &\big(\alpha_0|0\rangle + \alpha_1|1\rangle\big) \otimes \big(\beta_0|0\rangle + \beta_1|1\rangle\big) \\ &\quad = \alpha_0\beta_0|0\rangle \otimes |0\rangle + \alpha_0\beta_1|0\rangle \otimes |1\rangle + \alpha_1\beta_0|1\rangle \otimes |0\rangle + \alpha_1\beta_1|1\rangle \otimes |1\rangle \end{aligned}$$

to

$$\begin{aligned} &\alpha_0|0\rangle \otimes |y\rangle + \alpha_1|1\rangle \otimes (|1\rangle \oplus |y\rangle) \\ &\quad = \alpha_0|0\rangle \otimes \big(\beta_0|0\rangle + \beta_1|1\rangle\big) + \alpha_1|1\rangle \otimes \big(\beta_0|1\rangle + \beta_1|0\rangle\big) \\ &\quad = \alpha_0\beta_0|0\rangle \otimes |0\rangle + \alpha_0\beta_1|0\rangle \otimes |1\rangle + \alpha_1\beta_0|1\rangle \otimes |1\rangle + \alpha_1\beta_1|1\rangle \otimes |0\rangle \end{aligned}$$

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of **CNOT**)

The controlled-not gate is a 2-qubit gate defined by

$$\textbf{CNOT}: |x\rangle \otimes |y\rangle \mapsto \begin{cases} |x\rangle \otimes |y\rangle & \text{if } |x\rangle = |0\rangle, \\ |x\rangle \otimes (\mathrm{X}|y\rangle) & \text{if } |x\rangle = |1\rangle, \end{cases}$$

where X is the NOT gate. Write $|x\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle$ and $|y\rangle = \beta_0|0\rangle + \beta_1|1\rangle$. Then **CNOT** maps

$$\begin{aligned}&\big(\alpha_0|0\rangle + \alpha_1|1\rangle\big) \otimes \big(\beta_0|0\rangle + \beta_1|1\rangle\big) \\ &\quad = \alpha_0\beta_0|0\rangle \otimes |0\rangle + \alpha_0\beta_1|0\rangle \otimes |1\rangle + \alpha_1\beta_0|1\rangle \otimes |0\rangle + \alpha_1\beta_1|1\rangle \otimes |1\rangle\end{aligned}$$

to

$$\begin{aligned}&\alpha_0|0\rangle \otimes |y\rangle + \alpha_1|1\rangle \otimes (|1\rangle \oplus |y\rangle) \\ &\quad = \alpha_0|0\rangle \otimes \big(\beta_0|0\rangle + \beta_1|1\rangle\big) + \alpha_1|1\rangle \otimes \big(\beta_0|1\rangle + \beta_1|0\rangle\big) \\ &\quad = \alpha_0\beta_0|0\rangle \otimes |0\rangle + \alpha_0\beta_1|0\rangle \otimes |1\rangle + \alpha_1\beta_0|1\rangle \otimes |1\rangle + \alpha_1\beta_1|1\rangle \otimes |0\rangle\end{aligned}$$

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of **CNOT**)

The controlled-not gate is a 2-qubit gate defined by

$$\textbf{CNOT}: |x\rangle \otimes |y\rangle \mapsto \begin{cases} |x\rangle \otimes |y\rangle & \text{if } |x\rangle = |0\rangle\,, \\ |x\rangle \otimes (\mathrm{X}|y\rangle) & \text{if } |x\rangle = |1\rangle\,, \end{cases}$$

where X is the NOT gate. Write $|x\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle$ and $|y\rangle = \beta_0|0\rangle + \beta_1|1\rangle$. Then **CNOT** maps

$$\begin{aligned}&\big(\alpha_0|0\rangle + \alpha_1|1\rangle\big) \otimes \big(\beta_0|0\rangle + \beta_1|1\rangle\big) \\ &\quad = \alpha_0\beta_0|0\rangle \otimes |0\rangle + \alpha_0\beta_1|0\rangle \otimes |1\rangle + \alpha_1\beta_0|1\rangle \otimes |0\rangle + \alpha_1\beta_1|1\rangle \otimes |1\rangle\end{aligned}$$

to

$$\begin{aligned}&\alpha_0|0\rangle \otimes |y\rangle + \alpha_1|1\rangle \otimes (|1\rangle \oplus |y\rangle) \\ &\quad = \alpha_0|0\rangle \otimes \big(\beta_0|0\rangle + \beta_1|1\rangle\big) + \alpha_1|1\rangle \otimes \big(\beta_0|1\rangle + \beta_1|0\rangle\big) \\ &\quad = \alpha_0\beta_0|0\rangle \otimes |0\rangle + \alpha_0\beta_1|0\rangle \otimes |1\rangle + \alpha_1\beta_0|1\rangle \otimes |1\rangle + \alpha_1\beta_1|1\rangle \otimes |0\rangle\end{aligned}$$

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of **CNOT**)

The controlled-not gate is a 2-qubit gate defined by

$$\textbf{CNOT}: |x\rangle \otimes |y\rangle \mapsto \begin{cases} |x\rangle \otimes |y\rangle & \text{if } |x\rangle = |0\rangle\,, \\ |x\rangle \otimes (\mathrm{X}|y\rangle) & \text{if } |x\rangle = |1\rangle\,, \end{cases}$$

where X is the NOT gate. Write $|x\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle$ and $|y\rangle = \beta_0|0\rangle + \beta_1|1\rangle$. Then **CNOT** maps

$$\begin{aligned} &\big(\alpha_0|0\rangle + \alpha_1|1\rangle\big) \otimes \big(\beta_0|0\rangle + \beta_1|1\rangle\big) \\ &\quad = \alpha_0\beta_0|0\rangle\otimes|0\rangle + \alpha_0\beta_1|0\rangle\otimes|1\rangle + \alpha_1\beta_0|1\rangle\otimes|0\rangle + \alpha_1\beta_1|1\rangle\otimes|1\rangle \end{aligned}$$

to

$$\begin{aligned} &\alpha_0|0\rangle\otimes|y\rangle + \alpha_1|1\rangle\otimes(|1\rangle \oplus |y\rangle) \\ &\quad = \alpha_0|0\rangle\otimes\big(\beta_0|0\rangle + \beta_1|1\rangle\big) + \alpha_1|1\rangle\otimes\big(\beta_0|1\rangle + \beta_1|0\rangle\big) \\ &\quad = \alpha_0\beta_0|0\rangle\otimes|0\rangle + \alpha_0\beta_1|0\rangle\otimes|1\rangle + \alpha_1\beta_0|1\rangle\otimes|1\rangle + \alpha_1\beta_1|1\rangle\otimes|0\rangle \end{aligned}$$

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of **CNOT**)

The controlled-not gate is a 2-qubit gate defined by

$$\textbf{CNOT}: |x\rangle \otimes |y\rangle \mapsto \begin{cases} |x\rangle \otimes |y\rangle & \text{if } |x\rangle = |0\rangle, \\ |x\rangle \otimes (\mathrm{X}|y\rangle) & \text{if } |x\rangle = |1\rangle, \end{cases}$$

where X is the NOT gate. Write $|x\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle$ and $|y\rangle = \beta_0|0\rangle + \beta_1|1\rangle$. Then **CNOT** maps

$$\begin{aligned} &\big(\alpha_0|0\rangle + \alpha_1|1\rangle\big) \otimes \big(\beta_0|0\rangle + \beta_1|1\rangle\big) \\ &\quad = \alpha_0\beta_0 \quad |0\rangle \quad + \alpha_0\beta_1 \quad |1\rangle \quad + \alpha_1\beta_0 \quad |2\rangle \quad + \alpha_1\beta_1 \quad |3\rangle \end{aligned}$$

to

$$\begin{aligned} &\alpha_0|0\rangle \otimes |y\rangle + \alpha_1|1\rangle \otimes (|1\rangle \oplus |y\rangle) \\ &\quad = \alpha_0|0\rangle \otimes \big(\beta_0|0\rangle + \beta_1|1\rangle\big) + \alpha_1|1\rangle \otimes \big(\beta_0|1\rangle + \beta_1|0\rangle\big) \\ &\quad = \alpha_0\beta_0 \quad |0\rangle \quad + \alpha_0\beta_1 \quad |1\rangle \quad + \alpha_1\beta_0 \quad |3\rangle \quad + \alpha_1\beta_1 \quad |2\rangle \end{aligned}$$

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of **CNOT** (cont.))

so that

$$\textbf{CNOT}: \begin{bmatrix} \alpha_0\beta_0 \\ \alpha_0\beta_1 \\ \alpha_1\beta_0 \\ \alpha_1\beta_1 \end{bmatrix} \mapsto \begin{bmatrix} \alpha_0\beta_0 \\ \alpha_0\beta_1 \\ \alpha_1\beta_1 \\ \alpha_1\beta_0 \end{bmatrix}$$

for all $(\alpha_0, \beta_0), (\alpha_1, \beta_1)$ on the Bloch sphere. Therefore, the matrix representation (relative to the ordered basis $\mathcal{B}_2$) is a 4×4 matrix given by

$$\textbf{CNOT} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of **CNOT** (cont.))

so that

$$\textbf{CNOT}: \begin{bmatrix} \alpha_0\beta_0 \\ \alpha_0\beta_1 \\ \alpha_1\beta_0 \\ \alpha_1\beta_1 \end{bmatrix} \mapsto \begin{bmatrix} \alpha_0\beta_0 \\ \alpha_0\beta_1 \\ \alpha_1\beta_1 \\ \alpha_1\beta_0 \end{bmatrix}$$

for all $(\alpha_0, \beta_0), (\alpha_1, \beta_1)$ on the Bloch sphere. Therefore, the matrix representation (relative to the ordered basis $\mathcal{B}_2$) is a 4×4 matrix given by

$$\textbf{CNOT} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

§3.6 Tensor Product of Vector Spaces

Example (The TOFFOLI gate)

The TOFFOLI gate, also called the controlled-controlled-not gate, is a 3-qubit gate defined by

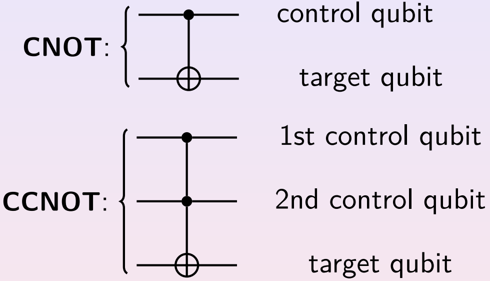
$$\textbf{CCNOT} : |x\rangle \otimes |y\rangle \otimes |z\rangle \mapsto \begin{cases} |x\rangle \otimes |y\rangle \otimes (\mathrm{X}|z\rangle) & \text{if } |x\rangle = |y\rangle = |1\rangle\,, \\ |x\rangle \otimes |y\rangle \otimes |z\rangle & \text{otherwise}\,, \end{cases}$$

where X is the NOT gate. The matrix representation of the TOFFOLI gate is a 8×8 matrix given by

$$\textbf{CCNOT} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

§3.6 Tensor Product of Vector Spaces

The quantum circuit symbols for **CNOT** and **CCNOT**:



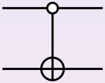
We always use • to denote a control qubit that activates the operation on the target qubit when the value is 1. Another kind of control qubit that activates the operation on the target qubit when the value is 0 is denoted by the symbol $\circ$.

§3.6 Tensor Product of Vector Spaces

For example, the 2-qubit quantum gate

$$|x\rangle \otimes |y\rangle \mapsto \begin{cases} |x\rangle \otimes |y\rangle & \text{if } |x\rangle = |1\rangle\,, \\ |x\rangle \otimes (\mathrm{X}|y\rangle) & \text{if } |x\rangle = |0\rangle\,, \end{cases}$$

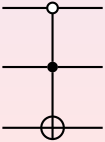
is symbolized by



and the 3-qubit quantum gate

$$|x\rangle \otimes |y\rangle \otimes |z\rangle \mapsto \begin{cases} |x\rangle \otimes |y\rangle \otimes (\mathrm{X}|z\rangle) & \text{if } |x\rangle = |0\rangle \text{ and } |y\rangle = |1\rangle\,, \\ |x\rangle \otimes |y\rangle \otimes |z\rangle & \text{otherwise}\,, \end{cases}$$

will be symbolized by



§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of $\textbf{CNOT}_{i,n}$)

In an n-qubit system, we use $\textbf{CNOT}_{i,j}$ to denote the contorlled-not gate whose controlled qubit is the i-th qubit while the target qubit is the j-th qubit; that is,

$$\textbf{CNOT}_{i,j}\big(|x_1\rangle\otimes|x_2\rangle\otimes\cdots\otimes|x_n\rangle\big) = \begin{cases} |x_1\rangle\otimes\cdots\otimes|x_{j-1}\rangle\otimes(\mathrm{X}|x_j\rangle)\otimes|x_{j+1}\rangle\otimes\cdots\otimes|x_n\rangle & \text{if } |x_i\rangle=|1\rangle, \\ |x_1\rangle\otimes\cdots\otimes|x_n\rangle & \text{if } |x_i\rangle=|0\rangle, \end{cases}$$

where X is the **NOT** gate. We note that $\textbf{CNOT}_{i,j}$ is perfectly defined operator as long as $i\neq j$ and $i,j\leqslant n$. On the other hand, the matrix representation for $\textbf{CNOT}_{i,j}$ is a $2^n\times 2^n$ matrix which essentially depends on the number of qubits that **CNOT** gate acts on. When talking about the matrix representation of $\textbf{CNOT}_{i,j}$, we always assume that it is a $2^k\times 2^k$ matrix, where $k=\max\{i,j\}$.

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of **CNOT**$_{i,n}$)

In an n-qubit system, we use **CNOT**$_{i,j}$ to denote the contorlled-not gate whose controlled qubit is the i-th qubit while the target qubit is the j-th qubit; that is,

$$\textbf{CNOT}_{i,j}\big(|x_1\rangle\otimes|x_2\rangle\otimes\cdots\otimes|x_n\rangle\big) = \begin{cases} |x_1\rangle\otimes\cdots\otimes|x_{j-1}\rangle\otimes(\mathrm{X}|x_j\rangle)\otimes|x_{j+1}\rangle\otimes\cdots\otimes|x_n\rangle & \text{if } |x_i\rangle=|1\rangle, \\ |x_1\rangle\otimes\cdots\otimes|x_n\rangle & \text{if } |x_i\rangle=|0\rangle, \end{cases}$$

where X is the **NOT** gate. We note that **CNOT**$_{i,j}$ is perfectly defined operator as long as $i\neq j$ and $i,j\leqslant n$. On the other hand, the matrix representation for **CNOT**$_{i,j}$ is a $2^n\times 2^n$ matrix which essentially depends on the number of qubits that **CNOT** gate acts on. When talking about the matrix representation of **CNOT**$_{i,j}$, we always assume that it is a $2^k\times 2^k$ matrix, where $k=\max\{i,j\}$.

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of $\mathbf{CNOT}_{i,n}$)

In an n-qubit system, we use $\mathbf{CNOT}_{i,j}$ to denote the contorlled-not gate whose controlled qubit is the i-th qubit while the target qubit is the j-th qubit; that is,

$$\mathbf{CNOT}_{i,j}\big(|x_1\rangle\otimes|x_2\rangle\otimes\cdots\otimes|x_n\rangle\big) = \begin{cases} |x_1\rangle\otimes\cdots\otimes|x_{j-1}\rangle\otimes(\mathrm{X}|x_j\rangle)\otimes|x_{j+1}\rangle\otimes\cdots\otimes|x_n\rangle & \text{if } |x_i\rangle=|1\rangle, \\ |x_1\rangle\otimes\cdots\otimes|x_n\rangle & \text{if } |x_i\rangle=|0\rangle, \end{cases}$$

where X is the **NOT** gate. We note that $\mathbf{CNOT}_{i,j}$ is perfectly defined operator as long as $i\neq j$ and $i,j\leqslant n$. On the other hand, the matrix representation for $\mathbf{CNOT}_{i,j}$ is a $2^n\times 2^n$ matrix which essentially depends on the number of qubits that **CNOT** gate acts on. When talking about the matrix representation of $\mathbf{CNOT}_{i,j}$, we always assume that it is a $2^k\times 2^k$ matrix, where $k=\max\{i,j\}$.

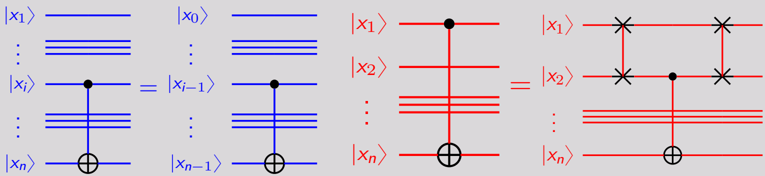
§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of $\textbf{CNOT}_{i,n}$ (cont.))

We first consider $\textbf{CNOT}_{i,n}$ on an n-qubit system, where $1 \leqslant i < n$. The keys for computing the matrix representation of $\textbf{CNOT}_{i,n}$ are the two identities

$$\begin{aligned} \textbf{CNOT}_{i,n} &= \mathrm{I}_2 \otimes \textbf{CNOT}_{i-1,n-1} \\ &= \mathrm{blkdiag}\left(\textbf{CNOT}_{i-1,n-1}, \textbf{CNOT}_{i-1,n-1}\right), \\ \textbf{CNOT}_{1,n} &= \textbf{SWAP}_{1,2} \cdot \textbf{CNOT}_{2,n} \cdot \textbf{SWAP}_{1,2}\,. \end{aligned}$$

The validity of the identities can be verified via the quantum circuits:



§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of $\mathbf{CNOT}_{i,n}$ (cont.))

We first show that for all $n \in \mathbb{N}$,

$$\mathbf{CNOT}_{1,n+1} = \operatorname{blkdiag}\big(\underbrace{\mathrm{I}_2, \mathrm{I}_2, \cdots, \mathrm{I}_2}_{2^{n-1} \text{ copies of } \mathrm{I}_2}, \underbrace{\mathrm{X}, \mathrm{X}, \cdots, \mathrm{X}}_{2^{n-1} \text{ copies of } \mathrm{X}}\big)\,. \tag{5}$$

To see (5), we note that $\mathbf{CNOT}_{1,2} = \mathbf{CNOT} = \operatorname{blkdiag}(\mathrm{I}_2, \mathrm{X})$, and

$$\begin{aligned}
\mathbf{CNOT}_{1,3} &= \mathbf{SWAP}_{1,2} \cdot \mathbf{CNOT}_{2,3} \cdot \mathbf{SWAP}_{1,2} \\
&= (\mathbf{SWAP} \otimes \mathrm{I}_2) \cdot (\mathrm{I}_2 \otimes \mathbf{CNOT}) \cdot (\mathbf{SWAP} \otimes \mathrm{I}_2) \\
&= \begin{bmatrix} \mathrm{I}_2 & & & \\ & & \mathrm{I}_2 & \\ & \mathrm{I}_2 & & \\ & & & \mathrm{I}_2 \end{bmatrix} \begin{bmatrix} \mathrm{I}_2 & & & \\ & \mathrm{X} & & \\ & & \mathrm{I}_2 & \\ & & & \mathrm{X} \end{bmatrix} \begin{bmatrix} \mathrm{I}_2 & & & \\ & & \mathrm{I}_2 & \\ & \mathrm{I}_2 & & \\ & & & \mathrm{I}_2 \end{bmatrix} \\
&= \operatorname{blkdiag}(\mathrm{I}_2, \mathrm{I}_2, \mathrm{X}, \mathrm{X})\,.
\end{aligned}$$

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of $\mathbf{CNOT}_{i,n}$ (cont.))

We first show that for all $n \in \mathbb{N}$,

$$\mathbf{CNOT}_{1,n+1} = \mathrm{blkdiag}\big(\underbrace{\mathrm{I}_2, \mathrm{I}_2, \cdots, \mathrm{I}_2}_{2^{n-1} \text{ copies of } \mathrm{I}_2}, \underbrace{\mathrm{X}, \mathrm{X}, \cdots, \mathrm{X}}_{2^{n-1} \text{ copies of } \mathrm{X}}\big)\,. \tag{5}$$

To see (5), we note that $\mathbf{CNOT}_{1,2} = \mathbf{CNOT} = \mathrm{blkdiag}(\mathrm{I}_2, \mathrm{X})$, and

$$\begin{aligned}
\mathbf{CNOT}_{1,3} &= \mathbf{SWAP}_{1,2} \cdot \mathbf{CNOT}_{2,3} \cdot \mathbf{SWAP}_{1,2} \\
&= (\mathbf{SWAP} \otimes \mathrm{I}_2) \cdot (\mathrm{I}_2 \otimes \mathbf{CNOT}) \cdot (\mathbf{SWAP} \otimes \mathrm{I}_2) \\
&= \begin{bmatrix} \mathrm{I}_2 & & & \\ & & \mathrm{I}_2 & \\ & \mathrm{I}_2 & & \\ & & & \mathrm{I}_2 \end{bmatrix}
\begin{bmatrix} \mathrm{I}_2 & & & \\ & \mathrm{X} & & \\ & & \mathrm{I}_2 & \\ & & & \mathrm{X} \end{bmatrix}
\begin{bmatrix} \mathrm{I}_2 & & & \\ & & \mathrm{I}_2 & \\ & \mathrm{I}_2 & & \\ & & & \mathrm{I}_2 \end{bmatrix} \\
&= \mathrm{blkdiag}(\mathrm{I}_2, \mathrm{I}_2, \mathrm{X}, \mathrm{X})\,.
\end{aligned}$$

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of $\textbf{CNOT}_{i,n}$ (cont.))

We first show that for all $n \in \mathbb{N}$,

$$\textbf{CNOT}_{1,n+1} = \mathrm{blkdiag}\Big(\underbrace{\mathrm{I}_2, \mathrm{I}_2, \cdots, \mathrm{I}_2}_{2^{n-1} \text{ copies of } \mathrm{I}_2}, \underbrace{\mathrm{X}, \mathrm{X}, \cdots, \mathrm{X}}_{2^{n-1} \text{ copies of } \mathrm{X}} \Big) . \tag{5}$$

To see (5), we note that $\textbf{CNOT}_{1,2} = \textbf{CNOT} = \mathrm{blkdiag}(\mathrm{I}_2, \mathrm{X})$, and

$$\begin{aligned}
\textbf{CNOT}_{1,3} &= \textbf{SWAP}_{1,2} \cdot \textbf{CNOT}_{2,3} \cdot \textbf{SWAP}_{1,2} \\
&= (\textbf{SWAP} \otimes \mathrm{I}_2) \cdot (\mathrm{I}_2 \otimes \textbf{CNOT}) \cdot (\textbf{SWAP} \otimes \mathrm{I}_2) \\
&= \begin{bmatrix} \mathrm{I}_2 & & & \\ & & \mathrm{I}_2 & \\ & \mathrm{I}_2 & & \\ & & & \mathrm{I}_2 \end{bmatrix} \begin{bmatrix} \mathrm{I}_2 & & & \\ & \mathrm{X} & & \\ & & \mathrm{I}_2 & \\ & & & \mathrm{X} \end{bmatrix} \begin{bmatrix} \mathrm{I}_2 & & & \\ & & \mathrm{I}_2 & \\ & \mathrm{I}_2 & & \\ & & & \mathrm{I}_2 \end{bmatrix} \\
&= \mathrm{blkdiag}(\mathrm{I}_2, \mathrm{I}_2, \mathrm{X}, \mathrm{X}) .
\end{aligned}$$

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of $\mathbf{CNOT}_{i,n}$ (cont.))

Suppose that (5) holds for the case $n = m$. If $n = m+1$, by writing

$$\mathrm{X}_{2^m} = \mathrm{blkdiag}(\underbrace{\mathrm{X}, \mathrm{X}, \cdots, \mathrm{X}}_{2^{m-1} \text{ copies of } \mathrm{X}})$$

so that $\mathbf{CNOT}_{1,m+1} = \mathrm{blkdiag}\big(\mathrm{I}_{2^m}, \mathrm{X}_{2^m}\big)$, we have

$$\begin{aligned}
\mathbf{CNOT}_{1,n+1} &= \mathbf{SWAP}_{1,2} \cdot \mathbf{CNOT}_{2,n+1} \cdot \mathbf{SWAP}_{1,2} \\
&= (\mathbf{SWAP} \otimes \mathrm{I}_{2^m}) \cdot (\mathrm{I}_2 \otimes \mathbf{CNOT}_{1,m+1}) \cdot (\mathbf{SWAP} \otimes \mathrm{I}_{2^m}) \\
&= \begin{bmatrix} \mathrm{I}_{2^m} & & & \\ & & \mathrm{I}_{2^m} & \\ & \mathrm{I}_{2^m} & & \\ & & & \mathrm{I}_{2^m} \end{bmatrix} \begin{bmatrix} \mathrm{I}_{2^m} & & & \\ & \mathrm{X}_{2^m} & & \\ & & \mathrm{I}_{2^m} & \\ & & & \mathrm{X}_{2^m} \end{bmatrix} \begin{bmatrix} \mathrm{I}_{2^m} & & & \\ & & \mathrm{I}_{2^m} & \\ & \mathrm{I}_{2^m} & & \\ & & & \mathrm{I}_{2^m} \end{bmatrix} \\
&= \mathrm{blkdiag}(\mathrm{I}_{2^m}, \mathrm{I}_{2^m}, \mathrm{X}_{2^m}, \mathrm{X}_{2^m}) = \mathrm{blkdiag}(\mathrm{I}_{2^{m+1}}, \mathrm{X}_{2^{m+1}}) .
\end{aligned}$$

Therefore, (5) is established by induction.

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of $\textbf{CNOT}_{i,n}$ (cont.))

Suppose that (5) holds for the case $n = m$. If $n = m + 1$, by writing

$$\mathrm{X}_{2^m} = \mathrm{blkdiag}(\underbrace{\mathrm{X}, \mathrm{X}, \cdots, \mathrm{X}}_{2^{m-1} \text{ copies of } \mathrm{X}})$$

so that $\textbf{CNOT}_{1,m+1} = \mathrm{blkdiag}\big(\mathrm{I}_{2^m}, \mathrm{X}_{2^m}\big)$, we have

$$\begin{aligned}\textbf{CNOT}_{1,n+1} &= \textbf{SWAP}_{1,2} \cdot \textbf{CNOT}_{2,n+1} \cdot \textbf{SWAP}_{1,2} \\ &= (\textbf{SWAP} \otimes \mathrm{I}_{2^m}) \cdot (\mathrm{I}_2 \otimes \textbf{CNOT}_{1,m+1}) \cdot (\textbf{SWAP} \otimes \mathrm{I}_{2^m}) \\ &= \begin{bmatrix} \mathrm{I}_{2^m} & & & \\ & & \mathrm{I}_{2^m} & \\ & \mathrm{I}_{2^m} & & \\ & & & \mathrm{I}_{2^m} \end{bmatrix} \begin{bmatrix} \mathrm{I}_{2^m} & & & \\ & \mathrm{X}_{2^m} & & \\ & & \mathrm{I}_{2^m} & \\ & & & \mathrm{X}_{2^m} \end{bmatrix} \begin{bmatrix} \mathrm{I}_{2^m} & & & \\ & & \mathrm{I}_{2^m} & \\ & \mathrm{I}_{2^m} & & \\ & & & \mathrm{I}_{2^m} \end{bmatrix} \\ &= \mathrm{blkdiag}(\mathrm{I}_{2^m}, \mathrm{I}_{2^m}, \mathrm{X}_{2^m}, \mathrm{X}_{2^m}) = \mathrm{blkdiag}(\mathrm{I}_{2^{m+1}}, \mathrm{X}_{2^{m+1}}) .\end{aligned}$$

Therefore, (5) is established by induction.

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of $\mathbf{CNOT}_{i,n}$ (cont.))

Having established that

$$\mathbf{CNOT}_{1,n+1} = \mathrm{blkdiag}\big(\underbrace{\mathrm{I}_2, \mathrm{I}_2, \cdots, \mathrm{I}_2}_{2^{n-1} \text{ copies of } \mathrm{I}_2}, \underbrace{\mathrm{X}, \mathrm{X}, \cdots, \mathrm{X}}_{2^{n-1} \text{ copies of } \mathrm{X}} \big) ,$$

$$\mathbf{CNOT}_{i,n+1} = \mathrm{blkdiag}(\mathbf{CNOT}_{i-1,n}, \mathbf{CNOT}_{i-1,n}) ,$$

we have

$$\mathbf{CNOT}_{1,3} = \mathrm{blkdiag}(\mathrm{I}_2, \mathrm{I}_2, \mathrm{X}, \mathrm{X}) ,$$
$$\mathbf{CNOT}_{2,3} = \mathrm{blkdiag}(\mathrm{I}_2, \mathrm{X}, \mathrm{I}_2, \mathrm{X}) .$$

The identities above further imply that

$$\mathbf{CNOT}_{1,4} = \mathrm{blkdiag}(\mathrm{I}_2, \mathrm{I}_2, \mathrm{I}_2, \mathrm{I}_2, \mathrm{X}, \mathrm{X}, \mathrm{X}, \mathrm{X}) ,$$
$$\mathbf{CNOT}_{2,4} = \mathrm{blkdiag}(\mathrm{I}_2, \mathrm{I}_2, \mathrm{X}, \mathrm{X}, \mathrm{I}_2, \mathrm{I}_2, \mathrm{X}, \mathrm{X}) ,$$
$$\mathbf{CNOT}_{3,4} = \mathrm{blkdiag}(\mathrm{I}_2, \mathrm{X}, \mathrm{I}_2, \mathrm{X}, \mathrm{I}_2, \mathrm{X}, \mathrm{I}_2, \mathrm{X}) ,$$

and

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of **CNOT**$_{i,n}$ (cont.))

Having established that

$$\mathbf{CNOT}_{1,n+1} = \mathrm{blkdiag}\big(\underbrace{\mathrm{I}_2, \mathrm{I}_2, \cdots, \mathrm{I}_2}_{2^{n-1}\text{ copies of }\mathrm{I}_2}, \underbrace{\mathrm{X}, \mathrm{X}, \cdots, \mathrm{X}}_{2^{n-1}\text{ copies of }\mathrm{X}}\big),$$

$$\mathbf{CNOT}_{i,n+1} = \mathrm{blkdiag}(\mathbf{CNOT}_{i-1,n}, \mathbf{CNOT}_{i-1,n}),$$

we have

$$\mathbf{CNOT}_{1,3} = \mathrm{blkdiag}(\mathrm{I}_2, \mathrm{I}_2, \mathrm{X}, \mathrm{X}),$$
$$\mathbf{CNOT}_{2,3} = \mathrm{blkdiag}(\mathrm{I}_2, \mathrm{X}, \mathrm{I}_2, \mathrm{X}).$$

The identities above further imply that

$$\mathbf{CNOT}_{1,4} = \mathrm{blkdiag}(\mathrm{I}_2, \mathrm{I}_2, \mathrm{I}_2, \mathrm{I}_2, \mathrm{X}, \mathrm{X}, \mathrm{X}, \mathrm{X}),$$
$$\mathbf{CNOT}_{2,4} = \mathrm{blkdiag}(\mathrm{I}_2, \mathrm{I}_2, \mathrm{X}, \mathrm{X}, \mathrm{I}_2, \mathrm{I}_2, \mathrm{X}, \mathrm{X}),$$
$$\mathbf{CNOT}_{3,4} = \mathrm{blkdiag}(\mathrm{I}_2, \mathrm{X}, \mathrm{I}_2, \mathrm{X}, \mathrm{I}_2, \mathrm{X}, \mathrm{I}_2, \mathrm{X}),$$

and

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of $\textbf{CNOT}_{i,n}$ (cont.))

$$\begin{aligned}
\textbf{CNOT}_{1,5} &= \mathrm{blkdiag}(\mathrm{I}_2, \mathrm{I}_2, \mathrm{I}_2, \mathrm{I}_2, \mathrm{I}_2, \mathrm{I}_2, \mathrm{I}_2, \mathrm{I}_2, \mathrm{X}, \mathrm{X}, \mathrm{X}, \mathrm{X}, \mathrm{X}, \mathrm{X}, \mathrm{X}, \mathrm{X}),\\
\textbf{CNOT}_{2,5} &= \mathrm{blkdiag}(\textbf{CNOT}_{1,4}, \textbf{CNOT}_{1,4})\\
&= \mathrm{blkdiag}(\mathrm{I}_2, \mathrm{I}_2, \mathrm{I}_2, \mathrm{I}_2, \mathrm{X}, \mathrm{X}, \mathrm{X}, \mathrm{X}, \mathrm{I}_2, \mathrm{I}_2, \mathrm{I}_2, \mathrm{I}_2, \mathrm{X}, \mathrm{X}, \mathrm{X}, \mathrm{X}),\\
\textbf{CNOT}_{3,5} &= \mathrm{blkdiag}(\textbf{CNOT}_{2,4}, \textbf{CNOT}_{2,4})\\
&= \mathrm{blkdiag}(\mathrm{I}_2, \mathrm{I}_2, \mathrm{X}, \mathrm{X}, \mathrm{I}_2, \mathrm{I}_2, \mathrm{X}, \mathrm{X}, \mathrm{I}_2, \mathrm{I}_2, \mathrm{X}, \mathrm{X}, \mathrm{I}_2, \mathrm{I}_2, \mathrm{X}, \mathrm{X}),\\
\textbf{CNOT}_{4,5} &= \mathrm{blkdiag}(\textbf{CNOT}_{3,4}, \textbf{CNOT}_{3,4})\\
&= \mathrm{blkdiag}(\mathrm{I}_2, \mathrm{X}, \mathrm{I}_2, \mathrm{X}, \mathrm{I}_2, \mathrm{X}, \mathrm{I}_2, \mathrm{X}, \mathrm{I}_2, \mathrm{X}, \mathrm{I}_2, \mathrm{X}, \mathrm{I}_2, \mathrm{X}, \mathrm{I}_2, \mathrm{X}).
\end{aligned}$$

In general, by defining $\mathrm{IX}_k = \mathrm{blkdiag}\big(\underbrace{\mathrm{I}_2, \cdots, \mathrm{I}_2}_{2^k \text{ copies of } \mathrm{I}_2}, \underbrace{\mathrm{X}, \cdots, \mathrm{X}}_{2^k \text{ copies of } \mathrm{X}}\big)$,

$$\textbf{CNOT}_{i,n} = \mathrm{blkdiag}\big(\underbrace{\mathrm{IX}_{n-i-1}, \cdots, \mathrm{IX}_{n-i-1}}_{2^{i-1} \text{ copies of } \mathrm{IX}_{n-i-1}}\big).$$

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of $\textbf{CNOT}_{i,n}$ (cont.))

$$\begin{aligned}
\textbf{CNOT}_{1,5} &= \mathrm{blkdiag}(\mathrm{I}_2, \mathrm{I}_2, \mathrm{I}_2, \mathrm{I}_2, \mathrm{I}_2, \mathrm{I}_2, \mathrm{I}_2, \mathrm{I}_2, \mathrm{X}, \mathrm{X}, \mathrm{X}, \mathrm{X}, \mathrm{X}, \mathrm{X}, \mathrm{X}, \mathrm{X}),\\
\textbf{CNOT}_{2,5} &= \mathrm{blkdiag}(\textbf{CNOT}_{1,4}, \textbf{CNOT}_{1,4})\\
&= \mathrm{blkdiag}(\mathrm{I}_2, \mathrm{I}_2, \mathrm{I}_2, \mathrm{I}_2, \mathrm{X}, \mathrm{X}, \mathrm{X}, \mathrm{X}, \mathrm{I}_2, \mathrm{I}_2, \mathrm{I}_2, \mathrm{I}_2, \mathrm{X}, \mathrm{X}, \mathrm{X}, \mathrm{X}),\\
\textbf{CNOT}_{3,5} &= \mathrm{blkdiag}(\textbf{CNOT}_{2,4}, \textbf{CNOT}_{2,4})\\
&= \mathrm{blkdiag}(\mathrm{I}_2, \mathrm{I}_2, \mathrm{X}, \mathrm{X}, \mathrm{I}_2, \mathrm{I}_2, \mathrm{X}, \mathrm{X}, \mathrm{I}_2, \mathrm{I}_2, \mathrm{X}, \mathrm{X}, \mathrm{I}_2, \mathrm{I}_2, \mathrm{X}, \mathrm{X}),\\
\textbf{CNOT}_{4,5} &= \mathrm{blkdiag}(\textbf{CNOT}_{3,4}, \textbf{CNOT}_{3,4})\\
&= \mathrm{blkdiag}(\mathrm{I}_2, \mathrm{X}, \mathrm{I}_2, \mathrm{X}, \mathrm{I}_2, \mathrm{X}, \mathrm{I}_2, \mathrm{X}, \mathrm{I}_2, \mathrm{X}, \mathrm{I}_2, \mathrm{X}, \mathrm{I}_2, \mathrm{X}, \mathrm{I}_2, \mathrm{X}).
\end{aligned}$$

In general, by defining $\mathrm{IX}_k = \mathrm{blkdiag}\Big(\underbrace{\mathrm{I}_2, \cdots, \mathrm{I}_2}_{2^k \text{ copies of } \mathrm{I}_2}, \underbrace{\mathrm{X}, \cdots, \mathrm{X}}_{2^k \text{ copies of } \mathrm{X}}\Big)$,

$$\textbf{CNOT}_{i,n} = \mathrm{blkdiag}\Big(\underbrace{\mathrm{IX}_{n-i-1}, \cdots, \mathrm{IX}_{n-i-1}}_{2^{i-1} \text{ copies of } \mathrm{IX}_{n-i-1}}\Big).$$

§3.6 Tensor Product of Vector Spaces

Remark: For $n \geqslant 1$, let $\left[\sigma_{i1}, \sigma_{i2}, \cdots, \sigma_{i2^n}\right]$ be the $(2^{n-i}+1)$-th row of the matrix H_n, where H_n is the **unnormalized** Walsh-Hadamard matrix, then

$$\textbf{CNOT}_{i,n+1} = \mathrm{blkdiag}\left(\mathrm{X}^{(1-\sigma_{i1})/2}, \mathrm{X}^{(1-\sigma_{i2})/2}, \cdots, \mathrm{X}^{(1-\sigma_{i2^n})/2}\right),$$

where $\mathrm{X}^0 \equiv \mathrm{I}_2$. In other words, with f denoting the matrix-valued function $f(1) = \mathrm{I}_2$ and $f(-1) = \mathrm{X}$,

$$\textbf{CNOT}_{i,n+1} = \mathrm{blkdiag}\left(f(\sigma_{i1}), f(\sigma_{i2}), \cdots, f(\sigma_{i2^n})\right).$$

The row vector $\left[\sigma_{i1}, \sigma_{i2}, \cdots, \sigma_{i2^n}\right]$ defined above is called the **symbol** of $\textbf{CNOT}_{i,n+1}$ in this note.

§3.6 Tensor Product of Vector Spaces

Remark: For $n \geqslant 1$, let $\left[\sigma_{i1}, \sigma_{i2}, \cdots, \sigma_{i2^n}\right]$ be the $(2^{n-i}+1)$-th row of the matrix H_n, where H_n is the **unnormalized** Walsh-Hadamard matrix, then

$$\textbf{CNOT}_{i,n+1} = \mathrm{blkdiag}\big(\mathrm{X}^{(1-\sigma_{i1})/2}, \mathrm{X}^{(1-\sigma_{i2})/2}, \cdots, \mathrm{X}^{(1-\sigma_{i2^n})/2}\big) ,$$

where $\mathrm{X}^0 \equiv \mathrm{I}_2$. In other words, with f denoting the matrix-valued function $f(1) = \mathrm{I}_2$ and $f(-1) = \mathrm{X}$,

$$\textbf{CNOT}_{i,n+1} = \mathrm{blkdiag}\big(f(\sigma_{i1}), f(\sigma_{i2}), \cdots, f(\sigma_{i2^n})\big) .$$

The row vector $\left[\sigma_{i1}, \sigma_{i2}, \cdots, \sigma_{i2^n}\right]$ defined above is called the **symbol** of **CNOT**$_{i,n+1}$ in this note.

§3.6 Tensor Product of Vector Spaces

Remark: For $n \geqslant 1$, let $\left[\sigma_{i1}, \sigma_{i2}, \cdots, \sigma_{i2^n}\right]$ be the $(2^{n-i}+1)$-th row of the matrix H_n, where H_n is the **unnormalized** Walsh-Hadamard matrix, then

$$\textbf{CNOT}_{i,n+1} = \mathrm{blkdiag}\left(\mathrm{X}^{(1-\sigma_{i1})/2}, \mathrm{X}^{(1-\sigma_{i2})/2}, \cdots, \mathrm{X}^{(1-\sigma_{i2^n})/2}\right),$$

where $\mathrm{X}^0 \equiv \mathrm{I}_2$. In other words, with f denoting the matrix-valued function $f(1) = \mathrm{I}_2$ and $f(-1) = \mathrm{X}$,

$$\textbf{CNOT}_{i,n+1} = \mathrm{blkdiag}\left(f(\sigma_{i1}), f(\sigma_{i2}), \cdots, f(\sigma_{i2^n})\right).$$

The row vector $\left[\sigma_{i1}, \sigma_{i2}, \cdots, \sigma_{i2^n}\right]$ defined above is called the **symbol** of $\textbf{CNOT}_{i,n+1}$ in this note.

§3.6 Tensor Product of Vector Spaces

Definition

An n-qubit quantum gate is called a **multi-controlled gate** if there exists some qubits, called control qubits, such that each value of the control qubits corresponds to a quantum gate acting on the rest of qubits, the target qubits.

In other words, rather than just applying a gate when all control bits are zero or one, a multi-controlled gate applies operation to the target qubits **can be different** for each of the 2^m possible classical values of the control qubits.

Remark: The **CCNOT** gate can be viewed as a multi-controlled gate since it applies identity gate to the target qubit when the control qubits are $|00\rangle$, $|01\rangle$ and $|10\rangle$.

§3.6 Tensor Product of Vector Spaces

Definition

An *n*-qubit quantum gate is called a **multi-controlled gate** if there exists some qubits, called control qubits, such that each value of the control qubits corresponds to a quantum gate acting on the rest of qubits, the target qubits.

In other words, rather than just applying a gate when all control bits are zero or one, a multi-controlled gate applies operation to the target qubits **can be different** for each of the 2^m possible classical values of the control qubits.

Remark: The **CCNOT** gate can be viewed as a multi-controlled gate since it applies identity gate to the target qubit when the control qubits are $|00\rangle$, $|01\rangle$ and $|10\rangle$.

§3.6 Tensor Product of Vector Spaces

In the following examples, we consider the matrix representation of some special multi-controlled gates.

Example (Multi-controlled gates)

Consider a multi-controlled gate given by

$$L(|x\rangle \otimes |y\rangle) = \begin{cases} |x\rangle \otimes U|y\rangle & \text{if } |x\rangle = |0\rangle\,, \\ |x\rangle \otimes V|y\rangle & \text{if } |x\rangle = |1\rangle\,. \end{cases}$$

where the control qubit $|x\rangle$ is a 1-qubit state, the target qubit $|y\rangle$ is an n-qubit state, and U, V are both n-qubit gates.

Write $|x\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle$, $|y\rangle = \beta_0|0\rangle + \cdots + \beta_{2^n-1}|2^n-1\rangle$, and $|\psi\rangle = |x\rangle \otimes |y\rangle$. Then

$$\begin{aligned} L|\psi\rangle = \alpha_0|0\rangle \otimes \big[U\big(\beta_0|0\rangle + \cdots + \beta_{2^n-1}|2^n-1\rangle\big)\big] \\ +\alpha_1|1\rangle \otimes \big[V\big(\beta_0|0\rangle + \cdots + \beta_{2^n-1}|2^n-1\rangle\big)\big]\,. \end{aligned}$$

§3.6 Tensor Product of Vector Spaces

In the following examples, we consider the matrix representation of some special multi-controlled gates.

Example (Multi-controlled gates)

Consider a multi-controlled gate given by

$$L(|x\rangle \otimes |y\rangle) = \begin{cases} |x\rangle \otimes U|y\rangle & \text{if } |x\rangle = |0\rangle\,, \\ |x\rangle \otimes V|y\rangle & \text{if } |x\rangle = |1\rangle\,. \end{cases}$$

where the control qubit $|x\rangle$ is a 1-qubit state, the target qubit $|y\rangle$ is an n-qubit state, and U, V are both n-qubit gates.

Write $|x\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle$, $|y\rangle = \beta_0|0\rangle + \cdots + \beta_{2^n-1}|2^n-1\rangle$, and $|\psi\rangle = |x\rangle \otimes |y\rangle$. Then

$$\begin{aligned} L|\psi\rangle = {} & \alpha_0|0\rangle \otimes \left[U\big(\beta_0|0\rangle + \cdots + \beta_{2^n-1}|2^n-1\rangle\big)\right] \\ & + \alpha_1|1\rangle \otimes \left[V\big(\beta_0|0\rangle + \cdots + \beta_{2^n-1}|2^n-1\rangle\big)\right]. \end{aligned}$$

§3.6 Tensor Product of Vector Spaces

Example (Multi-controlled gates (cont.))

Note that the matrix representation of $|\psi\rangle$ is given by

$$[\psi] = \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} \otimes \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{2^n-1} \end{bmatrix} = \begin{bmatrix} \alpha_0 \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_{2^n-1} \end{bmatrix} \\ \alpha_1 \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_{2^n-1} \end{bmatrix} \end{bmatrix}$$

so that $[L]$ satisfies

$$[L] : \begin{bmatrix} \alpha_0 \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_{2^n-1} \end{bmatrix} \\ \alpha_1 \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_{2^n-1} \end{bmatrix} \end{bmatrix} \mapsto \begin{bmatrix} \alpha_0 [U] \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_{2^n-1} \end{bmatrix} \\ \alpha_1 [V] \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_{2^n-1} \end{bmatrix} \end{bmatrix}.$$

§3.6 Tensor Product of Vector Spaces

Example (Multi-controlled gates (cont.))

To find the matrix representation of L, we let $\alpha_0 = \beta_{\ell-1} = 1$ for some fixed ℓ while $\alpha_i = \beta_j = 0$ if $i \neq 0$ and $j \neq \ell$ to obtain that the ℓ-th column of $[L]$ is given by

$$[L](:,\ell) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes U(:,\ell) = \begin{bmatrix} U(:,\ell) \\ \mathbf{0}_{2^m} \end{bmatrix}$$

and let $\alpha_1 = \beta_{\ell-1} = 1$ for some fixed ℓ while $\alpha_i = \beta_j = 0$ if $i \neq 1$ and $j \neq \ell$ to obtain that

$$[L](:,2^n+\ell) = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \otimes V(:,\ell) = \begin{bmatrix} \mathbf{0}_{2^m} \\ V(:,\ell) \end{bmatrix},$$

where $\mathbf{0}_{2^m}$ denotes the zero vectors in $\mathbb{C}^{2^m}$. This shows that

$$[L] = \begin{bmatrix} U & \mathbf{0} \\ \mathbf{0} & V \end{bmatrix} = \mathrm{blkdiag}(U,V).$$

§3.6 Tensor Product of Vector Spaces

Example (Multi-controlled gates (cont.))

To find the matrix representation of L, we let $\alpha_0 = \beta_{\ell-1} = 1$ for some fixed ℓ while $\alpha_i = \beta_j = 0$ if $i \neq 0$ and $j \neq \ell$ to obtain that the ℓ-th column of $[L]$ is given by

$$[L](:,\ell) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes U(:,\ell) = \begin{bmatrix} U(:,\ell) \\ \mathbf{0}_{2^m} \end{bmatrix}$$

and let $\alpha_1 = \beta_{\ell-1} = 1$ for some fixed ℓ while $\alpha_i = \beta_j = 0$ if $i \neq 1$ and $j \neq \ell$ to obtain that

$$[L](:,2^n+\ell) = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \otimes V(:,\ell) = \begin{bmatrix} \mathbf{0}_{2^m} \\ V(:,\ell) \end{bmatrix},$$

where $\mathbf{0}_{2^m}$ denotes the zero vectors in $\mathbb{C}^{2^m}$. This shows that

$$[L] = \begin{bmatrix} U & \mathbf{0} \\ \mathbf{0} & V \end{bmatrix} = \mathrm{blkdiag}(U,V).$$

§3.6 Tensor Product of Vector Spaces

Example (Multi-controlled gates (cont.))

To find the matrix representation of L, we let $\alpha_0 = \beta_{\ell-1} = 1$ for some fixed ℓ while $\alpha_i = \beta_j = 0$ if $i \neq 0$ and $j \neq \ell$ to obtain that the ℓ-th column of $[L]$ is given by

$$[L](:, \ell) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes U(:, \ell) = \begin{bmatrix} U(:, \ell) \\ \mathbf{0}_{2^m} \end{bmatrix}$$

and let $\alpha_1 = \beta_{\ell-1} = 1$ for some fixed ℓ while $\alpha_i = \beta_j = 0$ if $i \neq 1$ and $j \neq \ell$ to obtain that

$$[L](:, 2^n + \ell) = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \otimes V(:, \ell) = \begin{bmatrix} \mathbf{0}_{2^m} \\ V(:, \ell) \end{bmatrix},$$

where $\mathbf{0}_{2^m}$ denotes the zero vectors in $\mathbb{C}^{2^m}$. This shows that

$$[L] = \begin{bmatrix} U & \mathbf{0} \\ \mathbf{0} & V \end{bmatrix} = \mathrm{blkdiag}(U, V).$$

§3.6 Tensor Product of Vector Spaces

Example (Multi-controlled gates (cont.))

In general, if a multi-controlled $(n+1)$-qubit gate L is defined by

$$L(|x\rangle \otimes |y\rangle) = \begin{cases} |x\rangle \otimes U_0|y\rangle & \text{if } |x\rangle = |0\rangle\,, \\ |x\rangle \otimes U_1|y\rangle & \text{if } |x\rangle = |1\rangle\,, \\ \vdots & \vdots \\ |x\rangle \otimes U_{2^m-1}|y\rangle & \text{if } |x\rangle = |2^m-1\rangle\,; \end{cases}$$

that is, the controlled qubit $|x\rangle$ is an m-qubit state and $L(|x\rangle \otimes |y\rangle) = |x\rangle \otimes U_k|y\rangle$ if $|x\rangle = |k\rangle$. Then the matrix representation of L is

$$[L] = \mathrm{blkdiag}(U_0, U_1, \cdots, U_{2^m-1})$$

since by letting $|x\rangle = |k-1\rangle$ and $|y\rangle = |\ell-1\rangle$ for some $1 \leqslant k \leqslant 2^m$ and $1 \leqslant \ell \leqslant 2^{n-m+1}$, we have

$$[L](:, (k-1)2^m + \ell) = \mathbf{e}_k \otimes U(:, \ell)\,,$$

where $\{\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_{2^m}\}$ is the standard basis of $\mathbb{C}^{2^m}$.

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of $F^n_{n+1}(U)$)

Consider multi-controlled $(n+1)$-qubit gate L given by

$$L(|x_1\rangle\otimes\cdots\otimes|x_n\rangle\otimes|x_{n+1}\rangle)$$
$$=\begin{cases}|x_1\rangle\otimes\cdots\otimes|x_n\rangle\otimes(U_{0\cdots0}|x_{n+1}\rangle) & \text{if } |x_1\rangle\otimes\cdots\otimes|x_n\rangle=|0\rangle\otimes\cdots\otimes|0\rangle\,,\\ \vdots & \vdots \\ |x_1\rangle\otimes\cdots\otimes|x_n\rangle\otimes(U_{1\cdots1}|x_{n+1}\rangle) & \text{if } |x_1\rangle\otimes\cdots\otimes|x_n\rangle=|1\rangle\otimes\cdots\otimes|1\rangle\,.\end{cases}$$

where $U_{j_1\cdots j_n}$'s are 2×2 unitary matrices for all $j_n,\cdots,j_1\in\{0,1\}^n$, and the controlled qubits are the first n qubits. By identifying $(j_1\cdots j_n)_2$ with j or more precisely,

$$j=(j_1\cdots j_n)_2=2^{n-1}j_1+\cdots+2j_{n-1}+j_n\,,$$

we write $U_{j_1\cdots j_n}$ as U_j and $|j_1\rangle\otimes\cdots\otimes|j_n\rangle$ as $|j\rangle$ so that L can be simply written as

$$L(|x\rangle\otimes|x_{n+1}\rangle)=|x\rangle\otimes(U_j|x_{n+1}\rangle)\qquad\text{if}\quad |x\rangle=|j\rangle\,.$$

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of $F^n_{n+1}(U)$)

Consider multi-controlled $(n+1)$-qubit gate L given by

$$L(|x_1\rangle\otimes\cdots\otimes|x_n\rangle\otimes|x_{n+1}\rangle)$$

$$= \begin{cases} |x_1\rangle\otimes\cdots\otimes|x_n\rangle\otimes(U_{0\cdots0}|x_{n+1}\rangle) & \text{if } |x_1\rangle\otimes\cdots\otimes|x_n\rangle = |0\rangle\otimes\cdots\otimes|0\rangle\,, \\ \vdots & \vdots \\ |x_1\rangle\otimes\cdots\otimes|x_n\rangle\otimes(U_{1\cdots1}|x_{n+1}\rangle) & \text{if } |x_1\rangle\otimes\cdots\otimes|x_n\rangle = |1\rangle\otimes\cdots\otimes|1\rangle\,. \end{cases}$$

where $U_{j_1\cdots j_n}$'s are 2×2 unitary matrices for all $j_n,\cdots,j_1\in\{0,1\}^n$, and the controlled qubits are the first n qubits. By identifying $(j_1\cdots j_n)_2$ with j or more precisely,

$$j=(j_1\cdots j_n)_2=2^{n-1}j_1+\cdots+2j_{n-1}+j_n\,,$$

we write $U_{j_1\cdots j_n}$ as U_j and $|j_1\rangle\otimes\cdots\otimes|j_n\rangle$ as $|j\rangle$ so that L can be simply written as

$$L(|x\rangle\otimes|x_{n+1}\rangle)=|x\rangle\otimes(U_j|x_{n+1}\rangle) \qquad \text{if } \ |x\rangle=|j\rangle\,.$$

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of $F^n_{n+1}(U)$ (cont.))

Suppose that $U_j = \begin{bmatrix} u^{(j)}_{11} & u^{(j)}_{12} \\ u^{(j)}_{21} & u^{(j)}_{22} \end{bmatrix}$, $|x\rangle = \alpha_0|0\rangle + \cdots + \alpha_{2^n-1}|2^n-1\rangle$ and $|x_{n+1}\rangle = \beta_0|0\rangle + \beta_1|1\rangle$. Then

$$[L(|x\rangle \otimes |x_{n+1}\rangle)] = \begin{bmatrix} \alpha_0(u^{(0)}_{11}\beta_0 + u^{(0)}_{12}\beta_1) \\ \alpha_0(u^{(0)}_{21}\beta_0 + u^{(0)}_{22}\beta_1) \\ \alpha_1(u^{(1)}_{11}\beta_0 + u^{(1)}_{12}\beta_1) \\ \alpha_1(u^{(1)}_{21}\beta_0 + u^{(1)}_{22}\beta_1) \\ \vdots \\ \alpha_j(u^{(j)}_{11}\beta_0 + u^{(j)}_{12}\beta_1) \\ \alpha_j(u^{(j)}_{21}\beta_0 + u^{(j)}_{22}\beta_1) \\ \vdots \\ \alpha_{2^n-1}(u^{(2^n-1)}_{11}\beta_0 + u^{(2^n-1)}_{12}\beta_1) \\ \alpha_{2^n-1}(u^{(2^n-1)}_{21}\beta_0 + u^{(2^n-1)}_{22}\beta_1) \end{bmatrix}.$$

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of $F^n_{n+1}(U)$ (cont.))

Suppose that $U_j = \begin{bmatrix} u^{(j)}_{11} & u^{(j)}_{12} \\ u^{(j)}_{21} & u^{(j)}_{22} \end{bmatrix}$, $|x\rangle = \alpha_0|0\rangle + \cdots + \alpha_{2^n-1}|2^n-1\rangle$ and $|x_{n+1}\rangle = \beta_0|0\rangle + \beta_1|1\rangle$. Then

$$[L] : \begin{bmatrix} \alpha_0 \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \\ \alpha_1 \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \\ \vdots \\ \alpha_{2^n-1} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \end{bmatrix} \mapsto \begin{bmatrix} \alpha_0(u^{(0)}_{11}\beta_0 + u^{(0)}_{12}\beta_1) \\ \alpha_0(u^{(0)}_{21}\beta_0 + u^{(0)}_{22}\beta_1) \\ \alpha_1(u^{(1)}_{11}\beta_0 + u^{(1)}_{12}\beta_1) \\ \alpha_1(u^{(1)}_{21}\beta_0 + u^{(1)}_{22}\beta_1) \\ \vdots \\ \alpha_j(u^{(j)}_{11}\beta_0 + u^{(j)}_{12}\beta_1) \\ \alpha_j(u^{(j)}_{21}\beta_0 + u^{(j)}_{22}\beta_1) \\ \vdots \\ \alpha_{2^n-1}(u^{(2^n-1)}_{11}\beta_0 + u^{(2^n-1)}_{12}\beta_1) \\ \alpha_{2^n-1}(u^{(2^n-1)}_{21}\beta_0 + u^{(2^n-1)}_{22}\beta_1) \end{bmatrix}.$$

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of $F^n_{n+1}(U)$ (cont.))

Therefore,

$$\big[L(|x\rangle\otimes|x_{n+1}\rangle)\big] = \begin{bmatrix} u^{(0)}_{11} & u^{(0)}_{12} & & & & & & \\ u^{(0)}_{21} & u^{(0)}_{22} & & & & & & \\ & & u^{(1)}_{11} & u^{(1)}_{12} & & & & \\ & & u^{(1)}_{21} & u^{(1)}_{22} & & & & \\ & & & & \ddots & & & \\ & & & & & \ddots & & \\ & & & & & & u^{(2^n-1)}_{11} & u^{(2^n-1)}_{12} \\ & & & & & & u^{(2^n-1)}_{21} & u^{(2^n-1)}_{22} \end{bmatrix} \begin{bmatrix} \alpha_0 \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \\ \alpha_1 \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \\ \vdots \\ \vdots \\ \alpha_{2^n-1} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \end{bmatrix}.$$

The $2^{n+1} \times 2^{n+1}$ matrix is the matrix representation of L.

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of $F_1^n(U)$)

Consider multi-controlled $(n+1)$-qubit gate L given by

$$L(|x_0\rangle\otimes|x_1\rangle\otimes\cdots\otimes|x_n\rangle)$$
$$= \begin{cases} (U_{0\cdots0}|x_0\rangle)\otimes|x_1\rangle\otimes\cdots\otimes|x_n\rangle & \text{if } |x_1\rangle\otimes\cdots\otimes|x_n\rangle = |0\rangle\otimes\cdots\otimes|0\rangle\,, \\ \vdots & \vdots \\ (U_{1\cdots1}|x_0\rangle)\otimes|x_1\rangle\otimes\cdots\otimes|x_n\rangle & \text{if } |x_1\rangle\otimes\cdots\otimes|x_n\rangle = |1\rangle\otimes\cdots\otimes|1\rangle\,. \end{cases}$$

where $U_{j_1\cdots j_n}$'s are 2×2 unitary matrices for all $j_1,\cdots,j_n\in\{0,1\}^n$, and the control qubits are the last n qubits. By identifying $(j_1\cdots j_n)_2$ with j or more precisely,

$$j = (j_1\cdots j_n)_2 = 2^{n-1}j_1 + \cdots + 2j_{n-1} + j_n\,,$$

we write $U_{j_1\cdots j_n}$ as U_j and $|j_1\rangle\otimes\cdots\otimes|j_n\rangle$ as $|j\rangle$ so that L can be simply written as

$$L(|x_0\rangle\otimes|x\rangle) = (U_j|x_0\rangle)\otimes|x\rangle \qquad \text{if } |x\rangle = |j\rangle\,.$$

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of $F_1^n(U)$)

Consider multi-controlled $(n+1)$-qubit gate L given by

$$L(|x_0\rangle \otimes |x_1\rangle \otimes \cdots \otimes |x_n\rangle)$$
$$= \begin{cases} (U_{0\cdots 0}|x_0\rangle) \otimes |x_1\rangle \otimes \cdots \otimes |x_n\rangle & \text{if } |x_1\rangle \otimes \cdots \otimes |x_n\rangle = |0\rangle \otimes \cdots \otimes |0\rangle, \\ \vdots & \vdots \\ (U_{1\cdots 1}|x_0\rangle) \otimes |x_1\rangle \otimes \cdots \otimes |x_n\rangle & \text{if } |x_1\rangle \otimes \cdots \otimes |x_n\rangle = |1\rangle \otimes \cdots \otimes |1\rangle. \end{cases}$$

where $U_{j_1\cdots j_n}$'s are 2×2 unitary matrices for all $j_1, \cdots, j_n \in \{0,1\}^n$,

and the control qubits are the last n qubits. By identifying $(j_1 \cdots j_n)_2$ with j or more precisely,

$$j = (j_1 \cdots j_n)_2 = 2^{n-1} j_1 + \cdots + 2 j_{n-1} + j_n\,,$$

we write $U_{j_1\cdots j_n}$ as U_j and $|j_1\rangle \otimes \cdots \otimes |j_n\rangle$ as $|j\rangle$ so that L can be simply written as

$$L(|x_0\rangle \otimes |x\rangle) = (U_j|x_0\rangle) \otimes |x\rangle \qquad \text{if } \ |x\rangle = |j\rangle\,.$$

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of $F_1^n(U)$ (cont.))

Suppose that $U_j = \begin{bmatrix} u_{11}^{(j)} & u_{12}^{(j)} \\ u_{21}^{(j)} & u_{22}^{(j)} \end{bmatrix}$, $|x_0\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle$ and $|x\rangle = \beta_0|0\rangle + \cdots + \beta_{2^n-1}|2^n-1\rangle$. Then

$$[L(|x_0\rangle \otimes |x\rangle)] = \begin{bmatrix} (u_{11}^{(0)}\alpha_0 + u_{12}^{(0)}\alpha_1)\beta_0 \\ \vdots \\ (u_{11}^{(j)}\alpha_0 + u_{12}^{(j)}\alpha_1)\beta_j \\ \vdots \\ (u_{11}^{(2^n-1)}\alpha_0 + u_{12}^{(2^n-1)}\alpha_1)\beta_{2^n-1} \\ (u_{21}^{(0)}\alpha_0 + u_{22}^{(0)}\alpha_1)\beta_0 \\ \vdots \\ (u_{21}^{(j)}\alpha_0 + u_{22}^{(j)}\alpha_1)\beta_j \\ \vdots \\ (u_{21}^{(2^n-1)}\alpha_0 + u_{22}^{(2^n-1)}\alpha_1)\beta_{2^n-1} \end{bmatrix}.$$

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of $F_1^n(U)$ (cont.))

Suppose that $U_j = \begin{bmatrix} u_{11}^{(j)} & u_{12}^{(j)} \\ u_{21}^{(j)} & u_{22}^{(j)} \end{bmatrix}$, $|x_0\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle$ and $|x\rangle = \beta_0|0\rangle + \cdots + \beta_{2^n-1}|2^n-1\rangle$. Then

$$[L] : \begin{bmatrix} \alpha_0 \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_{2^n-1} \end{bmatrix} \\ \alpha_1 \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_{2^n-1} \end{bmatrix} \end{bmatrix} \mapsto \begin{bmatrix} (u_{11}^{(0)}\alpha_0 + u_{12}^{(0)}\alpha_1)\beta_0 \\ \vdots \\ (u_{11}^{(j)}\alpha_0 + u_{12}^{(j)}\alpha_1)\beta_j \\ \vdots \\ (u_{11}^{(2^n-1)}\alpha_0 + u_{12}^{(2^n-1)}\alpha_1)\beta_{2^n-1} \\ (u_{21}^{(0)}\alpha_0 + u_{22}^{(0)}\alpha_1)\beta_0 \\ \vdots \\ (u_{21}^{(j)}\alpha_0 + u_{22}^{(j)}\alpha_1)\beta_j \\ \vdots \\ (u_{21}^{(2^n-1)}\alpha_0 + u_{22}^{(2^n-1)}\alpha_1)\beta_{2^n-1} \end{bmatrix}.$$

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of $F_1^n(U)$ (cont.))

Therefore,

$$\big[L(|x_0\rangle\otimes|x\rangle)\big]=\begin{bmatrix}
u^{(0)}_{11} & & & & u^{(0)}_{12} & & & \\
& u^{(1)}_{11} & & & & u^{(1)}_{12} & & \\
& & \ddots & & & & \ddots & \\
& & & u^{(2^n-1)}_{11} & & & & u^{(2^n-1)}_{12} \\
u^{(0)}_{21} & & & & u^{(0)}_{22} & & & \\
& u^{(1)}_{21} & & & & u^{(1)}_{22} & & \\
& & \ddots & & & & \ddots & \\
& & & u^{(2^n-1)}_{21} & & & & u^{(2^n-1)}_{22}
\end{bmatrix}\begin{bmatrix}
\alpha_0\begin{bmatrix}\beta_0\\ \beta_1\\ \vdots\\ \beta_{2^n-1}\end{bmatrix}\\
\alpha_1\begin{bmatrix}\beta_0\\ \beta_1\\ \vdots\\ \beta_{2^n-1}\end{bmatrix}
\end{bmatrix}.$$

The $2^{n+1}\times 2^{n+1}$ matrix is the matrix representation of L.

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of $F_1^n(U)$ (cont.))

In particular, if U_j is a rotation matrix of the form

$$U_j = \mathrm{R}_y(2\theta_{j+1}) = \begin{bmatrix} \cos\theta_{j+1} & -\sin\theta_{j+1} \\ \sin\theta_{j+1} & \cos\theta_{j+1} \end{bmatrix}$$

(here we label U from 0 to $2^n - 1$ but label θ from 1 to 2^n), then

$$[L] = \begin{bmatrix} \cos\theta_1 & & & & -\sin\theta_1 & & & \\ & \cos\theta_2 & & & & -\sin\theta_2 & & \\ & & \ddots & & & & \ddots & \\ & & & \cos\theta_{2^n} & & & & -\sin\theta_{2^n} \\ \sin\theta_1 & & & & \cos\theta_1 & & & \\ & \sin\theta_2 & & & & \cos\theta_2 & & \\ & & \ddots & & & & \ddots & \\ & & & \sin\theta_{2^n} & & & & \cos\theta_{2^n} \end{bmatrix}.$$

A matrix of this form will play important role later.

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of $F_k^n(\mathrm{R}_z)$)

In this example we consider a special multi-controlled $(n+1)$-qubit gate A_j defined by

$$\begin{aligned} & A_j(|x_0\rangle \otimes \cdots \otimes |x_n\rangle) \\ & \quad = |x_0\rangle \otimes \cdots \otimes |x_{j-1}\rangle \otimes (\mathrm{R}_z(\theta_k)|x_j\rangle) \otimes |x_{j+1}\rangle \otimes \cdots \otimes |x_n\rangle \end{aligned}$$

if $(x_0 \cdots x_{j-1}x_{j+1} \cdots x_n)_2 = k$, where R_z is the rotation about z-axis given by

$$\mathrm{R}_z(\theta) = \begin{bmatrix} e^{-i\theta/2} & 0 \\ 0 & e^{i\theta/2} \end{bmatrix}.$$

This is a multi-controlled gate with n control qubits and the target qubit is the $|x_j\rangle$ qubit, and is sometimes denoted by $F_{j+1}^n(\mathrm{R}_z)$ (since the target qubit $|x_j\rangle$ is the $(j+1)$-th qubit counting from the highest/ left-most qubit).

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of $F_k^n(\mathrm{R}_z)$)

In this example we consider a special multi-controlled $(n+1)$-qubit gate A_j defined by

$$\begin{aligned} &A_j(|x_0\rangle \otimes \cdots \otimes |x_n\rangle) \\ &\quad = |x_0\rangle \otimes \cdots \otimes |x_{j-1}\rangle \otimes (\mathrm{R}_z(\theta_k)|x_j\rangle) \otimes |x_{j+1}\rangle \otimes \cdots \otimes |x_n\rangle \end{aligned}$$

if $(x_0 \cdots x_{j-1}x_{j+1} \cdots x_n)_2 = k$, where R_z is the rotation about z-axis given by

$$\mathrm{R}_z(\theta) = \begin{bmatrix} e^{-i\theta/2} & 0 \\ 0 & e^{i\theta/2} \end{bmatrix}.$$

This is a multi-controlled gate with n control qubits and the target qubit is the $|x_j\rangle$ qubit, and is sometimes denoted by $F_{j+1}^n(\mathrm{R}_z)$ (since the target qubit $|x_j\rangle$ is the $(j+1)$-th qubit counting from the highest/ left-most qubit).

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of $F_k^n(\mathrm{R}_z)$ (cont.))

Previous examples establish the case $j = 0$ and $j = n$, so we focus on the case $1 \leqslant j < n$. We first consider the case $j = 1$. In this case, we note that

$$A_{n-1} = \textbf{SWAP}_{n,n+1} \cdot A_n \cdot \textbf{SWAP}_{n,n+1} \,,$$

where the operator $\textbf{SWAP}_{n,n+1}$ swaps the position of the n-th and the $(n+1)$-th qubit, and A_n is the multi-controlled $(n+1)$-qubit gate introduced in previous example with $U_k = \mathrm{R}_z(\theta_k)$ and the target qubit is the $(n+1)$-th qubit. Previous example shows the matrix representation of A_n is given by

$$\begin{aligned}[A_n] &= \mathrm{blkdiag}\big(\mathrm{R}_z(\theta_1), \mathrm{R}_z(\theta_2), \cdots, \mathrm{R}_z(\theta_{2^n})\big)\\ &= \mathrm{diag}\big(e^{-i\theta_1/2}, e^{i\theta_1/2}, e^{-i\theta_2/2}, e^{i\theta_2/2}, \cdots, e^{-i\theta_{2^n}/2}, e^{i\theta_{2^n}/2}\big)\,;\end{aligned}$$

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of $F^n_k(\mathrm{R}_z)$ (cont.))

Previous examples establish the case $j = 0$ and $j = n$, so we focus on the case $1 \leqslant j < n$. We first consider the case $j = 1$. In this case, we note that

$$A_{n-1} = \mathbf{SWAP}_{n,n+1} \cdot A_n \cdot \mathbf{SWAP}_{n,n+1} \,,$$

where the operator **SWAP**$_{n,n+1}$ swaps the position of the n-th and the $(n+1)$-th qubit, and A_n is the multi-controlled $(n+1)$-qubit gate introduced in previous example with $U_k = \mathrm{R}_z(\theta_k)$ and the target qubit is the $(n+1)$-th qubit. Previous example shows the matrix representation of A_n is given by

$$\begin{aligned}[A_n] &= \mathrm{blkdiag}\big(\mathrm{R}_z(\theta_1), \mathrm{R}_z(\theta_2), \cdots, \mathrm{R}_z(\theta_{2^n})\big) \\ &= \mathrm{diag}\big(e^{-i\theta_1/2}, e^{i\theta_1/2}, e^{-i\theta_2/2}, e^{i\theta_2/2}, \cdots, e^{-i\theta_{2^n}/2}, e^{i\theta_{2^n}/2}\big)\,;\end{aligned}$$

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of $F_k^n(\mathrm{R}_z)$ (cont.))

Note that

$$\mathsf{SWAP}_{n,n+1} = \mathrm{I}_{2^{n-1}} \otimes \mathsf{SWAP} = \begin{bmatrix} \mathsf{SWAP} & & & \\ & \mathsf{SWAP} & & \\ & & \ddots & \\ & & & \mathsf{SWAP} \end{bmatrix}$$

and

$$\begin{aligned} &\mathsf{SWAP} \cdot \mathrm{diag}(a, b, c, d) \cdot \mathsf{SWAP} \\ &= \begin{bmatrix} 1 & & & \\ & & 1 & \\ & 1 & & \\ & & & 1 \end{bmatrix} \begin{bmatrix} a & & & \\ & b & & \\ & & c & \\ & & & d \end{bmatrix} \begin{bmatrix} 1 & & & \\ & & 1 & \\ & 1 & & \\ & & & 1 \end{bmatrix} \\ &= \begin{bmatrix} a & & & \\ & c & & \\ & & b & \\ & & & d \end{bmatrix} = \mathrm{diag}(a, c, b, d) \,. \end{aligned}$$

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of $F_k^n(\mathrm{R}_z)$ (cont.))

Note that

$$\textsf{SWAP}_{n,n+1} = \mathrm{I}_{2^{n-1}} \otimes \textsf{SWAP} = \begin{bmatrix} \textsf{SWAP} & & & \\ & \textsf{SWAP} & & \\ & & \ddots & \\ & & & \textsf{SWAP} \end{bmatrix}$$

and

$$\begin{aligned} &\textsf{SWAP} \cdot \mathrm{diag}(a, b, c, d) \cdot \textsf{SWAP} \\ &= \begin{bmatrix} 1 & & & \\ & & 1 & \\ & 1 & & \\ & & & 1 \end{bmatrix} \begin{bmatrix} a & & & \\ & b & & \\ & & c & \\ & & & d \end{bmatrix} \begin{bmatrix} 1 & & & \\ & & 1 & \\ & 1 & & \\ & & & 1 \end{bmatrix} \\ &= \begin{bmatrix} a & & & \\ & c & & \\ & & b & \\ & & & d \end{bmatrix} = \mathrm{diag}(a, c, b, d)\,. \end{aligned}$$

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of $F_k^n(\mathrm{R}_z)$ (cont.))

Therefore, using the fact that

$$A_{n-1} = \mathbf{SWAP}_{n,n+1} \cdot A_n \cdot \mathbf{SWAP}_{n,n+1} \,,$$

we find that

$$[A_{n-1}] = \begin{bmatrix} e^{-i\theta_1/2} &&&&&&&& \\ & e^{-i\theta_2/2} &&&&&&& \\ && e^{i\theta_1/2} &&&&&& \\ &&& e^{i\theta_2/2} &&&&& \\ &&&& e^{-i\theta_3/2} &&&& \\ &&&&& e^{-i\theta_4/2} &&& \\ &&&&&& e^{i\theta_3/2} && \\ &&&&&&& e^{i\theta_4/2} & \\ &&&&&&&& \ddots \end{bmatrix} .$$

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of $F_k^n(\mathrm{R}_z)$ (cont.))

We note that $[A_{n-1}]$ takes the form

$$[A_{n-1}] = \mathrm{blkdiag}(Q_1, Q_2, \cdots, Q_{2^{n-1}})\,,$$

where for each $1 \leqslant k \leqslant 2^{n-1}$,

$$Q_k = \mathrm{diag}\big(e^{-i\theta_{2k-1}/2}, e^{-i\theta_{2k}/2}, e^{i\theta_{2k-1}/2}, e^{i\theta_{2k}/2}\big)$$

for some $\theta_1, \cdots, \theta_{2^n} \in \mathbb{R}$.

In the following, for simplicity we only write the sign and the sub-index of the angle to express the matrix. For example, we write

$$[A_n] = \mathrm{diag}(-1, +1, -2, +2, \cdots, -2^n, +2^n)$$

and

$$[A_{n-1}] = \mathrm{diag}(-1, -2, +1, +2, -3, -4, +3, +4, \cdots, \\ -(2^n-1), -2^n, +(2^n-1), +2^n)\,.$$

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of $F_k^n(\mathrm{R}_z)$ (cont.))

We note that $[A_{n-1}]$ takes the form

$$[A_{n-1}] = \mathrm{blkdiag}(Q_1, Q_2, \cdots, Q_{2^{n-1}}),$$

where for each $1 \leqslant k \leqslant 2^{n-1}$,

$$Q_k = \mathrm{diag}\big(e^{-i\theta_{2k-1}/2}, e^{-i\theta_{2k}/2}, e^{i\theta_{2k-1}/2}, e^{i\theta_{2k}/2}\big)$$

for some $\theta_1, \cdots, \theta_{2^n} \in \mathbb{R}$.

In the following, for simplicity we only write the sign and the sub-index of the angle to express the matrix. For example, we write

$$[A_n] = \mathrm{diag}(-1, +1, -2, +2, \cdots, -2^n, +2^n)$$

and

$$[A_{n-1}] = \mathrm{diag}(-1, -2, +1, +2, -3, -4, +3, +4, \cdots, \\ -(2^n-1), -2^n, +(2^n-1), +2^n).$$

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of $F_k^n(\mathrm{R}_z)$ (cont.))

Now we consider A_{n-2}. Similar to the previous case, we have

$$A_{n-2} \equiv \textbf{SWAP}_{n-1,n} \cdot A_{n-1} \cdot \textbf{SWAP}_{n-1,n} \, .$$

Note that

$$\begin{aligned} \textbf{SWAP}_{n-1,n} &= \mathrm{I}_{2^{n-2}} \otimes \textbf{SWAP} \otimes \mathrm{I}_2 \\ &= \begin{bmatrix} \mathrm{I}_2 & & & & & & & & \\ & & \mathrm{I}_2 & & & & & & \\ & \mathrm{I}_2 & & & & & & & \\ & & & \mathrm{I}_2 & & & & & \\ & & & & \mathrm{I}_2 & & & & \\ & & & & & & \mathrm{I}_2 & & \\ & & & & & \mathrm{I}_2 & & & \\ & & & & & & & \mathrm{I}_2 & \\ & & & & & & & & \ddots \end{bmatrix} \end{aligned}$$

and

$$\begin{aligned} (\textbf{SWAP} \otimes \mathrm{I}_2) \cdot \mathrm{diag}(a, b, c, d, e, f, g, h) \cdot (\textbf{SWAP} \otimes \mathrm{I}_2) \\ = \mathrm{diag}(a, b, e, f, c, d, g, h) \, . \end{aligned}$$

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of $F_k^n(\mathrm{R}_z)$ (cont.))

Now we consider A_{n-2}. Similar to the previous case, we have

$$A_{n-2} \equiv \textbf{SWAP}_{n-1,n} \cdot A_{n-1} \cdot \textbf{SWAP}_{n-1,n} \,.$$

Note that

$$\begin{aligned}\textbf{SWAP}_{n-1,n} &= \mathrm{blkdiag}(\textbf{SWAP} \otimes \mathrm{I}_2, \cdots, \textbf{SWAP} \otimes \mathrm{I}_2) \\ &= \begin{bmatrix} \mathrm{I}_2 & & & & & & & & \\ & & \mathrm{I}_2 & & & & & & \\ & \mathrm{I}_2 & & & & & & & \\ & & & \mathrm{I}_2 & & & & & \\ & & & & \mathrm{I}_2 & & & & \\ & & & & & & \mathrm{I}_2 & & \\ & & & & & \mathrm{I}_2 & & & \\ & & & & & & & \mathrm{I}_2 & \\ & & & & & & & & \ddots \end{bmatrix}\end{aligned}$$

and

$$\begin{aligned}(\textbf{SWAP} \otimes \mathrm{I}_2) \cdot \mathrm{diag}(a, b, c, d, e, f, g, h) \cdot (\textbf{SWAP} \otimes \mathrm{I}_2) \\ = \mathrm{diag}(a, b, e, f, c, d, g, h) \,.\end{aligned}$$

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of $F_k^n(\mathrm{R}_z)$ (cont.))

Now we consider A_{n-2}. Similar to the previous case, we have

$$\color{blue} A_{n-2} \equiv \textbf{SWAP}_{n-1,n} \cdot A_{n-1} \cdot \textbf{SWAP}_{n-1,n} \,.$$

Note that

$$\begin{aligned}\textbf{SWAP}_{n-1,n} &= \operatorname{blkdiag}\big({\color{red}\textbf{SWAP} \otimes \mathrm{I}_2}, \cdots, {\color{red}\textbf{SWAP} \otimes \mathrm{I}_2}\big) \\ &= \begin{bmatrix} {\color{red}\mathrm{I}_2} & & & & & & & & \\ & & {\color{red}\mathrm{I}_2} & & & & & & \\ & {\color{red}\mathrm{I}_2} & & & & & & & \\ & & & {\color{red}\mathrm{I}_2} & & & & & \\ & & & & \mathrm{I}_2 & & & & \\ & & & & & & \mathrm{I}_2 & & \\ & & & & & \mathrm{I}_2 & & & \\ & & & & & & & \mathrm{I}_2 & \\ & & & & & & & & \ddots \end{bmatrix}\end{aligned}$$

and

$$\begin{aligned}&({\color{red}\textbf{SWAP} \otimes \mathrm{I}_2}) \cdot \operatorname{diag}(a, b, {\color{blue}c}, {\color{blue}d}, {\color{red}e}, {\color{red}f}, g, h) \cdot ({\color{red}\textbf{SWAP} \otimes \mathrm{I}_2}) \\ &\quad = \operatorname{diag}(a, b, {\color{red}e}, {\color{red}f}, {\color{blue}c}, {\color{blue}d}, g, h)\,.\end{aligned}$$

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of $F_k^n(\mathrm{R}_z)$ (cont.))

Now we consider A_{n-2}. Similar to the previous case, we have

$$A_{n-2} \equiv \textbf{SWAP}_{n-1,n} \cdot A_{n-1} \cdot \textbf{SWAP}_{n-1,n} \, .$$

Note that

$$\begin{aligned} \textbf{SWAP}_{n-1,n} &= \mathrm{blkdiag}(\textbf{SWAP} \otimes \mathrm{I}_2, \cdots, \textbf{SWAP} \otimes \mathrm{I}_2) \\ &= \begin{bmatrix} \mathrm{I}_2 & & & & & & & & \\ & & \mathrm{I}_2 & & & & & & \\ & \mathrm{I}_2 & & & & & & & \\ & & & \mathrm{I}_2 & & & & & \\ & & & & \mathrm{I}_2 & & & & \\ & & & & & & \mathrm{I}_2 & & \\ & & & & & \mathrm{I}_2 & & & \\ & & & & & & & \mathrm{I}_2 & \\ & & & & & & & & \ddots \end{bmatrix} \end{aligned}$$

and

$$\begin{aligned} &(\textbf{SWAP} \otimes \mathrm{I}_2) \cdot \mathrm{diag}(a, b, c, d, e, f, g, h) \cdot (\textbf{SWAP} \otimes \mathrm{I}_2) \\ &\quad = \mathrm{diag}(a, b, e, f, c, d, g, h) \, . \end{aligned}$$

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of $F_k^n(\mathrm{R}_z)$ (cont.))

Therefore, $[A_{n-2}]$ is obtained by

將 $[A_{n-1}]$ 的對角線元素連續八個視為一組，每一組的第三、第四個這對元素與第五、第六個這對元素交換;

thus using

$$[A_{n-1}] = \mathrm{diag}\big(-1, -2, +1, +2, -3, -4, +3, +4, \\ -5, -6, +5, +6, -7, -8, +7, +8, \cdots\big)$$

we find that

$$[A_{n-2}] = \mathrm{diag}\big(-1, -2, -3, -4, +1, +2, +3, +4, \\ -5, -6, -7, -8, +5, +6, +7, +8, \cdots\big)\,.$$

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of $F_k^n(\mathrm{R}_z)$ (cont.))

Therefore, $[A_{n-2}]$ is obtained by

將 $[A_{n-1}]$ 的對角線元素連續八個視為一組，每一組的第三、第四個這對元素與第五、第六個這對元素交換;

thus using

$$[A_{n-1}] = \mathrm{diag}\big(-1, -2, +1, +2, -3, -4, +3, +4, \\ -5, -6, +5, +6, -7, -8, +7, +8, \cdots\big)$$

we find that

$$[A_{n-2}] = \mathrm{diag}\big(-1, -2, -3, -4, +1, +2, +3, +4, \\ -5, -6, -7, -8, +5, +6, +7, +8, \cdots\big)\,.$$

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of $F_k^n(\mathrm{R}_z)$ (cont.))

We note that $[A_{n-2}]$ takes the form

$$\mathrm{blkdiag}\big(Q_1, Q_2, \cdots, Q_{2^{n-2}}\big)\,,$$

where for each $1 \leqslant k \leqslant 2^{n-2}$,

$$\begin{aligned}Q_k &= \mathrm{diag}\big(-(4k-3), -(4k-2), -(4k-1), -(4k),\\ &\qquad\quad +(4k-3), +(4k-2), +(4k-1), +(4k)\big)\\ &= \mathrm{diag}\big(e^{-i\theta_{4k-3}/2}, e^{-i\theta_{4k-2}/2}, e^{-i\theta_{4k-1}/2}, e^{-i\theta_{4k}/2},\\ &\qquad\quad e^{i\theta_{4k-3}/2}, e^{i\theta_{4k-2}/2}, e^{i\theta_{4k-1}/2}, e^{i\theta_{4k}/2}\big)\end{aligned}$$

for some $\theta_1, \cdots, \theta_{2^n} \in \mathbb{R}$.

§3.6 Tensor Product of Vector Spaces

Example (Matrix representation of $F_k^n(\mathrm{R}_z)$ (cont.))

In general, for each j we have

$$A_{j-1} = \mathbf{SWAP}_{j,j+1} \cdot A_j \cdot \mathbf{SWAP}_{j,j+1}$$

and the fact that $\mathbf{SWAP}_{j,j+1} = \mathrm{I}_{2^{j-1}} \otimes \mathbf{SWAP} \otimes \mathrm{I}_{2^{n-j}}$ implies that $[A_{n-j-1}]$ is obtained by

將 $[A_{n-j}]$ 的對角線元素連續 2^{j+2} 個視為一組，每一組等分為四塊，然後中間兩塊的元素交換；

so that

$$[A_{n-j}] = \mathrm{diag}\big(-1, \cdots, -2^j, +1, \cdots, +2^j, -(2^j+1), \cdots, -2^{j+1},\\ +(2^j+1), \cdots, +2^{j+1}, \cdots\big).$$

The identity above can be proved rigorously by induction.

§3.6 Tensor Product of Vector Spaces

Definition

An $(n+1)$-qubit gate L is called a **multi-controlled rotation gate of type $F^n_{j+1}(R_{\boldsymbol{a}})$** if there exist a unit vectors $\boldsymbol{a} \in \mathbb{R}^3$ and real numbers $\theta_1, \cdots, \theta_{2^n}$ such that

$$\begin{aligned} & L(|x_0\rangle \otimes \cdots \otimes |x_n\rangle) \\ & \quad = |x_0\rangle \otimes \cdots \otimes |x_{j-1}\rangle \otimes (R_{\boldsymbol{a}}(\theta_k)|x_j\rangle) \otimes |x_{j+1}\rangle \otimes \cdots \otimes |x_n\rangle \end{aligned}$$

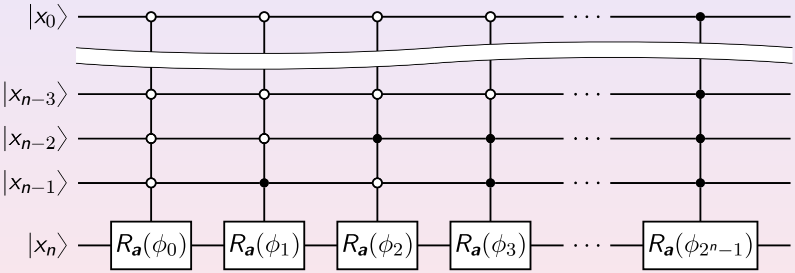
if $(x_0 \cdots x_{j-1}x_{j+1} \cdots x_n)_2 = k$, where for unit vector $\boldsymbol{a} = (a_x, a_y, a_z)$, $R_{\boldsymbol{a}}$ is a 1-qubit gate given by

$$R_{\boldsymbol{a}}(\phi) = \begin{bmatrix} \cos\frac{\phi}{2} + ia_z \sin\frac{\phi}{2} & (a_y + ia_x)\sin\frac{\phi}{2} \\ -(a_y - ia_x)\sin\frac{\phi}{2} & \cos\frac{\phi}{2} - ia_z\sin\frac{\phi}{2} \end{bmatrix}.$$

We also write $R_{\boldsymbol{a}}$ as R_y or R_z if $\boldsymbol{a} = (0, -1, 0)$ or $\boldsymbol{a} = (0, 0, -1)$.

§3.6 Tensor Product of Vector Spaces

Remark: One possible quantum circuit for a multi-controlled rotation gate of type $F^n_{n+1}(R_{\mathbf{a}})$, in term of 1-qubit quantum gate $R_{\mathbf{a}}(\phi)$, is given by



and quantum circuit for a multi-controlled rotation gate of type $F^n_j(R_{\mathbf{a}})$ can be constructed using **SWAP** gates and the quantum circuit given above.

§3.7 Unitary Decomposition

Unitary decomposition is the process of translating an arbitrary unitary gate into a specific (universal) set of single and two-qubit gates. Unitary decomposition is necessary because it is not otherwise possible to execute an arbitrary quantum gate on either a simulator or quantum accelerator.

In order to decompose all possible unitary matrices into quantum gates, a universal gate set is selected: rotations around the Y and Z axis by an arbitrary angle, the $\mathrm{R}_z(\theta)$ and $\mathrm{R}_y(\theta)$ gates, and the controlled-not, the **CNOT** gate whose matrix forms are given by

$$\mathrm{R}_y(\theta) = \begin{bmatrix} \cos\frac{\theta}{2} & -\sin\frac{\theta}{2} \\ \sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{bmatrix}, \ \mathrm{R}_z(\theta) = \begin{bmatrix} e^{-i\frac{\theta}{2}} & 0 \\ 0 & e^{i\frac{\theta}{2}} \end{bmatrix}, \ \textbf{CNOT} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

§3.7 Unitary Decomposition

Unitary decomposition is the process of translating an arbitrary unitary gate into a specific (universal) set of single and two-qubit gates. Unitary decomposition is necessary because it is not otherwise possible to execute an arbitrary quantum gate on either a simulator or quantum accelerator.

In order to decompose all possible unitary matrices into quantum gates, a universal gate set is selected: rotations around the Y and Z axis by an arbitrary angle, the $\mathrm{R}_z(\theta)$ and $\mathrm{R}_y(\theta)$ gates, and the controlled-not, the **CNOT** gate whose matrix forms are given by

$$\mathrm{R}_y(\theta) = \begin{bmatrix} \cos\frac{\theta}{2} & -\sin\frac{\theta}{2} \\ \sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{bmatrix}, \ \mathrm{R}_z(\theta) = \begin{bmatrix} e^{-i\frac{\theta}{2}} & 0 \\ 0 & e^{i\frac{\theta}{2}} \end{bmatrix}, \ \textbf{CNOT} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

§3.7 Unitary Decomposition

Unitary decomposition is the process of translating an arbitrary unitary gate into a specific (universal) set of single and two-qubit gates. Unitary decomposition is necessary because it is not otherwise possible to execute an arbitrary quantum gate on either a simulator or quantum accelerator.

In order to decompose all possible unitary matrices into quantum gates, a universal gate set is selected: rotations around the Y and Z axis by an arbitrary angle, the $\mathrm{R}_z(\theta)$ and $\mathrm{R}_y(\theta)$ gates, and the controlled-not, the **CNOT** gate whose matrix forms are given by

$$\mathrm{R}_y(\theta) = \begin{bmatrix} \cos\frac{\theta}{2} & -\sin\frac{\theta}{2} \\ \sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{bmatrix}, \ \mathrm{R}_z(\theta) = \begin{bmatrix} e^{-i\frac{\theta}{2}} & 0 \\ 0 & e^{i\frac{\theta}{2}} \end{bmatrix}, \ \textbf{CNOT} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

§3.7 Unitary Decomposition

§3.7.1 1-qubit gate decomposition

We first focus on expressing 1 qubit gates (or 2×2 unitary matrices) in terms of product of qubit gates from the set

$$\left\{ \mathrm{R}_y(\theta), \mathrm{R}_z(\theta), \mathrm{Ph}(\theta) \,\middle|\, \theta \in \mathbb{R} \right\},$$

where Ph is the global phase gate given by $\mathrm{Ph}(\theta) = \mathrm{diag}(e^{i\theta}, e^{i\theta})$.

Theorem

For every 1-qubit gate (that is, 2×2 unitary matrix) U, there exist real numbers δ, θ, ξ and η such that

$$U = \mathrm{Ph}(\delta)\mathrm{R}_z(\xi)\mathrm{R}_y(\theta)\mathrm{R}_z(\eta) = \mathrm{R}_z(\xi)\mathrm{R}_y(\theta)\mathrm{R}_z(\eta)\mathrm{Ph}(\delta)$$

$$= \begin{bmatrix} e^{i\delta} & 0 \\ 0 & e^{i\delta} \end{bmatrix} \begin{bmatrix} e^{-i\frac{\xi}{2}} & 0 \\ 0 & e^{i\frac{\xi}{2}} \end{bmatrix} \begin{bmatrix} \cos\frac{\theta}{2} & -\sin\frac{\theta}{2} \\ \sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{bmatrix} \begin{bmatrix} e^{-i\frac{\eta}{2}} & 0 \\ 0 & e^{i\frac{\eta}{2}} \end{bmatrix}.$$

§3.7 Unitary Decomposition

§3.7.1 1-qubit gate decomposition

We first focus on expressing 1 qubit gates (or 2×2 unitary matrices) in terms of product of qubit gates from the set

$$\{\mathrm{R}_y(\theta), \mathrm{R}_z(\theta), \mathrm{Ph}(\theta) \,\big|\, \theta \in \mathbb{R}\},$$

where Ph is the global phase gate given by $\mathrm{Ph}(\theta) = \mathrm{diag}(e^{i\theta}, e^{i\theta})$.

Theorem

For every 1-qubit gate (that is, 2×2 unitary matrix) U, there exist real numbers δ, θ, ξ and η such that

$$U = \mathrm{Ph}(\delta)\mathrm{R}_z(\xi)\mathrm{R}_y(\theta)\mathrm{R}_z(\eta) = \mathrm{R}_z(\xi)\mathrm{R}_y(\theta)\mathrm{R}_z(\eta)\mathrm{Ph}(\delta)$$

$$= \begin{bmatrix} e^{i\delta} & 0 \\ 0 & e^{i\delta} \end{bmatrix} \begin{bmatrix} e^{-i\frac{\xi}{2}} & 0 \\ 0 & e^{i\frac{\xi}{2}} \end{bmatrix} \begin{bmatrix} \cos\frac{\theta}{2} & -\sin\frac{\theta}{2} \\ \sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{bmatrix} \begin{bmatrix} e^{-i\frac{\eta}{2}} & 0 \\ 0 & e^{i\frac{\eta}{2}} \end{bmatrix}.$$

§3.7 Unitary Decomposition

Proof.

Let $U = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ be a 2×2 unitary matrix. Then there exists $\delta \in \mathbb{R}$ such that $\det(U) = e^{2i\delta}$. Define $V \equiv e^{-i\delta} U$. Then V is also a unitary matrix; thus using the fact that $V^\dagger = V^{-1}$ and $\det(V) = 1$ we find that V takes the form

$$V = \begin{bmatrix} \alpha & -\bar{\beta} \\ \beta & \bar{\alpha} \end{bmatrix}.$$

This further implies that U takes the form

$$U = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = e^{i\delta} \begin{bmatrix} \alpha & -\bar{\beta} \\ \beta & \bar{\alpha} \end{bmatrix}.$$

The fact that $|\alpha|^2 + |\beta|^2 = 1$ allows us to set $\alpha = e^{i\mu} \cos\frac{\theta}{2}$ and $\beta = e^{i\nu} \sin\frac{\theta}{2}$ for some μ, ν and $\theta \in \mathbb{R}$. □

§3.7 Unitary Decomposition

Proof.

Let $U = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ be a 2×2 unitary matrix. Then there exists $\delta \in \mathbb{R}$ such that $\det(U) = e^{2i\delta}$. Define $V \equiv e^{-i\delta} U$. Then V is also a unitary matrix; thus using the fact that $V^\dagger = V^{-1}$ and $\det(V) = 1$ we find that V takes the form

$$V = \begin{bmatrix} \alpha & -\bar{\beta} \\ \beta & \bar{\alpha} \end{bmatrix}.$$

This further implies that U takes the form

$$U = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = e^{i\delta} \begin{bmatrix} \alpha & -\bar{\beta} \\ \beta & \bar{\alpha} \end{bmatrix}.$$

The fact that $|\alpha|^2 + |\beta|^2 = 1$ allows us to set $\alpha = e^{i\mu} \cos \frac{\theta}{2}$ and $\beta = e^{i\nu} \sin \frac{\theta}{2}$ for some μ, ν and $\theta \in \mathbb{R}$. □

§3.7 Unitary Decomposition

Proof.

Let $U = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ be a 2×2 unitary matrix. Then there exists $\delta \in \mathbb{R}$ such that $\det(U) = e^{2i\delta}$. Define $V \equiv e^{-i\delta} U$. Then V is also a unitary matrix; thus using the fact that $V^\dagger = V^{-1}$ and $\det(V) = 1$ we find that V takes the form

$$V = \begin{bmatrix} \alpha & -\bar{\beta} \\ \beta & \bar{\alpha} \end{bmatrix}.$$

This further implies that U takes the form

$$U = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = e^{i\delta} \begin{bmatrix} \alpha & -\bar{\beta} \\ \beta & \bar{\alpha} \end{bmatrix}.$$

The fact that $|\alpha|^2 + |\beta|^2 = 1$ allows us to set $\alpha = e^{i\mu} \cos \frac{\theta}{2}$ and $\beta = e^{i\nu} \sin \frac{\theta}{2}$ for some μ, ν and $\theta \in \mathbb{R}$. □

§3.7 Unitary Decomposition

Proof.

Let $U = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ be a 2×2 unitary matrix. Then there exists $\delta \in \mathbb{R}$ such that $\det(U) = e^{2i\delta}$. Define $V \equiv e^{-i\delta} U$. Then V is also a unitary matrix; thus using the fact that $V^\dagger = V^{-1}$ and $\det(V) = 1$ we find that V takes the form

$$V = \begin{bmatrix} \alpha & -\bar{\beta} \\ \beta & \bar{\alpha} \end{bmatrix}.$$

This further implies that U takes the form

$$U = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = e^{i\delta} \begin{bmatrix} \alpha & -\bar{\beta} \\ \beta & \bar{\alpha} \end{bmatrix}.$$

The fact that $|\alpha|^2 + |\beta|^2 = 1$ allows us to set $\alpha = e^{i\mu} \cos\frac{\theta}{2}$ and $\beta = e^{i\nu} \sin\frac{\theta}{2}$ for some μ, ν and $\theta \in \mathbb{R}$. □

§3.7 Unitary Decomposition

Proof (cont.)

Let $\xi = \nu - \mu$ and $\eta = -\mu - \nu$. Then

$$
\begin{aligned}
&\mathrm{R}_z(\xi)\mathrm{R}_y(\theta)\mathrm{R}_z(\eta) \\
&\quad = \begin{bmatrix} e^{-i\frac{\xi}{2}} & 0 \\ 0 & e^{i\frac{\xi}{2}} \end{bmatrix} \begin{bmatrix} \cos\frac{\theta}{2} & -\sin\frac{\theta}{2} \\ \sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{bmatrix} \begin{bmatrix} e^{-i\frac{\eta}{2}} & 0 \\ 0 & e^{i\frac{\eta}{2}} \end{bmatrix} \\
&\quad = \begin{bmatrix} e^{-i\frac{\xi}{2}}\cos\frac{\theta}{2} & -e^{-i\frac{\xi}{2}}\sin\frac{\theta}{2} \\ e^{i\frac{\xi}{2}}\sin\frac{\theta}{2} & e^{i\frac{\xi}{2}}\cos\frac{\theta}{2} \end{bmatrix} \begin{bmatrix} e^{-i\frac{\eta}{2}} & 0 \\ 0 & e^{i\frac{\eta}{2}} \end{bmatrix} \\
&\quad = \begin{bmatrix} e^{i\frac{\xi+\eta}{2}}\cos\frac{\theta}{2} & -e^{-i\frac{\xi-\eta}{2}}\sin\frac{\theta}{2} \\ e^{i\frac{\xi-\eta}{2}}\sin\frac{\theta}{2} & e^{-i\frac{\xi+\eta}{2}}\cos\frac{\theta}{2} \end{bmatrix} = \begin{bmatrix} \alpha & -\bar{\beta} \\ \beta & \bar{\alpha} \end{bmatrix}
\end{aligned}
$$

which concludes the theorem. □

§3.7 Unitary Decomposition

- **Algorithm of 1-qubit gate decomposition**:

Let U be a 1-qubit gate (or equivalently, 2×2 unitary matrix).

Step 1: Find $\delta \in \mathbb{R}$ such that $\det(U) = e^{2i\delta}$.

Step 2: Find μ, ν, θ such that

$$e^{i\mu} \cos\frac{\theta}{2} = ae^{-i\delta} \quad \text{and} \quad e^{i\nu} \sin\frac{\theta}{2} = ce^{-i\delta}\,.$$

Step 3: $U = \mathrm{Ph}(\delta)\mathrm{R}_z(\nu - \mu)\mathrm{R}_y(\theta)\mathrm{R}_z(-\mu - \nu)$.

Example

$$\begin{aligned} \mathrm{X} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} &= \mathrm{Ph}\big(-\tfrac{\pi}{2}\big)\mathrm{R}_z\big(\tfrac{\pi}{2}\big)\mathrm{R}_y(\pi)\mathrm{R}_z\big(-\tfrac{\pi}{2}\big) \\ &= \begin{bmatrix} e^{-i\pi/2} & 0 \\ 0 & e^{-i\pi/2} \end{bmatrix} \begin{bmatrix} e^{-i\pi/4} & 0 \\ 0 & e^{i\pi/4} \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} e^{i\pi/4} & 0 \\ 0 & e^{-i\pi/4} \end{bmatrix}. \end{aligned}$$

§3.7 Unitary Decomposition

- **Algorithm of 1-qubit gate decomposition**:

Let U be a 1-qubit gate (or equivalently, 2×2 unitary matrix).

Step 1: Find $\delta \in \mathbb{R}$ such that $\det(U) = e^{2i\delta}$.

Step 2: Find μ, ν, θ such that

$$e^{i\mu} \cos \frac{\theta}{2} = ae^{-i\delta} \quad \text{and} \quad e^{i\nu} \sin \frac{\theta}{2} = ce^{-i\delta}.$$

Step 3: $U = \mathrm{Ph}(\delta)\mathrm{R}_z(\nu - \mu)\mathrm{R}_y(\theta)\mathrm{R}_z(-\mu - \nu)$.

Example

$$\mathrm{X} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \mathrm{Ph}\big(-\frac{\pi}{2}\big)\mathrm{R}_z\big(\frac{\pi}{2}\big)\mathrm{R}_y(\pi)\mathrm{R}_z\big(-\frac{\pi}{2}\big)$$

$$= \begin{bmatrix} e^{-i\pi/2} & 0 \\ 0 & e^{-i\pi/2} \end{bmatrix} \begin{bmatrix} e^{-i\pi/4} & 0 \\ 0 & e^{i\pi/4} \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} e^{i\pi/4} & 0 \\ 0 & e^{-i\pi/4} \end{bmatrix}.$$

§3.7 Unitary Decomposition

§3.7.2 Singular value decomposition (SVD)

Recall the spectral theorem from linear algebra given below:

Theorem (Spectral)

Let A be a Hermitian matrix; that is, $A = A^\dagger$. Then there exists unitary matrix U and a **real** *diagonal matrix D such that $A = UDU^\dagger$.*

We note that the columns of U are eigenvectors of A and the diagonal elements of D are eigenvalues of A.

Remark: The spectral theorem extends to a more general class of matrices, the normal matrices. One can show that A is normal (that is, $AA^\dagger = A^\dagger A$) if and only if there exists a unitary matrix U and a diagonal matrix D such that $A = UDU^\dagger$. Here the diagonal matrix D can be complex.

§3.7 Unitary Decomposition

§3.7.2 Singular value decomposition (SVD)

Recall the spectral theorem from linear algebra given below:

Theorem (Spectral)

Let A be a Hermitian matrix; that is, $A = A^\dagger$. Then there exists unitary matrix U and a **real** *diagonal matrix D such that $A = UDU^\dagger$.*

We note that the columns of U are eigenvectors of A and the diagonal elements of D are eigenvalues of A.

Remark: The spectral theorem extends to a more general class of matrices, the normal matrices. One can show that A is normal (that is, $AA^\dagger = A^\dagger A$) if and only if there exists a unitary matrix U and a diagonal matrix D such that $A = UDU^\dagger$. Here the diagonal matrix D can be complex.

§3.7 Unitary Decomposition

Let A be a complex $m \times n$ matrix. Then $A^\dagger A \in \mathbb{C}^{n\times n}$ and $AA^\dagger \in \mathbb{C}^{m\times m}$. Moreover,

1. $A^\dagger A$ and $AA^\dagger$ are both hermitian since
$$(A^\dagger A)^\dagger = A^\dagger (A^\dagger)^\dagger = A^\dagger A \quad \text{and} \quad (AA^\dagger)^\dagger = (A^\dagger)^\dagger A^\dagger = AA^\dagger .$$
2. $A^\dagger A$ and $AA^\dagger$ are both positive semi-definite since
$$\langle \boldsymbol{x}, A^\dagger A\boldsymbol{x}\rangle = \langle A\boldsymbol{x}, A\boldsymbol{x}\rangle = \|A\boldsymbol{x}\|^2 \geqslant 0 \qquad \forall\, \boldsymbol{x} \in \mathbb{C}^n$$
and
$$\langle \boldsymbol{x}, AA^\dagger\boldsymbol{x}\rangle = \langle A^\dagger\boldsymbol{x}, A^\dagger\boldsymbol{x}\rangle = \|A^\dagger\boldsymbol{x}\|^2 \geqslant 0 \qquad \forall\, \boldsymbol{x} \in \mathbb{C}^m .$$

Therefore, the Spectral Theorem implies that there exist $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n \geqslant 0$ and an orthonormal basis $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n\}$ of $\mathbb{C}^n$ such that
$$A^\dagger A\boldsymbol{v}_k = \lambda_k \boldsymbol{v}_k \qquad \forall\, 1 \leqslant k \leqslant n .$$

§3.7 Unitary Decomposition

Let A be a complex $m \times n$ matrix. Then $A^\dagger A \in \mathbb{C}^{n\times n}$ and $AA^\dagger \in \mathbb{C}^{m\times m}$. Moreover,

1. $A^\dagger A$ and $AA^\dagger$ are both hermitian since
$$(A^\dagger A)^\dagger = A^\dagger (A^\dagger)^\dagger = A^\dagger A \quad \text{and} \quad (AA^\dagger)^\dagger = (A^\dagger)^\dagger A^\dagger = AA^\dagger \,.$$
2. $A^\dagger A$ and $AA^\dagger$ are both positive semi-definite since
$$\langle \boldsymbol{x}, A^\dagger A\boldsymbol{x}\rangle = \langle A\boldsymbol{x}, A\boldsymbol{x}\rangle = \|A\boldsymbol{x}\|^2 \geqslant 0 \qquad \forall\, \boldsymbol{x} \in \mathbb{C}^n$$
and
$$\langle \boldsymbol{x}, AA^\dagger\boldsymbol{x}\rangle = \langle A^\dagger\boldsymbol{x}, A^\dagger\boldsymbol{x}\rangle = \|A^\dagger\boldsymbol{x}\|^2 \geqslant 0 \qquad \forall\, \boldsymbol{x} \in \mathbb{C}^m \,.$$

Therefore, the Spectral Theorem implies that there exist $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n \geqslant 0$ and an orthonormal basis $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n\}$ of $\mathbb{C}^n$ such that
$$A^\dagger A\boldsymbol{v}_k = \lambda_k \boldsymbol{v}_k \qquad \forall\, 1 \leqslant k \leqslant n \,.$$

§3.7 Unitary Decomposition

Let A be a complex $m \times n$ matrix. Then $A^\dagger A \in \mathbb{C}^{n\times n}$ and $AA^\dagger \in \mathbb{C}^{m\times m}$. Moreover,

1. $A^\dagger A$ and $AA^\dagger$ are both hermitian since
$$(A^\dagger A)^\dagger = A^\dagger (A^\dagger)^\dagger = A^\dagger A \quad \text{and} \quad (AA^\dagger)^\dagger = (A^\dagger)^\dagger A^\dagger = AA^\dagger\,.$$
2. $A^\dagger A$ and $AA^\dagger$ are both positive semi-definite since
$$\langle \boldsymbol{x}, A^\dagger A\boldsymbol{x}\rangle = \langle A\boldsymbol{x}, A\boldsymbol{x}\rangle = \|A\boldsymbol{x}\|^2 \geqslant 0 \qquad \forall\, \boldsymbol{x} \in \mathbb{C}^n$$
and
$$\langle \boldsymbol{x}, AA^\dagger\boldsymbol{x}\rangle = \langle A^\dagger\boldsymbol{x}, A^\dagger\boldsymbol{x}\rangle = \|A^\dagger\boldsymbol{x}\|^2 \geqslant 0 \qquad \forall\, \boldsymbol{x} \in \mathbb{C}^m\,.$$

Therefore, the Spectral Theorem implies that there exist $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n \geqslant 0$ and an orthonormal basis $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n\}$ of $\mathbb{C}^n$ such that
$$A^\dagger A\boldsymbol{v}_k = \lambda_k \boldsymbol{v}_k \qquad \forall\, 1 \leqslant k \leqslant n\,.$$

§3.7 Unitary Decomposition

Let A be a complex $m \times n$ matrix. Then $A^\dagger A \in \mathbb{C}^{n\times n}$ and $AA^\dagger \in \mathbb{C}^{m\times m}$. Moreover,

1. $A^\dagger A$ and $AA^\dagger$ are both hermitian since
$$(A^\dagger A)^\dagger = A^\dagger (A^\dagger)^\dagger = A^\dagger A \quad \text{and} \quad (AA^\dagger)^\dagger = (A^\dagger)^\dagger A^\dagger = AA^\dagger\,.$$
2. $A^\dagger A$ and $AA^\dagger$ are both positive semi-definite since
$$\langle \boldsymbol{x}, A^\dagger A\boldsymbol{x}\rangle = \langle A\boldsymbol{x}, A\boldsymbol{x}\rangle = \|A\boldsymbol{x}\|^2 \geqslant 0 \qquad \forall\, \boldsymbol{x} \in \mathbb{C}^n$$
and
$$\langle \boldsymbol{x}, AA^\dagger\boldsymbol{x}\rangle = \langle A^\dagger\boldsymbol{x}, A^\dagger\boldsymbol{x}\rangle = \|A^\dagger\boldsymbol{x}\|^2 \geqslant 0 \qquad \forall\, \boldsymbol{x} \in \mathbb{C}^m\,.$$

Therefore, the Spectral Theorem implies that there exist $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n \geqslant 0$ and an orthonormal basis $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n\}$ of $\mathbb{C}^n$ such that
$$A^\dagger A\boldsymbol{v}_k = \lambda_k \boldsymbol{v}_k \qquad \forall\, 1 \leqslant k \leqslant n\,.$$

§3.7 Unitary Decomposition

Let $\sigma_k = \sqrt{\lambda_k}$, and $r = \#\big\{1 \leqslant k \leqslant n \,\big|\, \lambda_k > 0\big\}$; that is, $A^\dagger A$ has r non-zero eigenvalues. For $1 \leqslant k \leqslant r$, define

$$\boldsymbol{u}_k = \frac{1}{\sigma_k} A \boldsymbol{v}_k \,.$$

1. $\boldsymbol{u}_k \neq \mathbf{0}$ for all $1 \leqslant k \leqslant r$. Moreover,

$$\|A\boldsymbol{v}_j\|^2 = \langle A\boldsymbol{v}_j, A\boldsymbol{v}_j\rangle = \langle \boldsymbol{v}_j, A^\dagger A\boldsymbol{v}_j\rangle = \langle \boldsymbol{v}_j, \lambda_j \boldsymbol{v}_j\rangle = \lambda_j\,;$$

thus the fact that $A^\dagger A$ and A have the same null space implies that $\big\{\boldsymbol{v}_{r+1}, \cdots, \boldsymbol{v}_n\big\}$ is an orthonormal basis of the null space of A.

2. $\big\{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_r\big\}$ is an orthonormal set since

$$\langle \boldsymbol{u}_k, \boldsymbol{u}_\ell\rangle = \frac{1}{\sigma_k\sigma_\ell}\langle A\boldsymbol{v}_k, A\boldsymbol{v}_\ell\rangle = \frac{1}{\sigma_k\sigma_\ell}\langle \boldsymbol{v}_k, A^\dagger A\boldsymbol{v}_\ell\rangle = \frac{\lambda_\ell}{\sigma_k\sigma_\ell}\langle \boldsymbol{v}_k, \boldsymbol{v}_\ell\rangle$$
$$= \frac{\sigma_\ell}{\sigma_k}\delta_{k\ell}\,.$$

§3.7 Unitary Decomposition

Let $\sigma_k = \sqrt{\lambda_k}$, and $r = \#\{1 \leqslant k \leqslant n \,|\, \lambda_k > 0\}$; that is, $A^\dagger A$ has r non-zero eigenvalues. For $1 \leqslant k \leqslant r$, define

$$\boldsymbol{u}_k = \frac{1}{\sigma_k} A\boldsymbol{v}_k \,.$$

1. $\boldsymbol{u}_k \neq \mathbf{0}$ for all $1 \leqslant k \leqslant r$. Moreover,

$$\|A\boldsymbol{v}_j\|^2 = \langle A\boldsymbol{v}_j, A\boldsymbol{v}_j\rangle = \langle \boldsymbol{v}_j, A^\dagger A\boldsymbol{v}_j\rangle = \langle \boldsymbol{v}_j, \lambda_j \boldsymbol{v}_j\rangle = \lambda_j \,;$$

   thus the fact that $A^\dagger A$ and A have the same null space implies that $\{\boldsymbol{v}_{r+1}, \cdots, \boldsymbol{v}_n\}$ is an orthonormal basis of the null space of A.

2. $\{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_r\}$ is an orthonormal set since

$$\langle \boldsymbol{u}_k, \boldsymbol{u}_\ell\rangle = \frac{1}{\sigma_k\sigma_\ell}\langle A\boldsymbol{v}_k, A\boldsymbol{v}_\ell\rangle = \frac{1}{\sigma_k\sigma_\ell}\langle \boldsymbol{v}_k, A^\dagger A\boldsymbol{v}_\ell\rangle = \frac{\lambda_\ell}{\sigma_k\sigma_\ell}\langle \boldsymbol{v}_k, \boldsymbol{v}_\ell\rangle$$
$$= \frac{\sigma_\ell}{\sigma_k}\delta_{k\ell} \,.$$

§3.7 Unitary Decomposition

Let $\sigma_k = \sqrt{\lambda_k}$, and $r = \#\{1 \leqslant k \leqslant n \,|\, \lambda_k > 0\}$; that is, $A^\dagger A$ has r non-zero eigenvalues. For $1 \leqslant k \leqslant r$, define

$$\boldsymbol{u}_k = \frac{1}{\sigma_k} A\boldsymbol{v}_k\,.$$

1. $\boldsymbol{u}_k \neq \mathbf{0}$ for all $1 \leqslant k \leqslant r$. Moreover,

   $$\|A\boldsymbol{v}_j\|^2 = \langle A\boldsymbol{v}_j, A\boldsymbol{v}_j\rangle = \langle \boldsymbol{v}_j, A^\dagger A\boldsymbol{v}_j\rangle = \langle \boldsymbol{v}_j, \lambda_j \boldsymbol{v}_j\rangle = \lambda_j\,;$$

   thus the fact that $A^\dagger A$ and A have the same null space implies that $\{\boldsymbol{v}_{r+1}, \cdots, \boldsymbol{v}_n\}$ is an orthonormal basis of the null space of A.

2. $\{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_r\}$ is an orthonormal set since

   $$\langle \boldsymbol{u}_k, \boldsymbol{u}_\ell\rangle = \frac{1}{\sigma_k\sigma_\ell}\langle A\boldsymbol{v}_k, A\boldsymbol{v}_\ell\rangle = \frac{1}{\sigma_k\sigma_\ell}\langle \boldsymbol{v}_k, A^\dagger A\boldsymbol{v}_\ell\rangle = \frac{\lambda_\ell}{\sigma_k\sigma_\ell}\langle \boldsymbol{v}_k, \boldsymbol{v}_\ell\rangle$$
   $$= \frac{\sigma_\ell}{\sigma_k}\delta_{k\ell}\,.$$

§3.7 Unitary Decomposition

Let $\sigma_k = \sqrt{\lambda_k}$, and $r = \#\{1 \leqslant k \leqslant n \,|\, \lambda_k > 0\}$; that is, $A^\dagger A$ has r non-zero eigenvalues. For $1 \leqslant k \leqslant r$, define

$$\boldsymbol{u}_k = \frac{1}{\sigma_k} A\boldsymbol{v}_k \,.$$

1. $\boldsymbol{u}_k \neq \mathbf{0}$ for all $1 \leqslant k \leqslant r$. Moreover,

$$\|A\boldsymbol{v}_j\|^2 = \langle A\boldsymbol{v}_j, A\boldsymbol{v}_j\rangle = \langle \boldsymbol{v}_j, A^\dagger A\boldsymbol{v}_j\rangle = \langle \boldsymbol{v}_j, \lambda_j \boldsymbol{v}_j\rangle = \lambda_j \,;$$

thus the fact that $A^\dagger A$ and A have the same null space implies that $\{\boldsymbol{v}_{r+1}, \cdots, \boldsymbol{v}_n\}$ is an orthonormal basis of the null space of A.

2. $\{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_r\}$ is an orthonormal set since

$$\langle \boldsymbol{u}_k, \boldsymbol{u}_\ell\rangle = \frac{1}{\sigma_k\sigma_\ell}\langle A\boldsymbol{v}_k, A\boldsymbol{v}_\ell\rangle = \frac{1}{\sigma_k\sigma_\ell}\langle \boldsymbol{v}_k, A^\dagger A\boldsymbol{v}_\ell\rangle = \frac{\lambda_\ell}{\sigma_k\sigma_\ell}\langle \boldsymbol{v}_k, \boldsymbol{v}_\ell\rangle$$

$$= \frac{\sigma_\ell}{\sigma_k}\delta_{k\ell} \,.$$

§3.7 Unitary Decomposition

3. $\{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_r\}$ are eigenvectors of $AA^\dagger$ with corresponding eigenvalues $\lambda_1, \cdots, \lambda_r$ since for $1 \leqslant j \leqslant r$,

$$AA^\dagger \boldsymbol{u}_j = AA^\dagger\Big(\frac{1}{\sigma_j}A\boldsymbol{v}_j\Big) = \frac{1}{\sigma_j}AA^\dagger A\boldsymbol{v}_j = \frac{1}{\sigma_j}A(\lambda_j\boldsymbol{v}_j) = \frac{\lambda_j}{\sigma_j}A\boldsymbol{v}_j \\ = \lambda_j\boldsymbol{u}_j\,.$$

By the fact that

$$r = \operatorname{rank}(A^\dagger A) = \operatorname{rank}(A) = \operatorname{rank}(A^\dagger) = \operatorname{rank}(AA^\dagger)\,,$$

the nullity (that is, the dimension of the null space) of $AA^\dagger$ is $m - r$; thus there exist an orthonormal set $\{\boldsymbol{u}_{r+1}, \cdots, \boldsymbol{u}_m\}$ in the null space of $AA^\dagger$. Then

$$AA^\dagger\boldsymbol{u}_j = \sigma_j^2\boldsymbol{u}_j \qquad \forall\, 1 \leqslant j \leqslant m\,.$$

Since $\{\boldsymbol{u}_{r+1}, \cdots, \boldsymbol{u}_m\}$ are eigenvectors of $AA^\dagger$ (corresponding to eigenvalue 0), $\{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_m\}$ is an orthonormal basis of $\mathbb{C}^m$.

§3.7 Unitary Decomposition

3. $\{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_r\}$ are eigenvectors of $AA^\dagger$ with corresponding eigenvalues $\lambda_1, \cdots, \lambda_r$ since for $1 \leqslant j \leqslant r$,

$$AA^\dagger \boldsymbol{u}_j = AA^\dagger\Big(\frac{1}{\sigma_j}A\boldsymbol{v}_j\Big) = \frac{1}{\sigma_j}AA^\dagger A\boldsymbol{v}_j = \frac{1}{\sigma_j}A(\lambda_j\boldsymbol{v}_j) = \frac{\lambda_j}{\sigma_j}A\boldsymbol{v}_j$$
$$= \lambda_j\boldsymbol{u}_j\,.$$

By the fact that

$$r = \text{rank}(A^\dagger A) = \text{rank}(A) = \text{rank}(A^\dagger) = \text{rank}(AA^\dagger)\,,$$

the nullity (that is, the dimension of the null space) of $AA^\dagger$ is $m - r$; thus there exist an orthonormal set $\{\boldsymbol{u}_{r+1}, \cdots, \boldsymbol{u}_m\}$ in the null space of $AA^\dagger$. Then

$$AA^\dagger \boldsymbol{u}_j = \sigma_j^2 \boldsymbol{u}_j \qquad \forall\, 1 \leqslant j \leqslant m\,.$$

Since $\{\boldsymbol{u}_{r+1}, \cdots, \boldsymbol{u}_m\}$ are eigenvectors of $AA^\dagger$ (corresponding to eigenvalue 0), $\{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_m\}$ is an orthonormal basis of $\mathbb{C}^m$.

§3.7 Unitary Decomposition

3. $\{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_r\}$ are eigenvectors of $AA^\dagger$ with corresponding eigenvalues $\lambda_1, \cdots, \lambda_r$ since for $1 \leqslant j \leqslant r$,

$$AA^\dagger \boldsymbol{u}_j = AA^\dagger\Big(\frac{1}{\sigma_j}A\boldsymbol{v}_j\Big) = \frac{1}{\sigma_j}AA^\dagger A\boldsymbol{v}_j = \frac{1}{\sigma_j}A(\lambda_j \boldsymbol{v}_j) = \frac{\lambda_j}{\sigma_j}A\boldsymbol{v}_j \\ = \lambda_j \boldsymbol{u}_j\,.$$

By the fact that

$$r = \operatorname{rank}(A^\dagger A) = \operatorname{rank}(A) = \operatorname{rank}(A^\dagger) = \operatorname{rank}(AA^\dagger)\,,$$

the nullity (that is, the dimension of the null space) of $AA^\dagger$ is $m-r$; thus there exist an orthonormal set $\{\boldsymbol{u}_{r+1}, \cdots, \boldsymbol{u}_m\}$ in the null space of $AA^\dagger$. Then

$$AA^\dagger \boldsymbol{u}_j = \sigma_j^2 \boldsymbol{u}_j \qquad \forall\, 1 \leqslant j \leqslant m\,.$$

Since $\{\boldsymbol{u}_{r+1}, \cdots, \boldsymbol{u}_m\}$ are eigenvectors of $AA^\dagger$ (corresponding to eigenvalue 0), $\{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_m\}$ is an orthonormal basis of $\mathbb{C}^m$.

§3.7 Unitary Decomposition

Let $U = \begin{bmatrix} \boldsymbol{u}_1 \vdots \boldsymbol{u}_2 \vdots \cdots \vdots \boldsymbol{u}_m \end{bmatrix}$ and $V = \begin{bmatrix} \boldsymbol{v}_1 \vdots \boldsymbol{v}_2 \vdots \cdots \vdots \boldsymbol{v}_n \end{bmatrix}$, as well as

$$\Sigma = \begin{bmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & \ddots \end{bmatrix}.$$

Then

$$\begin{aligned} AV &= A\begin{bmatrix} \boldsymbol{v}_1 \vdots \boldsymbol{v}_2 \vdots \cdots \vdots \boldsymbol{v}_n \end{bmatrix} = \begin{bmatrix} A\boldsymbol{v}_1 \vdots A\boldsymbol{v}_2 \vdots \cdots \vdots A\boldsymbol{v}_n \end{bmatrix} \\ &= \begin{bmatrix} \sigma_1\boldsymbol{u}_1 \vdots \sigma_2\boldsymbol{u}_2 \vdots \cdots \vdots \sigma_r\boldsymbol{u}_r \vdots \mathbf{0} \vdots \cdots \vdots \mathbf{0} \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{u}_1 \vdots \boldsymbol{u}_2 \vdots \cdots \vdots \boldsymbol{u}_m \end{bmatrix} \begin{bmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & \ddots \end{bmatrix} = U\Sigma\,. \end{aligned}$$

§3.7 Unitary Decomposition

The numbers $\sigma_1, \sigma_2, \cdots, \sigma_n$ are called the ***singular values*** of A.

The fact that U and V are unitary shows the following

Theorem

Let A be a complex $m \times n$ matrix. Then there exist unitary matrices $U \in \mathbb{C}^{m\times m}$ and $V \in \mathbb{C}^{n\times n}$ as well as an $m \times n$ matrix Σ of the form

$$\Sigma = \begin{bmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & \ddots \end{bmatrix},$$

where $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r > 0$, such that $A = U\Sigma V^\dagger$.

Remark: The decomposition $A = U\Sigma V^\dagger$ in the theorem above is called a ***singular value decomposition*** of A. We note that the singular decomposition of A is not unique.

§3.7 Unitary Decomposition

The numbers $\sigma_1, \sigma_2, \cdots, \sigma_n$ are called the ***singular values*** of A. The fact that U and V are unitary shows the following

Theorem

Let A be a complex $m \times n$ matrix. Then there exist unitary matrices $U \in \mathbb{C}^{m\times m}$ and $V \in \mathbb{C}^{n\times n}$ as well as an $m \times n$ matrix Σ of the form

$$\Sigma = \begin{bmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & \ddots \end{bmatrix},$$

where $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r > 0$, such that $A = U\Sigma V^\dagger$.

Remark: The decomposition $A = U\Sigma V^\dagger$ in the theorem above is called a ***singular value decomposition*** of A. We note that the singular decomposition of A is not unique.

§3.7 Unitary Decomposition

The numbers $\sigma_1, \sigma_2, \cdots, \sigma_n$ are called the ***singular values*** of A. The fact that U and V are unitary shows the following

Theorem

Let A be a complex $m \times n$ matrix. Then there exist unitary matrices $U \in \mathbb{C}^{m\times m}$ and $V \in \mathbb{C}^{n\times n}$ as well as an $m \times n$ matrix Σ of the form

$$\Sigma = \begin{bmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & \ddots \end{bmatrix},$$

where $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r > 0$, such that $A = U\Sigma V^\dagger$.

Remark: The decomposition $A = U\Sigma V^\dagger$ in the theorem above is called a ***singular value decomposition*** of A. We note that the singular decomposition of A is not unique.

§3.7 Unitary Decomposition

§3.7.3 CS decomposition

Theorem (Cosine-Sine decomposition)

For any 2×2 *partitioning*

$$Q = \begin{array}{cc} c_1 & c_2 \\ \left[\begin{array}{cc} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{array}\right] & \begin{array}{c} r_1 \\ r_2 \end{array} \end{array} \qquad n = c_1 + c_2 = r_1 + r_2\,, \tag{6}$$

of an $n \times n$ *unitary matrix* Q*, there exist unitary matrices* U_1*,* U_2*,* V_1*,* V_2 *such that*

$$U^\dagger Q V = \begin{bmatrix} U_1^\dagger & 0 \\ 0 & U_2^\dagger \end{bmatrix} \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \begin{bmatrix} V_1 & 0 \\ 0 & V_2 \end{bmatrix} = \begin{bmatrix} \mathrm{I} & & & \mathrm{O}_s^\dagger & & \\ & \mathrm{C} & & & -\mathrm{S} & \\ & & \mathrm{O}_c & & & -\mathrm{I} \\ \mathrm{O}_s & & & \mathrm{I} & & \\ & \mathrm{S} & & & \mathrm{C} & \\ & & \mathrm{I} & & & \mathrm{O}_c^\dagger \end{bmatrix},$$

where C *and* S *are diagonal matrices taking the form*

§3.7 Unitary Decomposition

Theorem (Cosine-Sine decomposition (cont.))

where C *and* S *are diagonal matrices taking the form*

$$\mathrm{C} = \mathrm{diag}(\gamma_1, \gamma_2, \cdots, \gamma_s), \qquad 1 > \gamma_1 \geqslant \gamma_2 \geqslant \cdots \geqslant \gamma_s > 0,$$

$$\mathrm{S} = \mathrm{diag}(\sigma_1, \sigma_2, \cdots, \sigma_s), \qquad 0 < \sigma_1 \leqslant \sigma_2 \leqslant \cdots \leqslant \sigma_s < 1,$$

and satisfying $\mathrm{C}^2 + \mathrm{S}^2 = \mathrm{I}$, *and* O_s, O_c *are matrices of zeros, and depending on Q and the partition, may have no row or no columns. Some of the identity matrices may be nonexistent, and no two of them need be equal. The four* C *and* S *matrices are square with the same dimension, and may be nonexistent.*

Remark: Since $\mathrm{C}^2 + \mathrm{S}^2 = \mathrm{I}$, there exists $\theta_1, \cdots, \theta_s$ such that $\gamma_k = \cos\theta_k$ and $\sigma_k = \sin\theta_k$ for all $1 \leqslant k \leqslant s$. This explains the name of the "cosine-sine" decomposition.

§3.7 Unitary Decomposition

Theorem (Cosine-Sine decomposition (cont.))

where C *and* S *are diagonal matrices taking the form*

$$\mathrm{C} = \mathrm{diag}(\gamma_1, \gamma_2, \cdots, \gamma_s)\,, \qquad 1 > \gamma_1 \geqslant \gamma_2 \geqslant \cdots \geqslant \gamma_s > 0\,,$$

$$\mathrm{S} = \mathrm{diag}(\sigma_1, \sigma_2, \cdots, \sigma_s)\,, \qquad 0 < \sigma_1 \leqslant \sigma_2 \leqslant \cdots \leqslant \sigma_s < 1\,,$$

and satisfying $\mathrm{C}^2 + \mathrm{S}^2 = \mathrm{I}$, *and* O_s, O_c *are matrices of zeros, and depending on Q and the partition, may have no row or no columns. Some of the identity matrices may be nonexistent, and no two of them need be equal. The four* C *and* S *matrices are square with the same dimension, and may be nonexistent.*

Remark: Since $\mathrm{C}^2 + \mathrm{S}^2 = \mathrm{I}$, there exists $\theta_1, \cdots, \theta_s$ such that $\gamma_k = \cos\theta_k$ and $\sigma_k = \sin\theta_k$ for all $1 \leqslant k \leqslant s$. This explains the name of the "cosine-sine" decomposition.

§3.7 Unitary Decomposition

Proof.

First we note that

$$\begin{bmatrix} U_1^\dagger & 0 \\ 0 & U_2^\dagger \end{bmatrix} \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \begin{bmatrix} V_1 & 0 \\ 0 & V_2 \end{bmatrix} = \begin{bmatrix} U_1^\dagger Q_{11} V_1 & U_1^\dagger Q_{12} V_2 \\ U_2^\dagger Q_{21} V_1 & U_2^\dagger Q_{22} V_2 \end{bmatrix}.$$

Choose unitary matrices U_1 and V_1 to give the usual singular value decomposition of Q_{11}, resulting in D_{11}. Choose unitary matrices U_2 and V_2 so that $D_{21} = U_2^\dagger Q_{21} V_1$ is lower triangular with **non-negative** real entries on the diagonals ending in the bottom right corners and $D_{12} = U_1^\dagger Q_{12} V_2$ is upper triangular with **non-positive** real entries on the diagonals ending in the bottom right corners. Define

$$D = \begin{bmatrix} U_1^\dagger & 0 \\ 0 & U_2^\dagger \end{bmatrix} \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \begin{bmatrix} V_1 & 0 \\ 0 & V_2 \end{bmatrix} = \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix}. \qquad (7)$$

□

§3.7 Unitary Decomposition

Proof.

First we note that

$$\begin{bmatrix} U_1^\dagger & 0 \\ 0 & U_2^\dagger \end{bmatrix} \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \begin{bmatrix} V_1 & 0 \\ 0 & V_2 \end{bmatrix} = \begin{bmatrix} U_1^\dagger Q_{11} V_1 & U_1^\dagger Q_{12} V_2 \\ U_2^\dagger Q_{21} V_1 & U_2^\dagger Q_{22} V_2 \end{bmatrix}.$$

Choose unitary matrices U_1 and V_1 to give the usual singular value decomposition of Q_{11}, resulting in D_{11}. Choose unitary matrices U_2 and V_2 so that $D_{21} = U_2^\dagger Q_{21} V_1$ is lower triangular with **non-negative** real entries on the diagonals ending in the bottom right corners and $D_{12} = U_1^\dagger Q_{12} V_2$ is upper triangular with **non-positive** real entries on the diagonals ending in the bottom right corners.

Define

$$D = \begin{bmatrix} U_1^\dagger & 0 \\ 0 & U_2^\dagger \end{bmatrix} \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \begin{bmatrix} V_1 & 0 \\ 0 & V_2 \end{bmatrix} = \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix}. \qquad (7)$$

□

§3.7 Unitary Decomposition

Proof.

First we note that

$$\begin{bmatrix} U_1^\dagger & 0 \\ 0 & U_2^\dagger \end{bmatrix} \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \begin{bmatrix} V_1 & 0 \\ 0 & V_2 \end{bmatrix} = \begin{bmatrix} U_1^\dagger Q_{11} V_1 & U_1^\dagger Q_{12} V_2 \\ U_2^\dagger Q_{21} V_1 & U_2^\dagger Q_{22} V_2 \end{bmatrix}.$$

Choose unitary matrices U_1 and V_1 to give the usual singular value decomposition of Q_{11}, resulting in D_{11}. Choose unitary matrices U_2 and V_2 so that $D_{21} = U_2^\dagger Q_{21} V_1$ is lower triangular with **non-negative** real entries on the diagonals ending in the bottom right corners and $D_{12} = U_1^\dagger Q_{12} V_2$ is upper triangular with **non-positive** real entries on the diagonals ending in the bottom right corners. Define

$$D = \begin{bmatrix} U_1^\dagger & 0 \\ 0 & U_2^\dagger \end{bmatrix} \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \begin{bmatrix} V_1 & 0 \\ 0 & V_2 \end{bmatrix} = \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix}. \tag{7}$$

□

§3.7 Unitary Decomposition

Proof (cont.)

Then D is unitary; thus the fact that any column (or row) of D has unit length implies that no singular value of D_{11} can exceed 1.

Therefore, D_{11} takes the form

$$D_{11} = \begin{bmatrix} \mathrm{I}_{k\times k} & & \\ & \mathrm{C}_{s\times s} & \\ & & \mathrm{O}_{p\times q} \end{bmatrix}$$

for some C taking the desired form, and the orthogonality of columns of D and the orthogonality of rows of D further show that D_{21} and D_{12} must take the form

$$D_{12} = \begin{bmatrix} \mathrm{O}_{k\times(c_2-s-p)} & & \\ & -\mathrm{S}_{s\times s} & \\ & & -\mathrm{I}_{p\times p} \end{bmatrix}, \quad D_{21} = \begin{bmatrix} \mathrm{O}_{(c_1-s-q)\times k} & & \\ & \mathrm{S}_{s\times s} & \\ & & \mathrm{I}_{q\times q} \end{bmatrix},$$

where $p = r_1 - k - s$ and $q = c_1 - k - s$. □

§3.7 Unitary Decomposition

Proof (cont.)

Then D is unitary; thus the fact that any column (or row) of D has unit length implies that no singular value of D_{11} can exceed 1. Therefore, D_{11} takes the form

$$D_{11} = \begin{bmatrix} \mathrm{I}_{k\times k} & & \\ & \mathrm{C}_{s\times s} & \\ & & \mathrm{O}_{p\times q} \end{bmatrix}$$

for some C taking the desired form, and the orthogonality of columns of D and the orthogonality of rows of D further show that D_{21} and D_{12} must take the form

$$D_{12} = \begin{bmatrix} \mathrm{O}_{k\times(c_2-s-p)} & & \\ & -\mathrm{S}_{s\times s} & \\ & & -\mathrm{I}_{p\times p} \end{bmatrix}, \quad D_{21} = \begin{bmatrix} \mathrm{O}_{(c_1-s-q)\times k} & & \\ & \mathrm{S}_{s\times s} & \\ & & \mathrm{I}_{q\times q} \end{bmatrix},$$

where $p = r_1 - k - s$ and $q = c_1 - k - s$. □

§3.7 Unitary Decomposition

Proof (cont.)

Then D is unitary; thus the fact that any column (or row) of D has unit length implies that no singular value of D_{11} can exceed 1. Therefore, D_{11} takes the form

$$D_{11} = \begin{bmatrix} \mathrm{I}_{k\times k} & & \\ & \mathrm{C}_{s\times s} & \\ & & \mathrm{O}_{p\times q} \end{bmatrix}$$

for some C taking the desired form, and the orthogonality of columns of D and the orthogonality of rows of D further show that D_{21} and D_{12} must take the form

$$D_{12} = \begin{bmatrix} \mathrm{O}_{k\times(c_2-s-p)} & & \\ & -\mathrm{S}_{s\times s} & \\ & & -\mathrm{I}_{p\times p} \end{bmatrix}, \quad D_{21} = \begin{bmatrix} \mathrm{O}_{(c_1-s-q)\times k} & & \\ & \mathrm{S}_{s\times s} & \\ & & \mathrm{I}_{q\times q} \end{bmatrix},$$

where $p = r_1 - k - s$ and $q = c_1 - k - s$. □

§3.7 Unitary Decomposition

Proof (cont.)

The fact that each column and each row of D has unit length also gives the form of D_{22} so that

$$D = \begin{bmatrix} \mathrm{I} & & & \mathrm{O}_s^\dagger & & \\ & \mathrm{C} & & & -\mathrm{S} & \\ & & \mathrm{O}_c & & & -\mathrm{I} \\ \mathrm{O}_s & & & \mathrm{K} & \mathrm{L} & \\ & \mathrm{S} & & \mathrm{M} & \mathrm{N} & \\ & & \mathrm{I} & & & \mathrm{O}_c^\dagger \end{bmatrix}$$

for some $(r_2 - s - q) \times (c_2 - s - p)$ matrix K, $(r_2 - s - q) \times s$ matrix L, $s \times (c_2 - s - p)$ matrix M and $s \times s$ matrix N. The orthogonality of the second and the fourth blocks of columns shows that $\mathrm{SM} = \mathrm{O}_{s\times(c_2-s-p)}$; thus $\mathrm{M} = \mathrm{O}_{s\times(c_2-s-p)}$ since S is non-singular. Similarly, the orthogonality of the second and the fourth blocks of rows shows that $\mathrm{L} = \mathrm{O}_{(r_2-s-q)\times s}$. □

§3.7 Unitary Decomposition

Proof (cont.)

The fact that each column and each row of D has unit length also gives the form of D_{22} so that

$$D = \begin{bmatrix} \mathrm{I} & & & \mathrm{O}_s^\dagger & & \\ & \mathrm{C} & & & -\mathrm{S} & \\ & & \mathrm{O}_c & & & -\mathrm{I} \\ \mathrm{O}_s & & & \mathrm{K} & \mathrm{L} & \\ & \mathrm{S} & & \mathrm{M} & \mathrm{N} & \\ & & \mathrm{I} & & & \mathrm{O}_c^\dagger \end{bmatrix}$$

for some $(r_2 - s - q) \times (c_2 - s - p)$ matrix K, $(r_2 - s - q) \times s$ matrix L, $s \times (c_2 - s - p)$ matrix M and $s \times s$ matrix N. The orthogonality of the second and the fourth blocks of columns shows that $\mathrm{SM} = \mathrm{O}_{s\times(c_2-s-p)}$; thus $\mathrm{M} = \mathrm{O}_{s\times(c_2-s-p)}$ since S is non-singular. Similarly, the orthogonality of the second and the fourth blocks of rows shows that $\mathrm{L} = \mathrm{O}_{(r_2-s-q)\times s}$. □

§3.7 Unitary Decomposition

Proof (cont.)

The fact that each column and each row of D has unit length also gives the form of D_{22} so that

$$D = \begin{bmatrix} \mathrm{I} & & & \mathrm{O}_s^\dagger & & \\ & \mathrm{C} & & & -\mathrm{S} & \\ & & \mathrm{O}_c & & & -\mathrm{I} \\ \mathrm{O}_s & & & \mathrm{K} & \mathrm{L} & \\ & \mathrm{S} & & \mathrm{M} & \mathrm{N} & \\ & & \mathrm{I} & & & \mathrm{O}_c^\dagger \end{bmatrix}$$

for some $(r_2 - s - q) \times (c_2 - s - p)$ matrix K, $(r_2 - s - q) \times s$ matrix L, $s \times (c_2 - s - p)$ matrix M and $s \times s$ matrix N. The orthogonality of the second and the fourth blocks of columns shows that $\mathrm{SM} = \mathrm{O}_{s\times(c_2-s-p)}$; thus $\mathrm{M} = \mathrm{O}_{s\times(c_2-s-p)}$ since S is non-singular. Similarly, the orthogonality of the second and the fourth blocks of rows shows that $\mathrm{L} = \mathrm{O}_{(r_2-s-q)\times s}$. □

§3.7 Unitary Decomposition

Proof (cont.)

Next, from the fifth and the second blocks of rows, $\mathrm{SC} - \mathrm{NS} = \mathrm{O}_{s\times s}$, so $\mathrm{N} = \mathrm{C}$ and we obtain that

$$D = \begin{bmatrix} \mathrm{I} & & & \mathrm{O}_s^\dagger & & \\ & \mathrm{C} & & & -\mathrm{S} & \\ & & \mathrm{O}_c & & & -\mathrm{I} \\ \mathrm{O}_s & & & \mathrm{K} & & \\ & \mathrm{S} & & & \mathrm{C} & \\ & & \mathrm{I} & & & \mathrm{O}_c^\dagger \end{bmatrix}$$

Finally, note that $r_2 - s - q = r_2 + k - c_1 = c_2 + k - r_1 = c_2 - s - p$ so that K is a square matrix. Together with the fact that $D^\dagger D = DD^\dagger = \mathrm{I}$, we find that $\mathrm{K}\mathrm{K}^\dagger = \mathrm{K}^\dagger\mathrm{K} = \mathrm{I}$ so that K is unitary and can be transformed to I without altering the rest of D by replacing U_2 with $U_2\,\mathrm{blkdiag}(K^\dagger, \mathrm{I}_{s\times s}, \mathrm{I}_{q\times q})$ in (7). □

§3.7 Unitary Decomposition

Proof (cont.)

Next, from the fifth and the second blocks of rows, $\mathrm{SC}-\mathrm{NS} = \mathrm{O}_{s\times s}$, so $\mathrm{N} = \mathrm{C}$ and we obtain that

$$D = \begin{bmatrix} \mathrm{I} & & & \mathrm{O}_s^\dagger & & \\ & \mathrm{C} & & & -\mathrm{S} & \\ & & \mathrm{O}_c & & & -\mathrm{I} \\ \mathrm{O}_s & & & \mathrm{K} & & \\ & \mathrm{S} & & & \mathrm{C} & \\ & & \mathrm{I} & & & \mathrm{O}_c^\dagger \end{bmatrix}$$

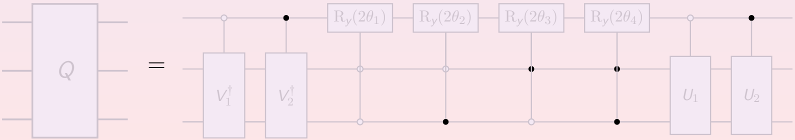
Finally, note that $r_2 - s - q = r_2 + k - c_1 = c_2 + k - r_1 = c_2 - s - p$ so that K is a square matrix. Together with the fact that $D^\dagger D = DD^\dagger = \mathrm{I}$, we find that $\mathrm{K}\mathrm{K}^\dagger = \mathrm{K}^\dagger\mathrm{K} = \mathrm{I}$ so that K is unitary and can be transformed to I without altering the rest of D by replacing U_2 with $U_2\,\mathrm{blkdiag}(K^\dagger, \mathrm{I}_{s\times s}, \mathrm{I}_{q\times q})$ in (7). □

§3.7 Unitary Decomposition

Remark: In quantum computing, for a $2^n \times 2^n$ unitary matrix Q (that is, Q is an n-qubit gate) we apply the CS decomposition for the case $c_1 = c_2 = r_1 = r_2 = 2^{n-1}$ and we have

$$\begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} = \begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix} \begin{bmatrix} \cos\theta_1 & & & -\sin\theta_1 & & \\ & \ddots & & & \ddots & \\ & & \cos\theta_{2^{n-1}} & & & -\sin\theta_{2^{n-1}} \\ \sin\theta_1 & & & \cos\theta_1 & & \\ & \ddots & & & \ddots & \\ & & \sin\theta_{2^{n-1}} & & & \cos\theta_{2^{n-1}} \end{bmatrix} \begin{bmatrix} V_1^\dagger & 0 \\ 0 & V_2^\dagger \end{bmatrix},$$

where $0 \leqslant \theta_1 \leqslant \theta_2 \leqslant \cdots \leqslant \theta_{2^{n-1}} \leqslant \dfrac{\pi}{2}$. In terms of quantum circuits, the case $n = 3$ can be illustrated as follows:

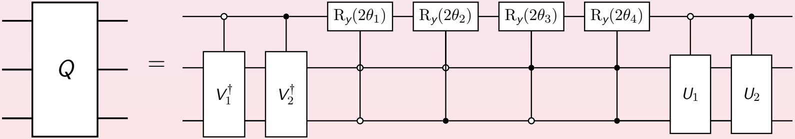

Q
$=$
$V_1^\dagger$
$V_2^\dagger$
$\mathrm{R}_y(2\theta_1)$
$\mathrm{R}_y(2\theta_2)$
$\mathrm{R}_y(2\theta_3)$
$\mathrm{R}_y(2\theta_4)$
U_1
U_2

§3.7 Unitary Decomposition

Remark: In quantum computing, for a $2^n \times 2^n$ unitary matrix Q (that is, Q is an n-qubit gate) we apply the CS decomposition for the case $c_1 = c_2 = r_1 = r_2 = 2^{n-1}$ and we have

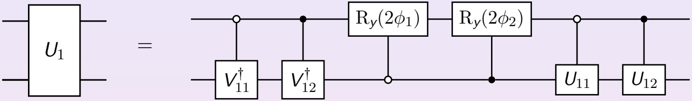
$$\begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} = \begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix} \begin{bmatrix} \cos\theta_1 & & & -\sin\theta_1 & & \\ & \ddots & & & \ddots & \\ & & \cos\theta_{2^{n-1}} & & & -\sin\theta_{2^{n-1}} \\ \sin\theta_1 & & & \cos\theta_1 & & \\ & \ddots & & & \ddots & \\ & & \sin\theta_{2^{n-1}} & & & \cos\theta_{2^{n-1}} \end{bmatrix} \begin{bmatrix} V_1^\dagger & 0 \\ 0 & V_2^\dagger \end{bmatrix},$$

where $0 \leqslant \theta_1 \leqslant \theta_2 \leqslant \cdots \leqslant \theta_{2^{n-1}} \leqslant \frac{\pi}{2}$. In terms of quantum circuits, the case $n = 3$ can be illustrated as follows:


Q
$=$
$V_1^\dagger$
$V_2^\dagger$
$\mathrm{R}_y(2\theta_1)$
$\mathrm{R}_y(2\theta_2)$
$\mathrm{R}_y(2\theta_3)$
$\mathrm{R}_y(2\theta_4)$
U_1
U_2

§3.7 Unitary Decomposition

The 2-qubit gates U_1, U_2, $V_1^\dagger$ and $V_2^\dagger$ can be further decomposed. For example,



for some $0 \leqslant \phi_1 \leqslant \phi_2 \leqslant \pi/2$ and quantum gates with matrix representations $V_{11}, V_{12}, U_{11}, U_{12}$ so that

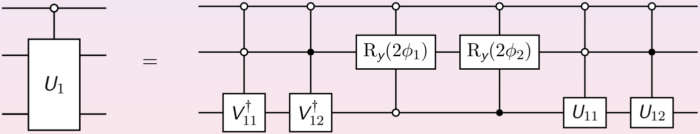
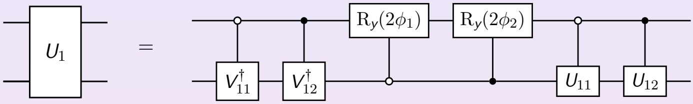


Figure 2: The decomposition of the controlled-U gate

Therefore, the CS decomposition essentially provides a way to express an n-qubit gate as the product of multi-controlled gates.

§3.7 Unitary Decomposition

The 2-qubit gates U_1, U_2, $V_1^\dagger$ and $V_2^\dagger$ can be further decomposed. For example,



for some $0 \leqslant \phi_1 \leqslant \phi_2 \leqslant \pi/2$ and quantum gates with matrix representations $V_{11}, V_{12}, U_{11}, U_{12}$ so that

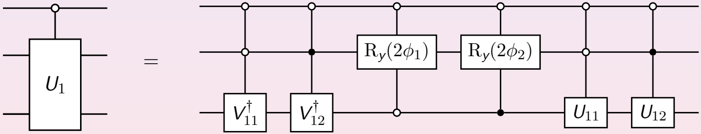


Figure 2: The decomposition of the controlled-U gate

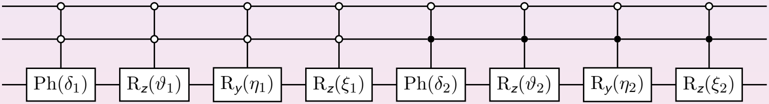
Therefore, the CS decomposition essentially provides a way to express an n-qubit gate as the product of multi-controlled gates.

§3.7 Unitary Decomposition

Recall that any 1-qubit gate can be decomposed further as the product of rotation gates R_y, R_z and the phase gate Ph. If the quantum gates U_{11} and U_{12} in Figure 2 can be expressed as

$$U_{11} = \mathrm{R}_z(\xi_1)\mathrm{R}_y(\eta_1)\mathrm{R}_z(\vartheta_1)\mathrm{Ph}(\delta_1), \quad U_{12} = \mathrm{R}_z(\xi_2)\mathrm{R}_y(\eta_2)\mathrm{R}_z(\vartheta_2)\mathrm{Ph}(\delta_2),$$

the controlled-U gates in Figure 2 can be further decomposed into



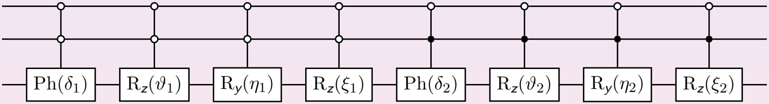
Without any further modification, we can express an n-qubit gate as the product of multi-controlled ***rotation*** gates, **at an expense of some not implementable phase gates**. In the next section, we talk about how to “cancel out” these phase gates and make an n-qubit gate indeed the product of multi-controlled rotation gates.

§3.7 Unitary Decomposition

Recall that any 1-qubit gate can be decomposed further as the product of rotation gates R_y, R_z and the phase gate Ph. If the quantum gates U_{11} and U_{12} in Figure 2 can be expressed as

$$U_{11} = \mathrm{R}_z(\xi_1)\mathrm{R}_y(\eta_1)\mathrm{R}_z(\vartheta_1)\mathrm{Ph}(\delta_1), \quad U_{12} = \mathrm{R}_z(\xi_2)\mathrm{R}_y(\eta_2)\mathrm{R}_z(\vartheta_2)\mathrm{Ph}(\delta_2),$$

the controlled-U gates in Figure 2 can be further decomposed into



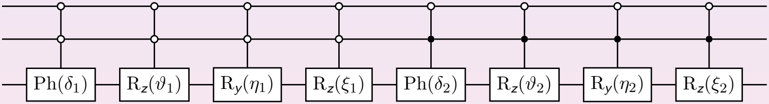
Without any further modification, we can express an n-qubit gate as the product of multi-controlled ***rotation*** gates, **at an expense of some not implementable phase gates**. In the next section, we talk about how to “cancel out” these phase gates and make an n-qubit gate indeed the product of multi-controlled rotation gates.

§3.7 Unitary Decomposition

Recall that any 1-qubit gate can be decomposed further as the product of rotation gates R_y, R_z and the phase gate Ph. If the quantum gates U_{11} and U_{12} in Figure 2 can be expressed as

$$U_{11} = \mathrm{R}_z(\xi_1)\mathrm{R}_y(\eta_1)\mathrm{R}_z(\vartheta_1)\mathrm{Ph}(\delta_1), \quad U_{12} = \mathrm{R}_z(\xi_2)\mathrm{R}_y(\eta_2)\mathrm{R}_z(\vartheta_2)\mathrm{Ph}(\delta_2),$$

the controlled-U gates in Figure 2 can be further decomposed into



Without any further modification, we can express an n-qubit gate as the product of multi-controlled ***rotation*** gates, **at an expense of some not implementable phase gates**. In the next section, we talk about how to “cancel out” these phase gates and make an n-qubit gate indeed the product of multi-controlled rotation gates.

§3.7 Unitary Decomposition

§3.7.4 Decomposition of arbitrary quantum gates

We note that the matrix $\begin{bmatrix} \mathrm{C} & -\mathrm{S} \\ \mathrm{S} & \mathrm{C} \end{bmatrix}$ commutes with the matrix $\begin{bmatrix} P & 0 \\ 0 & P \end{bmatrix}$ for all $2^{n-1} \times 2^{n-1}$ **diagonal unitary** matrix P; that is,

$$\begin{bmatrix} P & 0 \\ 0 & P \end{bmatrix}\begin{bmatrix} \mathrm{C} & -\mathrm{S} \\ \mathrm{S} & \mathrm{C} \end{bmatrix} = \begin{bmatrix} \mathrm{C} & -\mathrm{S} \\ \mathrm{S} & \mathrm{C} \end{bmatrix}\begin{bmatrix} P & 0 \\ 0 & P \end{bmatrix} \qquad \forall\, P = \mathrm{diag}(e^{i\phi_1}, \cdots, e^{i\phi_{2^{n-1}}}).$$

Therefore, if P is a diagonal unitary matrix,

$$\begin{aligned} \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} &= \begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix}\begin{bmatrix} \mathrm{C} & -\mathrm{S} \\ \mathrm{S} & \mathrm{C} \end{bmatrix}\begin{bmatrix} V_1^\dagger & 0 \\ 0 & V_2^\dagger \end{bmatrix} \\ &= \begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix}\begin{bmatrix} P & 0 \\ 0 & P \end{bmatrix}\begin{bmatrix} P^\dagger & 0 \\ 0 & P^\dagger \end{bmatrix}\begin{bmatrix} \mathrm{C} & -\mathrm{S} \\ \mathrm{S} & \mathrm{C} \end{bmatrix}\begin{bmatrix} V_1^\dagger & 0 \\ 0 & V_2^\dagger \end{bmatrix} \\ &= \begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix}\begin{bmatrix} P & 0 \\ 0 & P \end{bmatrix}\begin{bmatrix} \mathrm{C} & -\mathrm{S} \\ \mathrm{S} & \mathrm{C} \end{bmatrix}\begin{bmatrix} (V_1P)^\dagger & 0 \\ 0 & (V_2P)^\dagger \end{bmatrix}. \end{aligned}$$

The diagonal unitary matrix P will be chosen to "cancel out the phase gate" so that the matrix $\begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix}\begin{bmatrix} P & 0 \\ 0 & P \end{bmatrix}$ is a product of multi-controlled **rotation** gates.

§3.7 Unitary Decomposition

§3.7.4 Decomposition of arbitrary quantum gates

We note that the matrix $\begin{bmatrix} \mathrm{C} & -\mathrm{S} \\ \mathrm{S} & \mathrm{C} \end{bmatrix}$ commutes with the matrix $\begin{bmatrix} P & 0 \\ 0 & P \end{bmatrix}$ for all $2^{n-1} \times 2^{n-1}$ **diagonal unitary** matrix P; that is,

$$\begin{bmatrix} P & 0 \\ 0 & P \end{bmatrix}\begin{bmatrix} \mathrm{C} & -\mathrm{S} \\ \mathrm{S} & \mathrm{C} \end{bmatrix} = \begin{bmatrix} \mathrm{C} & -\mathrm{S} \\ \mathrm{S} & \mathrm{C} \end{bmatrix}\begin{bmatrix} P & 0 \\ 0 & P \end{bmatrix} \qquad \forall\, P = \mathrm{diag}(e^{i\phi_1}, \cdots, e^{i\phi_{2^{n-1}}}).$$

Therefore, if P is a diagonal unitary matrix,

$$\begin{aligned} \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} &= \begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix}\begin{bmatrix} \mathrm{C} & -\mathrm{S} \\ \mathrm{S} & \mathrm{C} \end{bmatrix}\begin{bmatrix} V_1^\dagger & 0 \\ 0 & V_2^\dagger \end{bmatrix} \\ &= \begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix}\begin{bmatrix} P & 0 \\ 0 & P \end{bmatrix}\begin{bmatrix} \mathrm{C} & -\mathrm{S} \\ \mathrm{S} & \mathrm{C} \end{bmatrix}\begin{bmatrix} P^\dagger & 0 \\ 0 & P^\dagger \end{bmatrix}\begin{bmatrix} V_1^\dagger & 0 \\ 0 & V_2^\dagger \end{bmatrix} \\ &= \begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix}\begin{bmatrix} P & 0 \\ 0 & P \end{bmatrix}\begin{bmatrix} \mathrm{C} & -\mathrm{S} \\ \mathrm{S} & \mathrm{C} \end{bmatrix}\begin{bmatrix} (V_1 P)^\dagger & 0 \\ 0 & (V_2 P)^\dagger \end{bmatrix}. \end{aligned}$$

The diagonal unitary matrix P will be chosen to "cancel out the phase gate" so that the matrix $\begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix}\begin{bmatrix} P & 0 \\ 0 & P \end{bmatrix}$ is a product of multi-controlled **rotation** gates.

§3.7 Unitary Decomposition

§3.7.4 Decomposition of arbitrary quantum gates

We note that the matrix $\begin{bmatrix} \mathrm{C} & -\mathrm{S} \\ \mathrm{S} & \mathrm{C} \end{bmatrix}$ commutes with the matrix $\begin{bmatrix} P & 0 \\ 0 & P \end{bmatrix}$ for all $2^{n-1} \times 2^{n-1}$ **diagonal unitary** matrix P; that is,

$$\begin{bmatrix} P & 0 \\ 0 & P \end{bmatrix}\begin{bmatrix} \mathrm{C} & -\mathrm{S} \\ \mathrm{S} & \mathrm{C} \end{bmatrix} = \begin{bmatrix} \mathrm{C} & -\mathrm{S} \\ \mathrm{S} & \mathrm{C} \end{bmatrix}\begin{bmatrix} P & 0 \\ 0 & P \end{bmatrix} \qquad \forall\, P = \mathrm{diag}(e^{i\phi_1}, \cdots, e^{i\phi_{2^{n-1}}}).$$

Therefore, if P is a diagonal unitary matrix,

$$\begin{aligned}
\begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} &= \begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix}\begin{bmatrix} \mathrm{C} & -\mathrm{S} \\ \mathrm{S} & \mathrm{C} \end{bmatrix}\begin{bmatrix} V_1^\dagger & 0 \\ 0 & V_2^\dagger \end{bmatrix} \\
&= \begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix}\begin{bmatrix} P & 0 \\ 0 & P \end{bmatrix}\begin{bmatrix} \mathrm{C} & -\mathrm{S} \\ \mathrm{S} & \mathrm{C} \end{bmatrix}\begin{bmatrix} P^\dagger & 0 \\ 0 & P^\dagger \end{bmatrix}\begin{bmatrix} V_1^\dagger & 0 \\ 0 & V_2^\dagger \end{bmatrix} \\
&= \begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix}\begin{bmatrix} P & 0 \\ 0 & P \end{bmatrix}\begin{bmatrix} \mathrm{C} & -\mathrm{S} \\ \mathrm{S} & \mathrm{C} \end{bmatrix}\begin{bmatrix} (V_1 P)^\dagger & 0 \\ 0 & (V_2 P)^\dagger \end{bmatrix}.
\end{aligned}$$

The diagonal unitary matrix P will be chosen to “cancel out the phase gate” so that the matrix $\begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix}\begin{bmatrix} P & 0 \\ 0 & P \end{bmatrix}$ is a product of multi-controlled **rotation** gates.

§3.7 Unitary Decomposition

§3.7.4 Decomposition of arbitrary quantum gates

We note that the matrix $\begin{bmatrix} \mathrm{C} & -\mathrm{S} \\ \mathrm{S} & \mathrm{C} \end{bmatrix}$ commutes with the matrix $\begin{bmatrix} P & 0 \\ 0 & P \end{bmatrix}$ for all $2^{n-1} \times 2^{n-1}$ **diagonal unitary** matrix P; that is,

$$\begin{bmatrix} P & 0 \\ 0 & P \end{bmatrix}\begin{bmatrix} \mathrm{C} & -\mathrm{S} \\ \mathrm{S} & \mathrm{C} \end{bmatrix} = \begin{bmatrix} \mathrm{C} & -\mathrm{S} \\ \mathrm{S} & \mathrm{C} \end{bmatrix}\begin{bmatrix} P & 0 \\ 0 & P \end{bmatrix} \qquad \forall\, P = \mathrm{diag}(e^{i\phi_1}, \cdots, e^{i\phi_{2^{n-1}}}).$$

Therefore, if P is a diagonal unitary matrix,

$$\begin{aligned} \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} &= \begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix}\begin{bmatrix} \mathrm{C} & -\mathrm{S} \\ \mathrm{S} & \mathrm{C} \end{bmatrix}\begin{bmatrix} V_1^\dagger & 0 \\ 0 & V_2^\dagger \end{bmatrix} \\ &= \begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix}\begin{bmatrix} P & 0 \\ 0 & P \end{bmatrix}\begin{bmatrix} \mathrm{C} & -\mathrm{S} \\ \mathrm{S} & \mathrm{C} \end{bmatrix}\begin{bmatrix} P^\dagger & 0 \\ 0 & P^\dagger \end{bmatrix}\begin{bmatrix} V_1^\dagger & 0 \\ 0 & V_2^\dagger \end{bmatrix} \\ &= \begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix}\begin{bmatrix} P & 0 \\ 0 & P \end{bmatrix}\begin{bmatrix} \mathrm{C} & -\mathrm{S} \\ \mathrm{S} & \mathrm{C} \end{bmatrix}\begin{bmatrix} (V_1P)^\dagger & 0 \\ 0 & (V_2P)^\dagger \end{bmatrix}. \end{aligned}$$

The diagonal unitary matrix P will be chosen to “cancel out the phase gate” so that the matrix $\begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix}\begin{bmatrix} P & 0 \\ 0 & P \end{bmatrix}$ is a product of multi-controlled **rotation** gates.

§3.7 Unitary Decomposition

Let U be an n-qubit quantum gate. By previous remark

$$U = \begin{bmatrix} U_{11}^1 & 0 \\ 0 & U_{12}^1 \end{bmatrix} \begin{bmatrix} P_{11}^1 & 0 \\ 0 & P_{11}^1 \end{bmatrix} \begin{bmatrix} C_{11}^1 & -S_{11}^1 \\ S_{11}^1 & C_{11}^1 \end{bmatrix} \begin{bmatrix} U_{21}^1 & 0 \\ 0 & U_{22}^1 \end{bmatrix},$$

where P_{11}^1 is a $2^{n-1} \times 2^{n-1}$ diagonal unitary matrix **to be determined**. The decomposition can be applied recursively to the **sub-matrices** U_{jk}^i until a 2×2 block-diagonal form is encountered. For example, we use the CS decomposition to write

$$U_{11}^1 = U_{111}^2 P_{111}^2 A_{11}^2 U_{112}^2 , \quad U_{12}^1 = U_{121}^2 P_{121}^2 A_{12}^2 U_{122}^2 ,$$

so that

$$\begin{aligned} \begin{bmatrix} U_{11}^1 & 0 \\ 0 & U_{12}^1 \end{bmatrix} &= \begin{bmatrix} U_{111}^2 P_{111}^2 A_{11}^2 U_{112}^2 & 0 \\ 0 & U_{121}^2 P_{121}^2 A_{12}^2 U_{122}^2 \end{bmatrix} \\ &= \begin{bmatrix} U_{111}^2 & 0 \\ 0 & U_{121}^2 \end{bmatrix} \begin{bmatrix} P_{111}^2 & 0 \\ 0 & P_{121}^2 \end{bmatrix} \begin{bmatrix} A_{11}^2 & \\ & A_{12}^2 \end{bmatrix} \begin{bmatrix} U_{112}^2 & 0 \\ 0 & U_{122}^2 \end{bmatrix}. \end{aligned}$$

§3.7 Unitary Decomposition

Let U be an n-qubit quantum gate. By previous remark

$$U = \begin{bmatrix} U^1_{11} & 0 \\ 0 & U^1_{12} \end{bmatrix} \begin{bmatrix} P^1_{11} & 0 \\ 0 & P^1_{11} \end{bmatrix} \begin{bmatrix} C^1_{11} & -S^1_{11} \\ S^1_{11} & C^1_{11} \end{bmatrix} \begin{bmatrix} U^1_{21} & 0 \\ 0 & U^1_{22} \end{bmatrix},$$

where P^1_{11} is a $2^{n-1} \times 2^{n-1}$ diagonal unitary matrix **to be determined**. The decomposition can be applied recursively to the **submatrices** U^i_{jk} until a 2×2 block-diagonal form is encountered. For example, we use the CS decomposition to write

$$U^1_{11} = U^2_{111} P^2_{111} A^2_{11} U^2_{112}\,, \quad U^1_{12} = U^2_{121} P^2_{121} A^2_{12} U^2_{122}\,,$$

so that

$$\begin{aligned} \begin{bmatrix} U^1_{11} & 0 \\ 0 & U^1_{12} \end{bmatrix} &= \begin{bmatrix} U^2_{111} P^2_{111} A^2_{11} U^2_{112} & 0 \\ 0 & U^2_{121} P^2_{121} A^2_{12} U^2_{122} \end{bmatrix} \\ &= \begin{bmatrix} U^2_{111} & 0 \\ 0 & U^2_{121} \end{bmatrix} \begin{bmatrix} P^2_{111} & 0 \\ 0 & P^2_{121} \end{bmatrix} \begin{bmatrix} A^2_{11} & \\ & A^2_{12} \end{bmatrix} \begin{bmatrix} U^2_{112} & 0 \\ 0 & U^2_{122} \end{bmatrix}. \end{aligned}$$

§3.7 Unitary Decomposition

In general, for $1 \leqslant i \leqslant n$ and $1 \leqslant j \leqslant 2^{i-1}$, we use the CS decomposition on each block of U_j^{i-1} to write

$$U_j^{i-1} = U_{2j-1}^i P_{2j-1}^i A_{2j-1}^i U_{2j}^i ,$$

where U_{2j-1}^i and U_{2j}^i are block diagonal matrices consisting of 2^i blocks of $2^{n-i} \times 2^{n-i}$ unitary matrices, P_{2j-1}^i is a block diagonal matrix of the form

$$P_{2j-1}^i = \mathrm{blkdiag}\big(Q_1^i, Q_1^i, Q_2^i, Q_2^i, \cdots, Q_{2^{i-1}}^i, Q_{2^{i-1}}^i\big)$$

for some $2^{n-i} \times 2^{n-i}$ diagonal unitary matrices $Q_1^i, \cdots, Q_{2^{i-1}}^i$ to be determined. We note that in principle we need to know P_{2j-1}^i first before we can decompose U_{2j}^i further since U_{2j}^i depends on P_{2j-1}^i.

§3.7 Unitary Decomposition

In general, for $1 \leqslant i \leqslant n$ and $1 \leqslant j \leqslant 2^{i-1}$, we use the CS decomposition on each block of U^{i-1}_j to write

$$U^{i-1}_j = U^i_{2j-1} P^i_{2j-1} A^i_{2j-1} U^i_{2j} ,$$

where U^i_{2j-1} and U^i_{2j} are block diagonal matrices consisting of 2^i blocks of $2^{n-i} \times 2^{n-i}$ unitary matrices, P^i_{2j-1} is a block diagonal matrix of the form

$$P^i_{2j-1} = \mathrm{blkdiag}\big(Q^i_1, Q^i_1, Q^i_2, Q^i_2, \cdots, Q^i_{2^{i-1}}, Q^i_{2^{i-1}}\big)$$

for some $2^{n-i} \times 2^{n-i}$ diagonal unitary matrices $Q^i_1, \cdots, Q^i_{2^{i-1}}$ to be determined. We note that in principle we need to know P^i_{2j-1} first before we can decompose U^i_{2j} further since U^i_{2j} depends on P^i_{2j-1}.

§3.7 Unitary Decomposition

Define $P^i_{2j} = P^{i-1}_j$ and $A^i_{2j} = A^{i-1}_j$. We then have the following sequence of decomposition

$$
\begin{aligned}
U &= U^1_1 P^1_1 A^1_1 U^1_2 = U^2_1 P^2_1 A^2_1 U^2_2 P^2_2 A^2_2 U^2_3 P^2_3 A^2_3 U^2_4 \\
&= U^2_1 P^2_1 A^2_1 U^2_2 P^2_2 A^2_2 U^2_3 P^2_3 A^2_3 U^2_4 \\
&= U^3_1 P^3_1 A^3_1 U^3_2 P^3_2 A^3_2 U^3_3 P^3_3 A^3_3 U^3_4 P^3_4 A^3_4 U^3_5 P^3_5 A^3_5 U^3_6 P^3_6 A^3_6 U^3_7 P^3_7 A^3_7 U^3_8 \\
&= \cdots\cdots\cdots \\
&= U^{n-1}_1 P^{n-1}_1 A^{n-1}_1 U^{n-1}_2 P^{n-1}_2 A^{n-1}_2 \cdots U^{n-1}_{2^{n-1}-1} P^{n-1}_{2^{n-1}-1} A^{n-1}_{2^{n-1}-1} U^{n-1}_{2^{n-1}} \\
&= \left(\prod_{j=1}^{2^{n-1}-1} U^{n-1}_j P^{n-1}_j A^{n-1}_j \right) U^{n-1}_{2^{n-1}} .
\end{aligned}
$$

Here the upper index denotes the level of recursion, whereas the lower index denotes the position of the matrix within the resulting matrix product.

§3.7 Unitary Decomposition

In the decomposition sequence above,

1. U_j^{n-1} takes the form
$$U_j^{n-1} = \mathrm{blkdiag}(U_1, U_2, \cdots, U_{2^{n-1}})$$
for some 2×2 unitary matrices $U_1, \cdots, U_{2^{n-1}}$.

2. For each $j \in \mathbb{N}$, let $\gamma(j)$ indicate the unique integer satisfying
$$j = 2^{\gamma(j)-1}(2k-1) \quad \text{for some } k \in \mathbb{N}\,.$$
Then
$$P_j^{n-1} = P_{2^{\gamma(j)-1}(2k-1)}^{n-1} = P_{2k-1}^{n-\gamma(j)}\,,$$
$$A_j^{n-1} = A_{2^{\gamma(j)-1}(2k-1)}^{n-1} = A_{2k-1}^{n-\gamma(j)}$$
which imply that P_j^{n-1} and A_j^{n-1} appear first time in the $(n-\gamma(j))$-th recursion of decompositions and do not appear in any previous recursion of decompositions.

§3.7 Unitary Decomposition

In the decomposition sequence above,

1. U_j^{n-1} takes the form

$$U_j^{n-1} = \mathrm{blkdiag}(U_1, U_2, \cdots, U_{2^{n-1}})$$

for some 2×2 unitary matrices $U_1, \cdots, U_{2^{n-1}}$.

2. For each $j \in \mathbb{N}$, let $\gamma(j)$ indicate the unique integer satisfying

$$j = 2^{\gamma(j)-1}(2k-1) \quad \text{for some } k \in \mathbb{N}\,.$$

Then

$$\begin{aligned} P_j^{n-1} &= P_{2^{\gamma(j)-1}(2k-1)}^{n-1} = P_{2k-1}^{n-\gamma(j)}\,, \\ A_j^{n-1} &= A_{2^{\gamma(j)-1}(2k-1)}^{n-1} = A_{2k-1}^{n-\gamma(j)} \end{aligned}$$

which imply that P_j^{n-1} and A_j^{n-1} appear first time in the $(n-\gamma(j))$-th recursion of decompositions and do not appear in any previous recursion of decompositions.

§3.7 Unitary Decomposition

Therefore, P_j^{n-1} takes the form

$$P_j^{n-1} = \operatorname{blkdiag}\big(Q_1, Q_1, \cdots, Q_{2^{n-\gamma(j)-1}}, Q_{2^{n-\gamma(j)-1}}\big)$$

for some $2^{\gamma(j)}\times 2^{\gamma(j)}$ diagonal unitary matrices $Q_1, \cdots, Q_{2^{n-\gamma(j)-1}}$, and A_j^{n-1} takes the form

$$A_j^{n-1} = \operatorname{blkdiag}\left(\begin{bmatrix} C_1 & -S_1 \\ S_1 & C_1 \end{bmatrix}, \cdots, \begin{bmatrix} C_{2^{n-\gamma(j)-1}} & -S_{2^{n-\gamma(j)-1}} \\ S_{2^{n-\gamma(j)-1}} & C_{2^{n-\gamma(j)-1}} \end{bmatrix}\right),$$

where for $1 \leqslant k \leqslant 2^{n-\gamma(j)-1}$,

$$C_k = \operatorname{diag}(\cos\theta_1^k, \cdots, \cos\theta_{2^{\gamma(j)}}^k),$$
$$S_k = \operatorname{diag}(\sin\theta_1^k, \cdots, \sin\theta_{2^{\gamma(j)}}^k)$$

for some $0 \leqslant \theta_1^k \leqslant \theta_2^k \leqslant \cdots \leqslant \theta_{2^{\gamma(j)}}^k \leqslant \dfrac{\pi}{2}$. We note that A_j^{n-1} is indeed a multi-controlled gate of type $F_{n-\gamma(j)}^n(\mathrm{R}_y)$.

§3.7 Unitary Decomposition

Therefore, P_j^{n-1} takes the form

$$P_j^{n-1} = \mathrm{blkdiag}\big(Q_1, Q_1, \cdots, Q_{2^{n-\gamma(j)-1}}, Q_{2^{n-\gamma(j)-1}}\big)$$

for some $2^{\gamma(j)} \times 2^{\gamma(j)}$ diagonal unitary matrices $Q_1, \cdots, Q_{2^{n-\gamma(j)-1}}$, and A_j^{n-1} takes the form

$$A_j^{n-1} = \mathrm{blkdiag}\left(\begin{bmatrix} C_1 & -S_1 \\ S_1 & C_1 \end{bmatrix}, \cdots, \begin{bmatrix} C_{2^{n-\gamma(j)-1}} & -S_{2^{n-\gamma(j)-1}} \\ S_{2^{n-\gamma(j)-1}} & C_{2^{n-\gamma(j)-1}} \end{bmatrix}\right),$$

where for $1 \leqslant k \leqslant 2^{n-\gamma(j)-1}$,

$$C_k = \mathrm{diag}\big(\cos\theta_1^k, \cdots, \cos\theta_{2^{\gamma(j)}}^k\big),$$
$$S_k = \mathrm{diag}\big(\sin\theta_1^k, \cdots, \sin\theta_{2^{\gamma(j)}}^k\big)$$

for some $0 \leqslant \theta_1^k \leqslant \theta_2^k \leqslant \cdots \leqslant \theta_{2^{\gamma(j)}}^k \leqslant \frac{\pi}{2}$. We note that A_j^{n-1} is indeed a multi-controlled gate of type $F_{n-\gamma(j)}^n(\mathrm{R}_y)$.

§3.7 Unitary Decomposition

3. P_j^{n-1} can be chosen according to U_j^{n-1} so that $U_j^{n-1}P_j^{n-1}$ is a product of multi-controlled rotation gates (which will be explained soon). On the other hand, for each $1 \leqslant j \leqslant 2^n - 1$ the block diagonal matrix U_{j+1}^{n-1} depends on P_k^{n-1} for all $1 \leqslant k \leqslant j$; thus we need to specify P_1^{n-1}, P_2^{n-1}, $\cdots$ successively in order to complete the decomposition.

Remark: In matlab®, γ can be implemented by

$$\gamma(j) = \mathbf{min}(\mathbf{find}(\mathbf{de2bi}(j) == 1)).$$

§3.7 Unitary Decomposition

3. P_j^{n-1} can be chosen according to U_j^{n-1} so that $U_j^{n-1}P_j^{n-1}$ is a product of multi-controlled rotation gates (which will be explained soon). On the other hand, for each $1 \leqslant j \leqslant 2^n-1$ the block diagonal matrix U_{j+1}^{n-1} depends on P_k^{n-1} for all $1 \leqslant k \leqslant j$; thus we need to specify P_1^{n-1}, P_2^{n-1}, $\cdots$ successively in order to complete the decomposition.

Remark: In matlab®, γ can be implemented by

$$\gamma(j) = \mathbf{min}(\mathbf{find}(\mathbf{de2bi}(j) == 1)).$$

§3.7 Unitary Decomposition

Now we determine P_1^{n-1}. Since U_1^{n-1} is a block diagonal matrix consisting of 2^{n-1} blocks of 2×2 unitary matrices $U_{11}^{n-1}, \cdots, U_{12^{n-1}}^{n-1}$; that is,

$$U_1^{n-1} = \begin{bmatrix} U_{11}^{n-1} & & & \\ & U_{12}^{n-1} & & \\ & & \ddots & \\ & & & U_{12^{n-1}}^{n-1} \end{bmatrix},$$

for each $1 \leqslant j \leqslant 2^{n-1}$ there exist $\delta_j, \xi_j, \theta_j, \eta_j$ such that

$$U_{1j}^{n-1} = \mathrm{R}_z(\xi_j)\mathrm{R}_y(\theta_j)\mathrm{R}_z(\eta_j)\mathrm{Ph}(\delta_j)\,;$$

thus U_1^{n-1} can be written as the following product

$$\begin{bmatrix} \mathrm{R}_z(\xi_1) & & \\ & \ddots & \\ & & \mathrm{R}_z(\xi_{2^{n-1}}) \end{bmatrix} \begin{bmatrix} \mathrm{R}_y(\theta_1) & & \\ & \ddots & \\ & & \mathrm{R}_y(\theta_{2^{n-1}}) \end{bmatrix} \begin{bmatrix} \mathrm{R}_z(\eta_1) & & \\ & \ddots & \\ & & \mathrm{R}_z(\eta_{2^{n-1}}) \end{bmatrix} \begin{bmatrix} \mathrm{Ph}(\delta_1) & & \\ & \ddots & \\ & & \mathrm{Ph}(\delta_{2^{n-1}}) \end{bmatrix}.$$

§3.7 Unitary Decomposition

Now we determine P_1^{n-1}. Since U_1^{n-1} is a block diagonal matrix consisting of 2^{n-1} blocks of 2×2 unitary matrices $U_{11}^{n-1}, \cdots, U_{12^{n-1}}^{n-1}$; that is,

$$U_1^{n-1} = \begin{bmatrix} U_{11}^{n-1} & & & \\ & U_{12}^{n-1} & & \\ & & \ddots & \\ & & & U_{12^{n-1}}^{n-1} \end{bmatrix},$$

for each $1 \leqslant j \leqslant 2^{n-1}$ there exist $\delta_j, \xi_j, \theta_j, \eta_j$ such that

$$U_{1j}^{n-1} = \mathrm{R}_z(\xi_j)\mathrm{R}_y(\theta_j)\mathrm{R}_z(\eta_j)\mathrm{Ph}(\delta_j)\,;$$

thus U_1^{n-1} can be written as the following product

$$\begin{bmatrix} \mathrm{R}_z(\xi_1) & & \\ & \ddots & \\ & & \mathrm{R}_z(\xi_{2^{n-1}}) \end{bmatrix} \begin{bmatrix} \mathrm{R}_y(\theta_1) & & \\ & \ddots & \\ & & \mathrm{R}_y(\theta_{2^{n-1}}) \end{bmatrix} \begin{bmatrix} \mathrm{R}_z(\eta_1) & & \\ & \ddots & \\ & & \mathrm{R}_z(\eta_{2^{n-1}}) \end{bmatrix} \begin{bmatrix} \mathrm{Ph}(\delta_1) & & \\ & \ddots & \\ & & \mathrm{Ph}(\delta_{2^{n-1}}) \end{bmatrix}.$$

§3.7 Unitary Decomposition

We note that the last matrix is **not** implementable since it consists of phase gates. The P matrix will be combined with this “phase gates” so that the combination is multi-controlled rotation gate.

For each $1 \leqslant j \leqslant 2^{n-2}$, let $\alpha_j = -\dfrac{\delta_{2j-1} + \delta_{2j}}{2}$. Define $Q_j = \mathrm{diag}(e^{i\alpha_j}, e^{i\alpha_j})$ and $\beta_j = \delta_{2j} - \delta_{2j-1}$. Then

$$\mathrm{Ph}(\delta_{2j-1}) Q_j = \mathrm{diag}\big(e^{-i\beta_j/2}, e^{-i\beta_j/2}\big)\,,$$
$$\mathrm{Ph}(\delta_{2j}) Q_j = \mathrm{diag}\big(e^{i\beta_j/2}, e^{i\beta_j/2}\big)$$

so that

$$\begin{bmatrix}\mathrm{Ph}(\delta_{2j-1}) & \\ & \mathrm{Ph}(\delta_{2j})\end{bmatrix}\begin{bmatrix}Q_j & \\ & Q_j\end{bmatrix} = \begin{bmatrix} e^{-i\beta_j/2} & & & \\ & e^{-i\beta_j/2} & & \\ & & e^{i\beta_j/2} & \\ & & & e^{i\beta_j/2}\end{bmatrix}.$$

§3.7 Unitary Decomposition

We note that the last matrix is **not** implementable since it consists of phase gates. The P matrix will be combined with this "phase gates" so that the combination is multi-controlled rotation gate.

For each $1 \leqslant j \leqslant 2^{n-2}$, let $\alpha_j = -\frac{\delta_{2j-1} + \delta_{2j}}{2}$. Define $Q_j = \mathrm{diag}(e^{i\alpha_j}, e^{i\alpha_j})$ and $\beta_j = \delta_{2j} - \delta_{2j-1}$. Then

$$\mathrm{Ph}(\delta_{2j-1})Q_j = \mathrm{diag}\left(e^{-i\beta_j/2}, e^{-i\beta_j/2}\right),$$
$$\mathrm{Ph}(\delta_{2j})Q_j = \mathrm{diag}\left(e^{i\beta_j/2}, e^{i\beta_j/2}\right)$$

so that

$$\begin{bmatrix} \mathrm{Ph}(\delta_{2j-1}) & \\ & \mathrm{Ph}(\delta_{2j}) \end{bmatrix} \begin{bmatrix} Q_j & \\ & Q_j \end{bmatrix} = \begin{bmatrix} e^{-i\beta_j/2} & & & \\ & e^{-i\beta_j/2} & & \\ & & e^{i\beta_j/2} & \\ & & & e^{i\beta_j/2} \end{bmatrix}.$$

§3.7 Unitary Decomposition

Therefore, by defining

$$P_1^{n-1} = \mathrm{blkdiag}\big(Q_1, Q_1, Q_2, Q_2, \cdots, Q_{2^{n-1}}, Q_{2^{n-1}}\big)$$

we have

$$\begin{aligned}&\mathrm{blkdiag}\big(\mathrm{Ph}(\delta_1), \cdots, \mathrm{Ph}(\delta_{2^{n-1}})\big) P_1^{n-1} \\ &\quad = \mathrm{diag}\big(e^{-i\beta_1/2}, e^{-i\beta_1/2}, e^{i\beta_1/2}, e^{i\beta_1/2}, \\ &\qquad\qquad e^{-i\beta_2/2}, e^{-i\beta_2/2}, e^{i\beta_2/2}, e^{i\beta_2/2}, \cdots \\ &\qquad\qquad \cdots, e^{-i\beta_{2^{n-2}}/2}, e^{-i\beta_{2^{n-2}}/2}, e^{i\beta_{2^{n-2}}/2}, e^{i\beta_{2^{n-2}}/2}\big)\end{aligned}$$

which is a multi-controlled gate of type $F_n^n(\mathrm{R}_z)$ (with $\theta_{2j-1} = \theta_{2j}$ for all $1 \leqslant j \leqslant 2^{n-1}$). This shows that $U_1^{n-1} P_1^{n-1}$ is a product of multi-controlled rotation gates in which the rotation gates involved are R_y and R_z.

§3.7 Unitary Decomposition

Therefore, by defining

$$P_1^{n-1} = \mathrm{blkdiag}\big(Q_1, Q_1, Q_2, Q_2, \cdots, Q_{2^{n-1}}, Q_{2^{n-1}}\big)$$

we have

$$\begin{aligned}&\mathrm{blkdiag}\big(\mathrm{Ph}(\delta_1), \cdots, \mathrm{Ph}(\delta_{2^{n-1}})\big) P_1^{n-1} \\ &\quad = \mathrm{diag}\big(e^{-i\beta_1/2}, e^{-i\beta_1/2}, e^{i\beta_1/2}, e^{i\beta_1/2}, \\ &\qquad\qquad e^{-i\beta_2/2}, e^{-i\beta_2/2}, e^{i\beta_2/2}, e^{i\beta_2/2}, \cdots \\ &\qquad\qquad \cdots, e^{-i\beta_{2^{n-2}}/2}, e^{-i\beta_{2^{n-2}}/2}, e^{i\beta_{2^{n-2}}/2}, e^{i\beta_{2^{n-2}}/2}\big)\end{aligned}$$

which is a multi-controlled gate of type $F_n^n(\mathrm{R}_z)$ (with $\theta_{2j-1} = \theta_{2j}$ for all $1 \leqslant j \leqslant 2^{n-1}$). This shows that $U_1^{n-1} P_1^{n-1}$ is a product of multi-controlled rotation gates in which the rotation gates involved are R_y and R_z.

§3.7 Unitary Decomposition

Suppose that $P_1^{n-1}, \cdots, P_{j-1}^{n-1}$ are specified so that $U_2^{n-1}, \cdots, U_j^{n-1}$ are determined accordingly. Since U_j^{n-1} is also a block diagonal matrix consisting of 2^{n-1} blocks of 2×2 unitary matrices U_{j1}^{n-1}, $\cdots$, $U_{j2^{n-1}}^{n-1}$, we can decompose U_j^{n-1} as

$$\begin{bmatrix} \mathrm{R}_z(\xi_1) & & \\ & \ddots & \\ & & \mathrm{R}_z(\xi_{2^{n-1}}) \end{bmatrix} \begin{bmatrix} \mathrm{R}_y(\theta_1) & & \\ & \ddots & \\ & & \mathrm{R}_y(\theta_{2^{n-1}}) \end{bmatrix} \begin{bmatrix} \mathrm{R}_z(\eta_1) & & \\ & \ddots & \\ & & \mathrm{R}_z(\eta_{2^{n-1}}) \end{bmatrix} \begin{bmatrix} \mathrm{Ph}(\delta_1) & & \\ & \ddots & \\ & & \mathrm{Ph}(\delta_{2^{n-1}}) \end{bmatrix}.$$

for some $\xi_1, \cdots, \xi_{2^{n-1}}$, $\theta_1, \cdots, \theta_{2^{n-1}}$, $\eta_1, \cdots, \eta_{2^{n-1}}$ and $\delta_1, \cdots, \delta_{2^{n-1}}$. We note that these ξ_j's, θ_j's, η_j's and δ_j's are in principle different from those values used in the decomposition of U_1^{n-1}, $\cdots$, U_{j-1}^{n-1}.

§3.7 Unitary Decomposition

Suppose that $P_1^{n-1}, \cdots, P_{j-1}^{n-1}$ are specified so that $U_2^{n-1}, \cdots, U_j^{n-1}$ are determined accordingly. Since U_j^{n-1} is also a block diagonal matrix consisting of 2^{n-1} blocks of 2×2 unitary matrices U_{j1}^{n-1}, $\cdots$, $U_{j2^{n-1}}^{n-1}$, we can decompose U_j^{n-1} as

$$\begin{bmatrix} \mathrm{R}_z(\xi_1) & & \\ & \ddots & \\ & & \mathrm{R}_z(\xi_{2^{n-1}}) \end{bmatrix} \begin{bmatrix} \mathrm{R}_y(\theta_1) & & \\ & \ddots & \\ & & \mathrm{R}_y(\theta_{2^{n-1}}) \end{bmatrix} \begin{bmatrix} \mathrm{R}_z(\eta_1) & & \\ & \ddots & \\ & & \mathrm{R}_z(\eta_{2^{n-1}}) \end{bmatrix} \begin{bmatrix} \mathrm{Ph}(\delta_1) & & \\ & \ddots & \\ & & \mathrm{Ph}(\delta_{2^{n-1}}) \end{bmatrix}.$$

for some $\xi_1, \cdots, \xi_{2^{n-1}}, \theta_1, \cdots, \theta_{2^{n-1}}, \eta_1, \cdots, \eta_{2^{n-1}}$ and $\delta_1, \cdots, \delta_{2^{n-1}}$.

We note that these ξ_j's, θ_j's, η_j's and δ_j's are in principle different from those values used in the decomposition of $U_1^{n-1}, \cdots, U_{j-1}^{n-1}$.

§3.7 Unitary Decomposition

Suppose that $P_1^{n-1}, \cdots, P_{j-1}^{n-1}$ are specified so that $U_2^{n-1}, \cdots, U_j^{n-1}$ are determined accordingly. Since U_j^{n-1} is also a block diagonal matrix consisting of 2^{n-1} blocks of 2×2 unitary matrices U_{j1}^{n-1}, $\cdots$, $U_{j2^{n-1}}^{n-1}$, we can decompose U_j^{n-1} as

$$\begin{bmatrix} \mathrm{R}_z(\xi_1) & & \\ & \ddots & \\ & & \mathrm{R}_z(\xi_{2^{n-1}}) \end{bmatrix} \begin{bmatrix} \mathrm{R}_y(\theta_1) & & \\ & \ddots & \\ & & \mathrm{R}_y(\theta_{2^{n-1}}) \end{bmatrix} \begin{bmatrix} \mathrm{R}_z(\eta_1) & & \\ & \ddots & \\ & & \mathrm{R}_z(\eta_{2^{n-1}}) \end{bmatrix} \begin{bmatrix} \mathrm{Ph}(\delta_1) & & \\ & \ddots & \\ & & \mathrm{Ph}(\delta_{2^{n-1}}) \end{bmatrix}.$$

for some $\xi_1, \cdots, \xi_{2^{n-1}}, \theta_1, \cdots, \theta_{2^{n-1}}, \eta_1, \cdots, \eta_{2^{n-1}}$ and $\delta_1, \cdots, \delta_{2^{n-1}}$. We note that these ξ_j's, θ_j's, η_j's and δ_j's are in principle different from those values used in the decomposition of U_1^{n-1}, $\cdots$, U_{j-1}^{n-1}.

§3.7 Unitary Decomposition

Next we look for P_j^{n-1}, a block diagonal matrix of the form

$$\mathrm{blkdiag}\Big(Q_1, Q_1, Q_2, Q_2, \cdots, Q_{2^{n-\gamma(j)-1}}, Q_{2^{n-\gamma(j)-1}}\Big)\,,$$

where each Q_k is a diagonal matrix of size $2^{\gamma(j)} \times 2^{\gamma(j)}$, so that $U_j^{n-1} P_j^{n-1}$ is the product of multi-controlled rotation gates.

For $1 \leqslant k \leqslant n - \gamma(j) - 1$ and $1 \leqslant \ell \leqslant 2^{\gamma(j)}$, let

$$\alpha_{(k-1)2^{\gamma(j)}+\ell} = -\frac{1}{2}\Big(\delta_{(k-1)2^{\gamma(j)}+[\frac{\ell+1}{2}]} + \delta_{(k-1)2^{\gamma(j)}+2^{\gamma(j)-1}+[\frac{\ell+1}{2}]}\Big)\,,$$

where $[\dfrac{\ell+1}{2}]$ in the sub-index denotes the largest integer which is not greater than $\dfrac{\ell+1}{2}$. Define

$$Q_k = \mathrm{diag}\Big(e^{i\alpha_{2^{(k-1)\gamma(j)}+1}}, \cdots, e^{i\alpha_{2^{k\gamma(j)}}}\Big)$$

and

$$P_j^{n-1} = \mathrm{blkdiag}\Big(Q_1, Q_1, Q_2, Q_2, \cdots, Q_{2^{n-\gamma(j)-1}}, Q_{2^{n-\gamma(j)-1}}\Big)\,.$$

§3.7 Unitary Decomposition

Next we look for P_j^{n-1}, a block diagonal matrix of the form

$$\mathrm{blkdiag}\big(Q_1, Q_1, Q_2, Q_2, \cdots, Q_{2^{n-\gamma(j)-1}}, Q_{2^{n-\gamma(j)-1}}\big)\,,$$

where each Q_k is a diagonal matrix of size $2^{\gamma(j)} \times 2^{\gamma(j)}$, so that $U_j^{n-1} P_j^{n-1}$ is the product of multi-controlled rotation gates.

For $1 \leqslant k \leqslant n - \gamma(j) - 1$ and $1 \leqslant \ell \leqslant 2^{\gamma(j)}$, let

$$\alpha_{(k-1)2^{\gamma(j)}+\ell} = -\frac{1}{2}\Big(\delta_{(k-1)2^{\gamma(j)}+[\frac{\ell+1}{2}]} + \delta_{(k-1)2^{\gamma(j)}+2^{\gamma(j)-1}+[\frac{\ell+1}{2}]}\Big)\,,$$

where $[\dfrac{\ell+1}{2}]$ in the sub-index denotes the largest integer which is not greater than $\dfrac{\ell+1}{2}$. Define

$$Q_k = \mathrm{diag}\big(e^{i\alpha_{2(k-1)\gamma(j)+1}}, \cdots, e^{i\alpha_{2^{k\gamma(j)}}}\big)$$

and

$$P_j^{n-1} = \mathrm{blkdiag}\big(Q_1, Q_1, Q_2, Q_2, \cdots, Q_{2^{n-\gamma(j)-1}}, Q_{2^{n-\gamma(j)-1}}\big)\,.$$

§3.7 Unitary Decomposition

Then $\mathrm{blkdiag}\big(\mathrm{Ph}(\delta_1),\cdots,\mathrm{Ph}(\delta_{2^{n-1}})\big)P_j^{n-1}$, with N denoting $2^{\gamma(j)}$, takes the form

$$\begin{aligned}&\mathrm{diag}\big(e^{-i\beta_1/2}, e^{-i\beta_2/2},\cdots, e^{-i\beta_N/2}, e^{i\beta_1/2}, e^{i\beta_2/2},\cdots, e^{i\beta_N/2},\\ &\quad e^{-i\beta_{N+1}/2}, e^{-\beta_{N+2}/2},\cdots, e^{-i\beta_{2N}/2}, e^{i\beta_{N+1}/2}, e^{\beta_{N+2}/2},\cdots, e^{i\beta_{2N}/2},\\ &\quad \cdots, e^{-i\beta_{2^{n-1}-N+1}/2}, e^{-\beta_{2^{n-1}-N+2}/2},\cdots, e^{-i\beta_{2^{n-1}}/2},\\ &\qquad\quad e^{i\beta_{2^{n-1}-N+1}/2}, e^{\beta_{2^{n-1}-N+2}/2},\cdots, e^{i\beta_{2^{n-1}}/2}\big)\end{aligned}$$

for some $\beta_1,\cdots,\beta_{2^{n-1}}\in\mathbb{R}$. This is a **multi-controlled gate of type** $F^n_{n-\gamma(j)}(\mathrm{R}_z)$. Therefore, $U_j^{n-1}P_j^{n-1}$ is the product of multi-controlled gates in which the rotation gates involved are R_y and R_z.

§3.7 Unitary Decomposition

Then $\mathrm{blkdiag}\big(\mathrm{Ph}(\delta_1),\cdots,\mathrm{Ph}(\delta_{2^{n-1}})\big)P_j^{n-1}$, with N denoting $2^{\gamma(j)}$, takes the form

$$\begin{aligned}\mathrm{diag}\big(&e^{-i\beta_1/2}, e^{-i\beta_2/2}, \cdots, e^{-i\beta_N/2}, e^{i\beta_1/2}, e^{i\beta_2/2}, \cdots, e^{i\beta_N/2},\\ &e^{-i\beta_{N+1}/2}, e^{-\beta_{N+2}/2}, \cdots, e^{-i\beta_{2N}/2}, e^{i\beta_{N+1}/2}, e^{\beta_{N+2}/2}, \cdots, e^{i\beta_{2N}/2},\\ &\cdots, e^{-i\beta_{2^{n-1}-N+1}/2}, e^{-\beta_{2^{n-1}-N+2}/2}, \cdots, e^{-i\beta_{2^{n-1}}/2},\\ &\qquad e^{i\beta_{2^{n-1}-N+1}/2}, e^{\beta_{2^{n-1}-N+2}/2}, \cdots, e^{i\beta_{2^{n-1}}/2}\big)\end{aligned}$$

for some $\beta_1,\cdots,\beta_{2^{n-1}}\in\mathbb{R}$. This is a **multi-controlled gate of type** $F^n_{n-\gamma(j)}(\mathrm{R}_z)$. Therefore, $U_j^{n-1}P_j^{n-1}$ is the product of multi-controlled gates in which the rotation gates involved are R_y and R_z.

§3.7 Unitary Decomposition

Let U be an n-qubit gate (or equivalently, $2^n \times 2^n$ unitary matrix). Recall that

$$U = \Big(\prod_{j=1}^{2^{n-1}-1} U_j^{n-1} P_j^{n-1} A_j^{n-1} \Big) U_{2^{n-1}}^{n-1} \,,$$

where U_j^{n-1} is a block diagonal of 2×2 matrix for all j, and A_j^i takes the form

$$A_j^{n-1} = \mathrm{blkdiag}\left(\begin{bmatrix} C_1 & -S_1 \\ S_1 & C_1 \end{bmatrix}, \cdots\cdots, \begin{bmatrix} C_{2^{n-\gamma(j)-1}} & -S_{2^{n-\gamma(j)-1}} \\ S_{2^{n-\gamma(j)-1}} & C_{2^{n-\gamma(j)-1}} \end{bmatrix} \right) ,$$

From the argument above, we know that $U_j^{n-1} P_j^{n-1}$ is the product of multi-controlled gates, while each A_j^{n-1} is a multi-controlled gate of type $F_{n-\gamma(j)}^n(\mathrm{R}_y)$. Therefore, in order to implement the quantum gate with matrix representation U using quantum circuits, it suffices to consider how to implement a multi-controlled gate in which the rotation gate involved is R_y or R_z.

§3.7 Unitary Decomposition

Let U be an n-qubit gate (or equivalently, $2^n \times 2^n$ unitary matrix). Recall that

$$U = \Big(\prod_{j=1}^{2^{n-1}-1} U_j^{n-1} P_j^{n-1} A_j^{n-1} \Big) U_{2^{n-1}}^{n-1} \,,$$

where U_j^{n-1} is a block diagonal of 2×2 matrix for all j, and A_j^i takes the form

$$A_j^{n-1} = \operatorname{blkdiag}\left(\begin{bmatrix} C_1 & -S_1 \\ S_1 & C_1 \end{bmatrix}, \cdots\cdots, \begin{bmatrix} C_{2^{n-\gamma(j)-1}} & -S_{2^{n-\gamma(j)-1}} \\ S_{2^{n-\gamma(j)-1}} & C_{2^{n-\gamma(j)-1}} \end{bmatrix} \right) ,$$

From the argument above, we know that $U_j^{n-1} P_j^{n-1}$ is the product of multi-controlled gates, while each A_j^{n-1} is a multi-controlled gate of type $F_{n-\gamma(j)}^n(\mathrm{R}_y)$. Therefore, in order to implement the quantum gate with matrix representation U using quantum circuits, it suffices to consider how to implement a multi-controlled gate in which the rotation gate involved is R_y or R_z.

§3.7 Unitary Decomposition

Theorem

Each $2^{n+1} \times 2^{n+1}$ unitary matrix can be expressed as the product of multi-controlled rotation gates of type $F_k^n(\mathrm{R}_y)$ and $F_k^n(\mathrm{R}_z)$, $k = 1, 2, \cdots, n+1$.

Recall that the multi-controlled rotation gates of type $F_{j+1}^n(R_{\boldsymbol{a}})$ are $(n+1)$-qubit gates L satisfying

$$L(|x_0\rangle\otimes\cdots\otimes|x_n\rangle) = |x_0\rangle\otimes\cdots\otimes|x_{j-1}\rangle\otimes(R_{\boldsymbol{a}}(\phi_k)|x_j\rangle)\otimes|x_{j+1}\rangle\otimes\cdots\otimes|x_0\rangle$$

if $(x_0\cdots x_{j-1}x_{j+1}\cdots x_n)_2 = k$, where $\boldsymbol{a} = (a_x, a_y, a_z)$ is a unit vector in $\mathbb{R}^3$ and $R_{\boldsymbol{a}}$ is a 1-qubit gate given by

$$R_{\boldsymbol{a}}(\phi) = \begin{bmatrix} \cos\frac{\phi}{2} + ia_z\sin\frac{\phi}{2} & (a_y + ia_x)\sin\frac{\phi}{2} \\ -(a_y - ia_x)\sin\frac{\phi}{2} & \cos\frac{\phi}{2} - ia_z\sin\frac{\phi}{2} \end{bmatrix}. \tag{8}$$

We also write $R_{\boldsymbol{a}}$ as R_y or R_z if $\boldsymbol{a} = (0, -1, 0)$ or $\boldsymbol{a} = (0, 0, -1)$.

§3.8 Implementation of Multi-Controlled Rotation Gates

In this section we are concerned with the implementation of multi-controlled rotation gates of type $F^n_{n+1}(R_{\boldsymbol{a}})$ with unit vector $\boldsymbol{a} = (0, a_y, a_z)$ using quantum circuits. The implementation of multi-controlled gate of this type is the building block of the implementation of general quantum gates. We note that multi-controlled rotation gate of type $F^n_k(R_{\boldsymbol{a}})$, where $1 \leqslant k \leqslant n$, can be obtained by applying several swap operations on multi-controlled rotation gate of type $F^n_{n+1}(R_{\boldsymbol{a}})$; thus arbitrary multi-controlled rotation gates can also be implemented even though we only focus on the case of $F^n_{n+1}(R_{\boldsymbol{a}})$.

§3.8 Implementation of Multi-Controlled Rotation Gates

In this section we are concerned with the implementation of multi-controlled rotation gates of type $F^n_{n+1}(R_{\boldsymbol{a}})$ with unit vector $\boldsymbol{a} = (0, a_y, a_z)$ using quantum circuits. The implementation of multi-controlled gate of this type is the building block of the implementation of general quantum gates. We note that multi-controlled rotation gate of type $F^n_k(R_{\boldsymbol{a}})$, where $1 \leqslant k \leqslant n$, can be obtained by applying several swap operations on multi-controlled rotation gate of type $F^n_{n+1}(R_{\boldsymbol{a}})$; thus arbitrary multi-controlled rotation gates can also be implemented even though we only focus on the case of $F^n_{n+1}(R_{\boldsymbol{a}})$.

§3.8 Implementation of Multi-Controlled Rotation Gates

Recall that the matrix representation of multi-controlled rotation gates of type $F^n_{n+1}(R_{\boldsymbol{a}})$ takes the form

$$R = \mathrm{blkdiag}\big(R_{\boldsymbol{a}}(\phi_1), \cdots, R_{\boldsymbol{a}}(\phi_n)\big) \quad \text{with} \quad \boldsymbol{a} = (0, a_y, a_z)\,, \qquad (9)$$

where for a given unit vector $\boldsymbol{a} = (a_x, a_y, a_z)$ and angle ϕ, the rotation matrix $R_{\boldsymbol{a}}(\phi)$ is given by (8) or equivalently,

$$R_{\boldsymbol{a}}(\phi) = \mathrm{I}\cos\frac{\phi}{2} + i\,(a_x\sigma_x + a_y\sigma_y + a_z\sigma_z)\sin\frac{\phi}{2}\,,$$

in which σ_x, σ_y and σ_z are the Pauli matrices

$$\sigma_x = \mathrm{X} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_y = \mathrm{Y} = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma_z = \mathrm{Z} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Therefore, it suffices to consider the implementation of a unitary matrix of form (9).

§3.8 Implementation of Multi-Controlled Rotation Gates

- **Some properties of $R_{\boldsymbol{a}}(\phi)$:**

1. $\mathrm{R}_y(\phi) = R_{(0,1,0)}(-\phi) = R_{(0,-1,0)}(\phi)$ for all angle ϕ.
2. $\mathrm{R}_z(\phi) = R_{(0,0,1)}(-\phi) = R_{(0,0,-1)}(\phi)$ for all angle ϕ.
3. $R_{\boldsymbol{a}}(\phi)^\dagger = R_{\boldsymbol{a}}(-\phi)$ for all unit vectors $\boldsymbol{a} \in \mathbb{R}^3$ and angle ϕ.
4. $R_{\boldsymbol{a}}(\phi)$ is unitary for all unit vectors $\boldsymbol{a} \in \mathbb{R}^3$ and angle ϕ.
5. $R_{\boldsymbol{a}}(\theta)R_{\boldsymbol{a}}(\phi) = R_{\boldsymbol{a}}(\theta + \phi)$ for all unit vectors $\boldsymbol{a} \in \mathbb{R}$ and angles θ, ϕ.
6. $\mathrm{X}R_{\boldsymbol{a}}(\phi)\mathrm{X} = R_{\boldsymbol{a}}(-\phi)$ for all unit vectorss $\boldsymbol{a} = (0, a_y, a_z) \in \mathbb{R}^3$ and angle ϕ.

Such operator $R_{\boldsymbol{a}}(\phi)$ is called the rotation (of a qubit) about the three-dimensional vector $\boldsymbol{a}$ with angle ϕ (on the Bloch sphere).

§3.8 Implementation of Multi-Controlled Rotation Gates

We use the following example to illustrate the idea of how a quantum gate of the form in (9) is implemented using quantum circuits.

Example (4-qubit gate decomposition)

Let $\boldsymbol{a} = (0, a_y, a_z)$ be a unit vector in $\mathbb{R}^3$. Consider the multi-controlled 4-qubit gate given by

$$|k\rangle \otimes |y\rangle \mapsto |k\rangle \otimes (R_{\boldsymbol{a}}(\alpha_{k+1})|y\rangle) \qquad \text{for all } 0 \leqslant k \leqslant 7$$

whose matrix representation is given by

$$\mathrm{blkdiag}(R_{\boldsymbol{a}}(\alpha_1), R_{\boldsymbol{a}}(\alpha_2), \cdots, R_{\boldsymbol{a}}(\alpha_8)) .$$

§3.8 Implementation of Multi-Controlled Rotation Gates

Example (4-qubit gate decomposition (cont.))

Goal: Find $C_1, C_2, \cdots, C_k \in \{\mathbf{CNOT}_{1,4}, \mathbf{CNOT}_{2,4}, \mathbf{CNOT}_{3,4}\}$ and $R_1, \cdots, R_k$ of the form $R_j = \text{blkdiag}\Big(\underbrace{R_{\boldsymbol{a}}(\theta_j), R_{\boldsymbol{a}}(\theta_j), \cdots, R_{\boldsymbol{a}}(\theta_j)}_{8 \text{ copies of } R_{\boldsymbol{a}}(\theta_j)}\Big)$

(which is the matrix representation of $\mathrm{I}_2 \otimes \mathrm{I}_2 \otimes \mathrm{I}_2 \otimes R_{\boldsymbol{a}}(\theta_j)$) so that

$$\text{blkdiag}(R_{\boldsymbol{a}}(\alpha_1), \cdots, R_{\boldsymbol{a}}(\alpha_8)) = C_8 R_8 C_7 R_7 \cdots C_2 R_2 C_1 R_1 .$$

Recall that $\mathbf{CNOT}_{i,4}$ denotes the controlled-not gate whose control qubit is the i-th qubit while the target qubit is the 4-th qubit, and the matrix representation of $\mathbf{CNOT}_{i,4}$ are given by

$$\mathbf{CNOT}_{1,4} = \text{blkdiag}(\mathrm{I}_2, \mathrm{I}_2, \mathrm{I}_2, \mathrm{I}_2, \mathrm{X}, \mathrm{X}, \mathrm{X}, \mathrm{X}) ,$$
$$\mathbf{CNOT}_{2,4} = \text{blkdiag}(\mathrm{I}_2, \mathrm{I}_2, \mathrm{X}, \mathrm{X}, \mathrm{I}_2, \mathrm{I}_2, \mathrm{X}, \mathrm{X}) ,$$
$$\mathbf{CNOT}_{3,4} = \text{blkdiag}(\mathrm{I}_2, \mathrm{X}, \mathrm{I}_2, \mathrm{X}, \mathrm{I}_2, \mathrm{X}, \mathrm{I}_2, \mathrm{X}) .$$

§3.8 Implementation of Multi-Controlled Rotation Gates

Example (4-qubit gate decomposition (cont.))

Goal: Find $C_1, C_2, \cdots, C_k \in \{\mathbf{CNOT}_{1,4}, \mathbf{CNOT}_{2,4}, \mathbf{CNOT}_{3,4}\}$ and $R_1, \cdots, R_k$ of the form $R_j = \text{blkdiag}\big(\underbrace{R_{\boldsymbol{a}}(\theta_j), R_{\boldsymbol{a}}(\theta_j), \cdots, R_{\boldsymbol{a}}(\theta_j)}_{8 \text{ copies of } R_{\boldsymbol{a}}(\theta_j)}\big)$ (which is the matrix representation of $\mathrm{I}_2 \otimes \mathrm{I}_2 \otimes \mathrm{I}_2 \otimes R_{\boldsymbol{a}}(\theta_j)$) so that

$$\text{blkdiag}(R_{\boldsymbol{a}}(\alpha_1), \cdots, R_{\boldsymbol{a}}(\alpha_8)) = C_8 R_8 C_7 R_7 \cdots C_2 R_2 C_1 R_1 \,.$$

Recall that $\mathbf{CNOT}_{i,4}$ denotes the controlled-not gate whose control qubit is the i-th qubit while the target qubit is the 4-th qubit, and the matrix representation of $\mathbf{CNOT}_{i,4}$ are given by

$$\begin{aligned} \mathbf{CNOT}_{1,4} &= \text{blkdiag}(\mathrm{I}_2, \mathrm{I}_2, \mathrm{I}_2, \mathrm{I}_2, \mathrm{X}, \mathrm{X}, \mathrm{X}, \mathrm{X}) \,, \\ \mathbf{CNOT}_{2,4} &= \text{blkdiag}(\mathrm{I}_2, \mathrm{I}_2, \mathrm{X}, \mathrm{X}, \mathrm{I}_2, \mathrm{I}_2, \mathrm{X}, \mathrm{X}) \,, \\ \mathbf{CNOT}_{3,4} &= \text{blkdiag}(\mathrm{I}_2, \mathrm{X}, \mathrm{I}_2, \mathrm{X}, \mathrm{I}_2, \mathrm{X}, \mathrm{I}_2, \mathrm{X}) \,. \end{aligned}$$

§3.8 Implementation of Multi-Controlled Rotation Gates

Example (4-qubit gate decomposition (cont.))

Define $\widetilde{R}_k$ by

$$\widetilde{R}_k = C_8 C_7 \cdots C_k R_k C_k C_{k+1} \cdots C_8 \,.$$

Since $C_j C_j = \mathrm{I}_{16}$ and $C_j C_k = C_k C_j$ for all $1 \leqslant j, k \leqslant 8$, we find that

$$\begin{aligned} &\widetilde{R}_8 \widetilde{R}_7 \cdots \widetilde{R}_1 \\ &\quad = (C_8 R_8 C_8)(C_8 C_7 R_7 C_7 C_8)(C_8 C_7 C_6 R_6 C_6 C_7 C_8) \cdots \\ &\qquad \cdots (C_8 \cdots C_1 R_1 C_1 \cdots C_8) \\ &\quad = (C_8 R_8)(C_7 R_7 C_7)(C_7 C_6 R_6 C_6 C_7) \cdots (C_7 \cdots C_1 R_1 C_1 \cdots C_8) \\ &\quad = (C_8 R_8)(C_7 R_7)(C_6 R_6 C_6) \cdots (C_6 \cdots C_1 R_1 C_1 \cdots C_8) \\ &\quad = \cdots\cdots\cdots \\ &\quad = (C_8 R_8)(C_7 R_7)(C_6 R_6) \cdots (C_1 R_1)(C_1 \cdots C_8) \end{aligned}$$

so that

$$C_8 R_8 C_7 R_7 \cdots C_1 R_1 = \widetilde{R}_8 \widetilde{R}_7 \cdots \widetilde{R}_1 \cdot (C_1 C_2 \cdots C_8) \,.$$

§3.8 Implementation of Multi-Controlled Rotation Gates

Example (4-qubit gate decomposition (cont.))

By the fact that $\mathrm{X}R_{\mathbf{a}}(\theta)\mathrm{X} = R_{\mathbf{a}}(-\theta)$ for all $\theta \in \mathbb{R}$, for all $\phi_1, \phi_2, \cdots, \phi_8 \in \mathbb{R}$ we have we have

$$\mathbf{CNOT}_{1,4} \cdot \mathrm{blkdiag}\big(R_{\mathbf{a}}(\phi_1), R_{\mathbf{a}}(\phi_2), \cdots, R_{\mathbf{a}}(\phi_8)\big) \cdot \mathbf{CNOT}_{1,4}$$
$$= \mathrm{blkdiag}\big(R_{\mathbf{a}}(\phi_1), R_{\mathbf{a}}(\phi_2), R_{\mathbf{a}}(\phi_3), R_{\mathbf{a}}(\phi_4), R_{\mathbf{a}}(-\phi_5), R_{\mathbf{a}}(-\phi_6), R_{\mathbf{a}}(-\phi_7), R_{\mathbf{a}}(-\phi_8)\big),$$
$$\mathbf{CNOT}_{2,4} \cdot \mathrm{blkdiag}\big(R_{\mathbf{a}}(\phi_1), R_{\mathbf{a}}(\phi_2), \cdots, R_{\mathbf{a}}(\phi_8)\big) \cdot \mathbf{CNOT}_{2,4}$$
$$= \mathrm{blkdiag}\big(R_{\mathbf{a}}(\phi_1), R_{\mathbf{a}}(\phi_2), R_{\mathbf{a}}(-\phi_3), R_{\mathbf{a}}(-\phi_4), R_{\mathbf{a}}(\phi_5), R_{\mathbf{a}}(\phi_6), R_{\mathbf{a}}(-\phi_7), R_{\mathbf{a}}(-\phi_8)\big),$$
$$\mathbf{CNOT}_{3,4} \cdot \mathrm{blkdiag}\big(R_{\mathbf{a}}(\phi_1), R_{\mathbf{a}}(\phi_2), \cdots, R_{\mathbf{a}}(\phi_8)\big) \cdot \mathbf{CNOT}_{3,4}$$
$$= \mathrm{blkdiag}\big(R_{\mathbf{a}}(\phi_1), R_{\mathbf{a}}(-\phi_2), R_{\mathbf{a}}(\phi_3), R_{\mathbf{a}}(-\phi_4), R_{\mathbf{a}}(\phi_5), R_{\mathbf{a}}(-\phi_6), R_{\mathbf{a}}(\phi_7), R_{\mathbf{a}}(-\phi_8)\big).$$

Therefore, $\widetilde{R}_k$ must take the form

$$\mathrm{blkdiag}\big(R_{\mathbf{a}}(b_{k1}\theta_k), R_{\mathbf{a}}(b_{k2}\theta_k), \cdots, R_{\mathbf{a}}(b_{k8}\theta_k)\big),$$

where $b_{kj} = \pm 1$ and b_{kj} is determined by $C_k, \cdots, C_8$.

§3.8 Implementation of Multi-Controlled Rotation Gates

Example (4-qubit gate decomposition (cont.))

By the fact that $\mathrm{X}R_{\boldsymbol{a}}(\theta)\mathrm{X} = R_{\boldsymbol{a}}(-\theta)$ for all $\theta \in \mathbb{R}$, for all $\phi_1, \phi_2, \cdots, \phi_8 \in \mathbb{R}$ we have we have

$$\begin{aligned}
&\textbf{CNOT}_{1,4} \cdot \mathrm{blkdiag}\big(R_{\boldsymbol{a}}(\phi_1), R_{\boldsymbol{a}}(\phi_2), \cdots, R_{\boldsymbol{a}}(\phi_8)\big) \cdot \textbf{CNOT}_{1,4} \\
&\quad = \mathrm{blkdiag}\big(R_{\boldsymbol{a}}(\phi_1), R_{\boldsymbol{a}}(\phi_2), R_{\boldsymbol{a}}(\phi_3), R_{\boldsymbol{a}}(\phi_4), R_{\boldsymbol{a}}(-\phi_5), R_{\boldsymbol{a}}(-\phi_6), R_{\boldsymbol{a}}(-\phi_7), R_{\boldsymbol{a}}(-\phi_8)\big), \\
&\textbf{CNOT}_{2,4} \cdot \mathrm{blkdiag}\big(R_{\boldsymbol{a}}(\phi_1), R_{\boldsymbol{a}}(\phi_2), \cdots, R_{\boldsymbol{a}}(\phi_8)\big) \cdot \textbf{CNOT}_{2,4} \\
&\quad = \mathrm{blkdiag}\big(R_{\boldsymbol{a}}(\phi_1), R_{\boldsymbol{a}}(\phi_2), R_{\boldsymbol{a}}(-\phi_3), R_{\boldsymbol{a}}(-\phi_4), R_{\boldsymbol{a}}(\phi_5), R_{\boldsymbol{a}}(\phi_6), R_{\boldsymbol{a}}(-\phi_7), R_{\boldsymbol{a}}(-\phi_8)\big), \\
&\textbf{CNOT}_{3,4} \cdot \mathrm{blkdiag}\big(R_{\boldsymbol{a}}(\phi_1), R_{\boldsymbol{a}}(\phi_2), \cdots, R_{\boldsymbol{a}}(\phi_8)\big) \cdot \textbf{CNOT}_{3,4} \\
&\quad = \mathrm{blkdiag}\big(R_{\boldsymbol{a}}(\phi_1), R_{\boldsymbol{a}}(-\phi_2), R_{\boldsymbol{a}}(\phi_3), R_{\boldsymbol{a}}(-\phi_4), R_{\boldsymbol{a}}(\phi_5), R_{\boldsymbol{a}}(-\phi_6), R_{\boldsymbol{a}}(\phi_7), R_{\boldsymbol{a}}(-\phi_8)\big).
\end{aligned}$$

Therefore, $\widetilde{R}_k$ must take the form

$$\mathrm{blkdiag}\big(R_{\boldsymbol{a}}(b_{k1}\theta_k), R_{\boldsymbol{a}}(b_{k2}\theta_k), \cdots, R_{\boldsymbol{a}}(b_{k8}\theta_k)\big),$$

where $b_{kj} = \pm 1$ and b_{kj} is determined by $C_k, \cdots, C_8$.

§3.8 Implementation of Multi-Controlled Rotation Gates

Example (4-qubit gate decomposition (cont.))

In fact, if $\boldsymbol{r}_j$ is the symbol of C_j, then

$$\boldsymbol{b}_k \equiv \left[\, b_{k1}, b_{k2}, \cdots, b_{k8} \,\right] = \boldsymbol{r}_k \mathbin{.*} \cdots \mathbin{.*} \boldsymbol{r}_7 \mathbin{.*} \boldsymbol{r}_8$$

where $.*$ denotes the Hadamard product. We recall that the symbols of **CNOT** is connected the 8×8 Hadamard matrix

$$\mathrm{M} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{bmatrix}.$$

If the k-th row of M is denoted by S_{k-1}, then the symbols for $\textbf{CNOT}_{3,4}$, $\textbf{CNOT}_{2,4}$ and $\textbf{CNOT}_{1,4}$ are S_1, S_2 and S_4, respectively.

§3.8 Implementation of Multi-Controlled Rotation Gates

Example (4-qubit gate decomposition (cont.))

In fact, if $\boldsymbol{r}_j$ is the symbol of C_j, then

$$\boldsymbol{b}_k \equiv \left[b_{k1}, b_{k2}, \cdots, b_{k8} \right] = \boldsymbol{r}_k .* \cdots .* \boldsymbol{r}_7 .* \boldsymbol{r}_8$$

where $.*$ denotes the Hadamard product. We recall that the symbols of **CNOT** is connected the 8×8 Hadamard matrix

$$\mathrm{M} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{bmatrix}.$$

If the k-th row of M is denoted by S_{k-1}, then the symbols for **CNOT**$_{3,4}$, **CNOT**$_{2,4}$ and **CNOT**$_{1,4}$ are S_1, S_2 and S_4, respectively.

§3.8 Implementation of Multi-Controlled Rotation Gates

Example (4-qubit gate decomposition (cont.))

Note that the identity $R_{\boldsymbol{a}}(\theta)R_{\boldsymbol{a}}(\phi) = R_{\boldsymbol{a}}(\theta+\phi)$ implies that

$$\begin{aligned}
\widetilde{R}_8\widetilde{R}_7\cdots\widetilde{R}_1 &= \mathrm{blkdiag}\big(R_{\boldsymbol{a}}(b_{81}\theta_8), R_{\boldsymbol{a}}(b_{82}\theta_8),\cdots, R_{\boldsymbol{a}}(b_{88}\theta_8)\big)\\
&\quad\cdot\mathrm{blkdiag}\big(R_{\boldsymbol{a}}(b_{71}\theta_7), R_{\boldsymbol{a}}(b_{72}\theta_7),\cdots, R_{\boldsymbol{a}}(b_{78}\theta_7)\big)\\
&\quad\cdots\cdot\mathrm{blkdiag}\big(R_{\boldsymbol{a}}(b_{11}\theta_1), R_{\boldsymbol{a}}(b_{12}\theta_1),\cdots, R_{\boldsymbol{a}}(b_{18}\theta_1)\big)\\
&= \mathrm{blkdiag}\big(R_{\boldsymbol{a}}(b_{81}\theta_8 + b_{71}\theta_7 + \cdots + b_{11}\theta_1),\\
&\qquad\qquad R_{\boldsymbol{a}}(b_{82}\theta_8 + b_{72}\theta_7 + \cdots + b_{12}\theta_1),\cdots,\\
&\qquad\qquad R_{\boldsymbol{a}}(b_{88}\theta_8 + b_{78}\theta_8 + \cdots + b_{18}\theta_1)\big)\,.
\end{aligned}$$

This computation motivates us to choose $\boldsymbol{r}_1,\cdots,\boldsymbol{r}_8 \in \{S_1, S_2, S_4\}$ such that $\boldsymbol{b}_k$ given by $\boldsymbol{b}_k = \boldsymbol{r}_k .* \cdots .* \boldsymbol{r}_7 .* \boldsymbol{r}_8$ satisfies

1. $\boldsymbol{b}_1 = S_0$ (if so, then $C_1C_2\cdots C_8 = \mathrm{I}_{16}$ which implies that $C_8R_8C_7R_7\cdots C_1R_1 = \widetilde{R}_8\widetilde{R}_7\cdots\widetilde{R}_1$).
2. The collection $\{\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_8\}$ is linearly independent.

§3.8 Implementation of Multi-Controlled Rotation Gates

Example (4-qubit gate decomposition (cont.))

Note that the identity $R_{\boldsymbol{a}}(\theta)R_{\boldsymbol{a}}(\phi) = R_{\boldsymbol{a}}(\theta+\phi)$ implies that

$$\begin{aligned}
\widetilde{R}_8\widetilde{R}_7\cdots\widetilde{R}_1 &= \mathrm{blkdiag}\big(R_{\boldsymbol{a}}(b_{81}\theta_8), R_{\boldsymbol{a}}(b_{82}\theta_8), \cdots, R_{\boldsymbol{a}}(b_{88}\theta_8)\big)\\
&\quad\cdot \mathrm{blkdiag}\big(R_{\boldsymbol{a}}(b_{71}\theta_7), R_{\boldsymbol{a}}(b_{72}\theta_7), \cdots, R_{\boldsymbol{a}}(b_{78}\theta_7)\big)\\
&\quad\cdots\cdot \mathrm{blkdiag}\big(R_{\boldsymbol{a}}(b_{11}\theta_1), R_{\boldsymbol{a}}(b_{12}\theta_1), \cdots, R_{\boldsymbol{a}}(b_{18}\theta_1)\big)\\
&= \mathrm{blkdiag}\big(R_{\boldsymbol{a}}(b_{81}\theta_8 + b_{71}\theta_7 + \cdots + b_{11}\theta_1),\\
&\qquad\qquad R_{\boldsymbol{a}}(b_{82}\theta_8 + b_{72}\theta_7 + \cdots + b_{12}\theta_1), \cdots,\\
&\qquad\qquad R_{\boldsymbol{a}}(b_{88}\theta_8 + b_{78}\theta_8 + \cdots + b_{18}\theta_1)\big)\,.
\end{aligned}$$

This computation motivates us to choose $\boldsymbol{r}_1, \cdots, \boldsymbol{r}_8 \in \{S_1, S_2, S_4\}$ such that $\boldsymbol{b}_k$ given by $\boldsymbol{b}_k = \boldsymbol{r}_k \,.\!*\, \cdots \,.\!*\, \boldsymbol{r}_7 \,.\!*\, \boldsymbol{r}_8$ satisfies

1. $\boldsymbol{b}_1 = S_0$ (if so, then $C_1C_2\cdots C_8 = \mathrm{I}_{16}$ which implies that $C_8R_8C_7R_7\cdots C_1R_1 = \widetilde{R}_8\widetilde{R}_7\cdots\widetilde{R}_1$).
2. The collection $\{\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_8\}$ is linearly independent.

§3.8 Implementation of Multi-Controlled Rotation Gates

Example (4-qubit gate decomposition (cont.))

If we are able to find such $\boldsymbol{r}_1, \cdots, \boldsymbol{r}_8$, then we choose $\theta_1, \cdots, \theta_8$ satisfying

$$\begin{bmatrix} b_{11} & b_{21} & \cdots & b_{71} & b_{81} \\ b_{12} & b_{22} & \cdots & b_{72} & b_{82} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ b_{18} & b_{28} & \cdots & b_{78} & b_{88} \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_8 \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_8 \end{bmatrix}$$

whose solvability is guaranteed by property 2 above. Such θ_k's will then verify that

$$\begin{aligned} C_8 R_8 C_7 R_7 \cdots C_1 R_1 &= \widetilde{R}_8 \widetilde{R}_7 \cdots \widetilde{R}_1 \\ &= \operatorname{blkdiag}\big(R_{\boldsymbol{a}}(\alpha_1), R_{\boldsymbol{a}}(\alpha_2), \cdots, R_{\boldsymbol{a}}(\alpha_8)\big) \end{aligned}$$

and our goal is achieved.

§3.8 Implementation of Multi-Controlled Rotation Gates

Example (4-qubit gate decomposition (cont.))

If we are able to find such $\boldsymbol{r}_1, \cdots, \boldsymbol{r}_8$, then we choose $\theta_1, \cdots, \theta_8$ satisfying

$$\begin{bmatrix} b_{11} & b_{21} & \cdots & b_{71} & b_{81} \\ b_{12} & b_{22} & \cdots & b_{72} & b_{82} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ b_{18} & b_{28} & \cdots & b_{78} & b_{88} \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_8 \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_8 \end{bmatrix}$$

whose solvability is guaranteed by property 2 above. Such θ_k's will then verify that

$$\begin{aligned} C_8 R_8 C_7 R_7 \cdots C_1 R_1 &= \widetilde{R}_8 \widetilde{R}_7 \cdots \widetilde{R}_1 \\ &= \mathrm{blkdiag}\big(R_{\boldsymbol{a}}(\alpha_1), R_{\boldsymbol{a}}(\alpha_2), \cdots, R_{\boldsymbol{a}}(\alpha_8)\big) \end{aligned}$$

and our goal is achieved.

§3.8 Implementation of Multi-Controlled Rotation Gates

Example (4-qubit gate decomposition (cont.))

Finally, let us explain how to find $\boldsymbol{r}_1, \boldsymbol{r}_2, \cdots, \boldsymbol{r}_8 \in \{S_1, S_2, S_4\}$ satisfying the two properties above. First we note that

$$\mathrm{M} = \begin{bmatrix} \mathbf{ones}(1,8) \equiv S_0 \\ S_1 \\ S_2 \\ S_1 \ .* \ S_2 \\ S_4 \\ S_1 \ .* \ S_4 \\ S_2 \ .* \ S_4 \\ S_1 \ .* \ S_2 \ .* \ S_4 \end{bmatrix}.$$

Therefore, all rows of M can be generated by S_1, S_2 and S_4 using the Hadamard product $.*$ and we have

$$S_i \ .* \ S_j = S_{i+j} \quad \forall\, i, j \in \{1, 2, 4\} \quad \text{and} \quad S_1 \ .* \ S_2 \ .* \ S_4 = S_7 \,.$$

§3.8 Implementation of Multi-Controlled Rotation Gates

Example (4-qubit gate decomposition (cont.))

This allows us to write

$$S_\ell = S_1^{\ell_0} .* S_2^{\ell_1} .* S_4^{\ell_2} \quad \text{if} \quad 0 \leqslant \ell = (\ell_2\ell_1\ell_0)_2 \leqslant 7\,,$$

where $S_k^0 \equiv S_0$ for $k = 1, 2, 4$. By the fact that $S_i .* S_i = S_0$ and $S_i .* S_j = S_j .* S_i$ for $S_i, S_j \in \{S_1, S_2, S_4\}$,

$$S_i .* S_j = S_1^{i_0 \oplus j_0} .* S_2^{i_1 \oplus j_1} .* S_4^{i_2 \oplus j_2} \quad \text{if} \quad i = (i_2 i_1 i_0)_2, j = (j_2 j_1 j_0)_2\,,$$

where $\oplus$ is the addition in $\mathbb{Z}_2$. Writing $S_{(\ell_2\ell_1\ell_0)_2}$ instead of S_ℓ if $\ell = (\ell_2\ell_1\ell_0)_2$, we have

$$S_{(i_2 i_1 i_0)_2} .* S_{(j_2 j_1 j_0)_2} = S_{(k_2 k_1 k_0)_2} \quad \text{if} \quad i_\ell, j_\ell \in \{0, 1\} \text{ and } k_\ell = i_\ell \oplus j_\ell\,.$$

Example: $S_3 .* S_5 = S_{(011)_2} .* S_{(101)_2} = S_{(110)_2} = S_6$. Note that $S_3 .* S_5 \neq S_8$.

§3.8 Implementation of Multi-Controlled Rotation Gates

Example (4-qubit gate decomposition (cont.))

This allows us to write

$$S_\ell = S_1^{\ell_0} \mathbin{.*} S_2^{\ell_1} \mathbin{.*} S_4^{\ell_2} \quad \text{if} \quad 0 \leqslant \ell = (\ell_2\ell_1\ell_0)_2 \leqslant 7\,,$$

where $S_k^0 \equiv S_0$ for $k = 1, 2, 4$. By the fact that $S_i \mathbin{.*} S_i = S_0$ and $S_i \mathbin{.*} S_j = S_j \mathbin{.*} S_i$ for $S_i, S_j \in \{S_1, S_2, S_4\}$,

$$S_i \mathbin{.*} S_j = S_1^{i_0 \oplus j_0} \mathbin{.*} S_2^{i_1 \oplus j_1} \mathbin{.*} S_4^{i_2 \oplus j_2} \quad \text{if} \quad i = (i_2 i_1 i_0)_2, j = (j_2 j_1 j_0)_2\,,$$

where $\oplus$ is the addition in $\mathbb{Z}_2$. Writing $S_{(\ell_2\ell_1\ell_0)_2}$ instead of S_ℓ if $\ell = (\ell_2\ell_1\ell_0)_2$, we have

$$S_{(i_2 i_1 i_0)_2} \mathbin{.*} S_{(j_2 j_1 j_0)_2} = S_{(k_2 k_1 k_0)_2} \quad \text{if} \quad i_\ell, j_\ell \in \{0, 1\} \text{ and } k_\ell = i_\ell \oplus j_\ell\,.$$

Example: $S_3 \mathbin{.*} S_5 = S_{(011)_2} \mathbin{.*} S_{(101)_2} = S_{(110)_2} = S_6$. Note that $S_3 \mathbin{.*} S_5 \neq S_8$.

§3.8 Implementation of Multi-Controlled Rotation Gates

Example (4-qubit gate decomposition (cont.))

This allows us to write

$$S_\ell = S_1^{\ell_0} .* S_2^{\ell_1} .* S_4^{\ell_2} \quad \text{if} \quad 0 \leqslant \ell = (\ell_2\ell_1\ell_0)_2 \leqslant 7\,,$$

where $S_k^0 \equiv S_0$ for $k = 1, 2, 4$. By the fact that $S_i .* S_i = S_0$ and $S_i .* S_j = S_j .* S_i$ for $S_i, S_j \in \{S_1, S_2, S_4\}$,

$$S_i .* S_j = S_1^{i_0 \oplus j_0} .* S_2^{i_1 \oplus j_1} .* S_4^{i_2 \oplus j_2} \quad \text{if} \quad i = (i_2 i_1 i_0)_2, j = (j_2 j_1 j_0)_2\,,$$

where $\oplus$ is the addition in $\mathbb{Z}_2$. Writing $S_{(\ell_2\ell_1\ell_0)_2}$ instead of S_ℓ if $\ell = (\ell_2\ell_1\ell_0)_2$, we have

$$S_{(i_2 i_1 i_0)_2} .* S_{(j_2 j_1 j_0)_2} = S_{(k_2 k_1 k_0)_2} \quad \text{if} \quad i_\ell, j_\ell \in \{0, 1\} \text{ and } k_\ell = i_\ell \oplus j_\ell\,.$$

Example: $S_3 .* S_5 = S_{(011)_2} .* S_{(101)_2} = S_{(110)_2} = S_6$. Note that $S_3 .* S_5 \neq S_8$.

§3.8 Implementation of Multi-Controlled Rotation Gates

Example (4-qubit gate decomposition (cont.))

This allows us to write

$$S_\ell = S_1^{\ell_0} \mathbin{.*} S_2^{\ell_1} \mathbin{.*} S_4^{\ell_2} \quad \text{if} \quad 0 \leqslant \ell = (\ell_2\ell_1\ell_0)_2 \leqslant 7\,,$$

where $S_k^0 \equiv S_0$ for $k = 1, 2, 4$. By the fact that $S_i \mathbin{.*} S_i = S_0$ and $S_i \mathbin{.*} S_j = S_j \mathbin{.*} S_i$ for $S_i, S_j \in \{S_1, S_2, S_4\}$,

$$S_i \mathbin{.*} S_j = S_1^{i_0 \oplus j_0} \mathbin{.*} S_2^{i_1 \oplus j_1} \mathbin{.*} S_4^{i_2 \oplus j_2} \quad \text{if} \quad i = (i_2 i_1 i_0)_2, j = (j_2 j_1 j_0)_2\,,$$

where $\oplus$ is the addition in $\mathbb{Z}_2$. Writing $S_{(\ell_2\ell_1\ell_0)_2}$ instead of S_ℓ if $\ell = (\ell_2\ell_1\ell_0)_2$, we have

$$S_{(i_2 i_1 i_0)_2} \mathbin{.*} S_{(j_2 j_1 j_0)_2} = S_{(k_2 k_1 k_0)_2} \quad \text{if} \quad i_\ell, j_\ell \in \{0, 1\} \text{ and } k_\ell = i_\ell \oplus j_\ell\,.$$

Example: $S_3 \mathbin{.*} S_5 = S_{(011)_2} \mathbin{.*} S_{(101)_2} = S_{(110)_2} = S_6$. Note that $S_3 \mathbin{.*} S_5 \neq S_8$.

§3.8 Implementation of Multi-Controlled Rotation Gates

Example (4-qubit gate decomposition (cont.))

By identifying $S_{(\ell_2\ell_1\ell_0)_2}$ as (ℓ_2, ℓ_1, ℓ_0), we find that the group $(\{S_0, S_1, \cdots, S_7\}, .*)$ is isomorphic to the group $(\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2, \oplus)$, where $\oplus$ on $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$ is given by

$$(i_2, i_1, i_0) \oplus (j_2, j_2, j_0) \equiv (i_2 \oplus j_2, i_1 \oplus j_1, i_0 \oplus j_0), \quad i_\ell, j_\ell \in \{0, 1\};$$

that is, there exists a bijection $\varphi : \{S_0, \cdots, S_7\} \to \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$ given by $\varphi(S_{(k_2k_1k_0)_2}) = (k_2, k_1, k_0)$ such that

$$\varphi(S_{(i_2i_1i_0)_2} .* S_{(j_2j_1j_0)_2}) = \varphi(S_{(i_2i_1i_0)_2}) \oplus \varphi(S_{(j_2j_1j_0)_2}).$$

"**Definition**": A group is **a set equipped with an operation** that combines any two elements to form a third element while being associative as well as having an identity element and inverse elements.

§3.8 Implementation of Multi-Controlled Rotation Gates

Example (4-qubit gate decomposition (cont.))

By identifying $S_{(\ell_2\ell_1\ell_0)_2}$ as (ℓ_2, ℓ_1, ℓ_0), we find that the group $(\{S_0, S_1, \cdots, S_7\}, .*)$ is isomorphic to the group $(\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2, \oplus)$, where $\oplus$ on $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$ is given by

$$(i_2, i_1, i_0) \oplus (j_2, j_2, j_0) \equiv (i_2 \oplus j_2, i_1 \oplus j_1, i_0 \oplus j_0), \quad i_\ell, j_\ell \in \{0, 1\};$$

that is, there exists a bijection $\varphi : \{S_0, \cdots, S_7\} \to \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$ given by $\varphi(S_{(k_2k_1k_0)_2}) = (k_2, k_1, k_0)$ such that

$$\varphi\big(S_{(i_2i_1i_0)_2} .* S_{(j_2j_1j_0)_2}\big) = \varphi\big(S_{(i_2i_1i_0)_2}\big) \oplus \varphi\big(S_{(j_2j_1j_0)_2}\big).$$

Definition: A group is a set G equipped with an operation $*$ such that

1. $a * b \in G \;\forall\, a, b \in G$; 2. $(a * b) * c = a * (b * c) \;\forall\, a, b, c \in G$;
3. $\exists\, e \in G \ni a * e = e * a = a$ for all $a \in G$;
4. $\forall\, a \in G \;\exists b \in G \ni a * b = b * a = e$.

§3.8 Implementation of Multi-Controlled Rotation Gates

Example (4-qubit gate decomposition (cont.))

By identifying $S_{(\ell_2\ell_1\ell_0)_2}$ as (ℓ_2, ℓ_1, ℓ_0), we find that the group $(\{S_0, S_1, \cdots, S_7\}, .*)$ is isomorphic to the group $(\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2, \oplus)$, where $\oplus$ on $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$ is given by

$$(i_2, i_1, i_0) \oplus (j_2, j_2, j_0) \equiv (i_2 \oplus j_2, i_1 \oplus j_1, i_0 \oplus j_0)\,, \quad i_\ell, j_\ell \in \{0, 1\}\,;$$

that is, there exists a bijection $\varphi : \{S_0, \cdots, S_7\} \to \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$ given by $\varphi(S_{(k_2k_1k_0)_2}) = (k_2, k_1, k_0)$ such that

$$\varphi\big(S_{(i_2i_1i_0)_2} .* S_{(j_2j_1j_0)_2}\big) = \varphi\big(S_{(i_2i_1i_0)_2}\big) \oplus \varphi\big(S_{(j_2j_1j_0)_2}\big)\,.$$

Definition: A group is a set G equipped with an operation $*$ such that

1. $a * b \in G \;\forall\, a, b \in G$; 2. $(a * b) * c = a * (b * c) \;\forall\, a, b, c \in G$;
3. $\exists\, e \in G \ni a * e = e * a = a$ for all $a \in G$;
4. $\forall\, a \in G\; \exists b \in G \ni a * b = b * a = e$.

§3.8 Implementation of Multi-Controlled Rotation Gates

Example (4-qubit gate decomposition (cont.))

Recall that

1. $\boldsymbol{r}_j$ are symbols of $C_j \in \big\{\mathbf{CNOT}_{1,4}, \mathbf{CNOT}_{2,4}, \mathbf{CNOT}_{3,4}\big\}$; thus $\boldsymbol{r}_j \in \{S_1, S_2, S_4\}$ which implies that $\boldsymbol{r}_j = S_{(x_j y_j z_j)_2}$ for some $x_j, y_j, z_j \in \{0, 1\}$ with the property that one and only one of x_j, y_j, z_j is 1.
2. $\boldsymbol{b}_k = \boldsymbol{r}_k .* \cdots .* \boldsymbol{r}_8$ for all $1 \leqslant k \leqslant 8$, and we look for $\boldsymbol{r}_j$ such that $\boldsymbol{b}_1 = S_0$ and $\{\boldsymbol{b}_1, \cdots, \boldsymbol{b}_8\}$ are linearly independent (so that $\{\boldsymbol{b}_1, \cdots, \boldsymbol{b}_8\} = \{S_0, \cdots, S_7\}$).

Since

$$(x_k, y_k, z_k) \oplus \cdots \oplus (x_8, y_8, z_8) = (x_k \oplus \cdots \oplus x_8, y_k \oplus \cdots \oplus y_8, z_k \oplus \cdots \oplus z_8),$$

we find that $\varphi(\boldsymbol{b}_k)$ and $\varphi(\boldsymbol{b}_{k+1})$, the correspondence of $\boldsymbol{b}_k$ and $\boldsymbol{b}_{k+1}$ in $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$, differs by only one slot/bit.

§3.8 Implementation of Multi-Controlled Rotation Gates

Example (4-qubit gate decomposition (cont.))

Therefore, we need to arrange $S_0, \cdots, S_7$ in an order such that adjacent $\varphi(S_j)$ differs by one slot/bit. This motivates the idea of the reflected binary code (also called Gray code) which is a scheme for listing all n-bit binary numbers so that successive numbers differ in exactly one bit. The 3-qubit reflected Gray code is given by $[0, 1, 3, 2, 6, 7, 5, 4]$. We list these numbers in terms of binary representation in the following table and one can see that adjacent numbers differ by one bit.

$j = (j_2\, j_1\, j_0)_2$	0	1	3	2	6	7	5	4	0
j_2	0	0	0	0	1	1	1	1	0
j_1	0	0	1	1	1	1	0	0	0
j_0	0	1	1	0	0	1	1	0	0

§3.8 Implementation of Multi-Controlled Rotation Gates

Example (4-qubit gate decomposition (cont.))

Therefore, we need to arrange $S_0, \cdots, S_7$ in an order such that adjacent $\varphi(S_j)$ differs by one slot/bit. This motivates the idea of the reflected binary code (also called Gray code) which is a scheme for listing all n-bit binary numbers so that successive numbers differ in exactly one bit. The 3-qubit reflected Gray code is given by $[0, 1, 3, 2, 6, 7, 5, 4]$. We list these numbers in terms of binary representation in the following table and one can see that adjacent numbers differ by one bit.

$j = (j_2 j_1 j_0)_2$	0	1	3	2	6	7	5	4	0
j_2	0	0	0	0	1	1	1	1	0
j_1	0	0	1	1	1	1	0	0	0
j_0	0	1	1	0	0	1	1	0	0

§3.8 Implementation of Multi-Controlled Rotation Gates

Example (4-qubit gate decomposition (cont.))

Therefore, we need to arrange $S_0, \cdots, S_7$ in an order such that adjacent $\varphi(S_j)$ differs by one slot/bit. This motivates the idea of the reflected binary code (also called Gray code) which is a scheme for listing all *n*-bit binary numbers so that successive numbers differ in exactly one bit. The 3-qubit reflected Gray code is given by $[0, 1, 3, 2, 6, 7, 5, 4]$. We list these numbers in terms of binary representation in the following table and one can see that adjacent numbers differ by one bit.

$j = (j_2\, j_1\, j_0)_2$	0	1	3	2	6	7	5	4	0
j_2	0	0	0	0	1	1	1	1	**0**
j_1	0	0	1	1	1	1	0	0	0
j_0	0	1	1	0	0	1	1	0	0

§3.8 Implementation of Multi-Controlled Rotation Gates

Example (4-qubit gate decomposition (cont.))

From the table above, $\boldsymbol{b}_1$, $\boldsymbol{b}_2$, $\cdots$, $\boldsymbol{b}_8$ correspond to $(0,0,0)$, $(0,0,1)$, $\cdots$, $(1,0,0)$ in $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$. How do we find $\boldsymbol{r}_j$? Note that $\boldsymbol{b}_k = \boldsymbol{r}_k .* \boldsymbol{b}_{k+1}$; thus

$$\boldsymbol{r}_k = \boldsymbol{r}_k .* \boldsymbol{b}_{k+1} .* \boldsymbol{b}_{k+1} = \boldsymbol{b}_k .* \boldsymbol{b}_{k+1} \quad \forall\, 1 \leqslant j \leqslant 7\,, \ \boldsymbol{r}_8 = \boldsymbol{b}_8\,.$$

Therefore, $\boldsymbol{r}_1$ corresponds to the element $(0,0,0) \oplus (0,0,1)$ in $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$, $\boldsymbol{r}_2$ corresponds to the element $(0,0,1) \oplus (0,1,1)$ in $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$, and etc. This implies that $\boldsymbol{r}_1 = S_1$ and $\boldsymbol{r}_2 = S_2$, and so on. Note that the addition in fact indicates the bit where $\boldsymbol{b}_k$ and $\boldsymbol{b}_{k+1}$ differ (which is shown as boldface colored 0 or 1 in the table). Moreover, the position of the different bit is in fact the position of the controlled qubit in the CNOT gate (for example, the bit where $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ differs locates in the 3rd qubit; thus $\boldsymbol{r}_1 = \textbf{CNOT}_{3,4}$).

§3.8 Implementation of Multi-Controlled Rotation Gates

Example (4-qubit gate decomposition (cont.))

From the table above, $\boldsymbol{b}_1$, $\boldsymbol{b}_2$, $\cdots$, $\boldsymbol{b}_8$ correspond to $(0,0,0)$, $(0,0,1)$, $\cdots$, $(1,0,0)$ in $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$. How do we find $\boldsymbol{r}_j$? Note that $\boldsymbol{b}_k = \boldsymbol{r}_k \,.\!*\, \boldsymbol{b}_{k+1}$; thus

$$\boldsymbol{r}_k = \boldsymbol{r}_k \,.\!*\, \boldsymbol{b}_{k+1} \,.\!*\, \boldsymbol{b}_{k+1} = \boldsymbol{b}_k \,.\!*\, \boldsymbol{b}_{k+1} \quad \forall\, 1 \leqslant j \leqslant 7\,, \;\; \boldsymbol{r}_8 = \boldsymbol{b}_8\,.$$

Therefore, $\boldsymbol{r}_1$ corresponds to the element $(0,0,0) \oplus (0,0,1)$ in $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$, $\boldsymbol{r}_2$ corresponds to the element $(0,0,1) \oplus (0,1,1)$ in $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$, and etc. This implies that $\boldsymbol{r}_1 = S_1$ and $\boldsymbol{r}_2 = S_2$, and so on. Note that the addition in fact indicates the bit where $\boldsymbol{b}_k$ and $\boldsymbol{b}_{k+1}$ differ (which is shown as boldface colored 0 or 1 in the table). Moreover, the position of the different bit is in fact the position of the controlled qubit in the CNOT gate (for example, the bit where $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ differs locates in the 3rd qubit; thus $\boldsymbol{r}_1 = \textbf{CNOT}_{3,4}$).

§3.8 Implementation of Multi-Controlled Rotation Gates

Example (4-qubit gate decomposition (cont.))

From the table above, $\boldsymbol{b}_1$, $\boldsymbol{b}_2$, $\cdots$, $\boldsymbol{b}_8$ correspond to $(0,0,0)$, $(0,0,1)$, $\cdots$, $(1,0,0)$ in $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$. How do we find $\boldsymbol{r}_j$? Note that $\boldsymbol{b}_k = \boldsymbol{r}_k .* \boldsymbol{b}_{k+1}$; thus

$$\boldsymbol{r}_k = \boldsymbol{r}_k .* \boldsymbol{b}_{k+1} .* \boldsymbol{b}_{k+1} = \boldsymbol{b}_k .* \boldsymbol{b}_{k+1} \quad \forall\, 1 \leqslant j \leqslant 7\,,\ \boldsymbol{r}_8 = \boldsymbol{b}_8\,.$$

Therefore, $\boldsymbol{r}_1$ corresponds to the element $(0,0,0) \oplus (0,0,1)$ in $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$, $\boldsymbol{r}_2$ corresponds to the element $(0,0,1) \oplus (0,1,1)$ in $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$, and etc. This implies that $\boldsymbol{r}_1 = S_1$ and $\boldsymbol{r}_2 = S_2$, and so on. Note that the addition in fact indicates the bit where $\boldsymbol{b}_k$ and $\boldsymbol{b}_{k+1}$ differ (which is shown as boldface colored 0 or 1 in the table). Moreover, the position of the different bit is in fact the position of the controlled qubit in the CNOT gate (for example, the bit where $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ differs locates in the 3rd qubit; thus $\boldsymbol{r}_1 = \textbf{CNOT}_{3,4}$).

§3.8 Implementation of Multi-Controlled Rotation Gates

Example (4-qubit gate decomposition (cont.))

From the table above, $\boldsymbol{b}_1$, $\boldsymbol{b}_2$, $\cdots$, $\boldsymbol{b}_8$ correspond to $(0,0,0)$, $(0,0,1)$, $\cdots$, $(1,0,0)$ in $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$. How do we find $\boldsymbol{r}_j$? Note that $\boldsymbol{b}_k = \boldsymbol{r}_k .* \boldsymbol{b}_{k+1}$; thus

$$\boldsymbol{r}_k = \boldsymbol{r}_k .* \boldsymbol{b}_{k+1} .* \boldsymbol{b}_{k+1} = \boldsymbol{b}_k .* \boldsymbol{b}_{k+1} \quad \forall\, 1 \leqslant j \leqslant 7\,,\ \boldsymbol{r}_8 = \boldsymbol{b}_8\,.$$

Therefore, $\boldsymbol{r}_1$ corresponds to the element $(0,0,0) \oplus (0,0,1)$ in $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$, $\boldsymbol{r}_2$ corresponds to the element $(0,0,1) \oplus (0,1,1)$ in $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$, and etc. This implies that $\boldsymbol{r}_1 = S_1$ and $\boldsymbol{r}_2 = S_2$, and so on. Note that the addition in fact indicates the bit where $\boldsymbol{b}_k$ and $\boldsymbol{b}_{k+1}$ differ (which is shown as boldface colored 0 or 1 in the table). Moreover, the position of the different bit is in fact the position of the controlled qubit in the CNOT gate (for example, the bit where $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ differs locates in the 3rd qubit; thus $\boldsymbol{r}_1 = \mathbf{CNOT}_{3,4}$).

§3.8 Implementation of Multi-Controlled Rotation Gates

Example (4-qubit gate decomposition (cont.))

From the table above, $\boldsymbol{b}_1$, $\boldsymbol{b}_2$, $\cdots$, $\boldsymbol{b}_8$ correspond to $(0,0,0)$, $(0,0,1)$, $\cdots$, $(1,0,0)$ in $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$. How do we find $\boldsymbol{r}_j$? Note that $\boldsymbol{b}_k = \boldsymbol{r}_k .* \boldsymbol{b}_{k+1}$; thus

$$\boldsymbol{r}_k = \boldsymbol{r}_k .* \boldsymbol{b}_{k+1} .* \boldsymbol{b}_{k+1} = \boldsymbol{b}_k .* \boldsymbol{b}_{k+1} \quad \forall\, 1 \leqslant j \leqslant 7\,, \;\; \boldsymbol{r}_8 = \boldsymbol{b}_8\,.$$

Therefore, $\boldsymbol{r}_1$ corresponds to the element $(0,0,0) \oplus (0,0,1)$ in $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$, $\boldsymbol{r}_2$ corresponds to the element $(0,0,1) \oplus (0,1,1)$ in $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$, and etc. This implies that $\boldsymbol{r}_1 = S_1$ and $\boldsymbol{r}_2 = S_2$, and so on. Note that the addition in fact indicates the bit where $\boldsymbol{b}_k$ and $\boldsymbol{b}_{k+1}$ differ (which is shown as boldface colored 0 or 1 in the table). Moreover, the position of the different bit is in fact the position of the controlled qubit in the CNOT gate (for example, the bit where $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ differs locates in the 3rd qubit; thus $\boldsymbol{r}_1 = \textbf{CNOT}_{3,4}$).

§3.8 Implementation of Multi-Controlled Rotation Gates

Example (4-qubit gate decomposition (cont.))

Therefore, a choice of $C_1, C_2, \cdots, C_8$ can be

$$C_1 = \textbf{CNOT}_{3,4}, C_2 = \textbf{CNOT}_{2,4}, C_3 = \textbf{CNOT}_{3,4}, C_4 = \textbf{CNOT}_{1,4},$$

$$C_5 = \textbf{CNOT}_{3,4}, C_6 = \textbf{CNOT}_{2,4}, C_7 = \textbf{CNOT}_{1,4}, C_8 = \textbf{CNOT}_{1,4},$$

and recall that $R_1, R_2, \cdots, R_8$ are given by

$$R_k = \operatorname{blkdiag}\big(R_{\boldsymbol{a}}(\theta_k), \cdots, R_{\boldsymbol{a}}(\theta_k)\big).$$

Such C_k's and R_k's fulfill our goal

$$\operatorname{blkdiag}(R_{\boldsymbol{a}}(\alpha_1), R_{\boldsymbol{a}}(\alpha_2), \cdots, R_{\boldsymbol{a}}(\alpha_8)) = C_8 R_8 C_7 R_7 \cdots C_2 R_2 C_1 R_1 .$$

We summarize the discussion in the previous example and state the general procedure of the decomposition of multi-controlled $(n+1)$-qubit gate (with first n-qubit as control qubits) as follows.

§3.8 Implementation of Multi-Controlled Rotation Gates

Example (4-qubit gate decomposition (cont.))

Therefore, a choice of $C_1, C_2, \cdots, C_8$ can be

$$C_1 = \textbf{CNOT}_{3,4}, C_2 = \textbf{CNOT}_{2,4}, C_3 = \textbf{CNOT}_{3,4}, C_4 = \textbf{CNOT}_{1,4},$$

$$C_5 = \textbf{CNOT}_{3,4}, C_6 = \textbf{CNOT}_{2,4}, C_7 = \textbf{CNOT}_{1,4}, C_8 = \textbf{CNOT}_{1,4},$$

and recall that $R_1, R_2, \cdots, R_8$ are given by

$$R_k = \mathrm{blkdiag}\big(R_{\boldsymbol{a}}(\theta_k), \cdots, R_{\boldsymbol{a}}(\theta_k)\big) .$$

Such C_k's and R_k's fulfill our goal

$$\mathrm{blkdiag}(R_{\boldsymbol{a}}(\alpha_1), R_{\boldsymbol{a}}(\alpha_2), \cdots, R_{\boldsymbol{a}}(\alpha_8)) = C_8 R_8 C_7 R_7 \cdots C_2 R_2 C_1 R_1 .$$

We summarize the discussion in the previous example and state the general procedure of the decomposition of multi-controlled $(n+1)$-qubit gate (with first n-qubit as control qubits) as follows.

§3.8 Implementation of Multi-Controlled Rotation Gates

Example (4-qubit gate decomposition (cont.))

Therefore, a choice of $C_1, C_2, \cdots, C_8$ can be

$$C_1 = \textbf{CNOT}_{3,4}, C_2 = \textbf{CNOT}_{2,4}, C_3 = \textbf{CNOT}_{3,4}, C_4 = \textbf{CNOT}_{1,4},$$
$$C_5 = \textbf{CNOT}_{3,4}, C_6 = \textbf{CNOT}_{2,4}, C_7 = \textbf{CNOT}_{1,4}, C_8 = \textbf{CNOT}_{1,4},$$

and recall that $R_1, R_2, \cdots, R_8$ are given by

$$R_k = \mathrm{blkdiag}\big(R_{\boldsymbol{a}}(\theta_k), \cdots, R_{\boldsymbol{a}}(\theta_k)\big)\,.$$

Such C_k's and R_k's fulfill our goal

$$\mathrm{blkdiag}(R_{\boldsymbol{a}}(\alpha_1), R_{\boldsymbol{a}}(\alpha_2), \cdots, R_{\boldsymbol{a}}(\alpha_8)) = C_8 R_8 C_7 R_7 \cdots C_2 R_2 C_1 R_1\,.$$

We summarize the discussion in the previous example and state the general procedure of the decomposition of multi-controlled $(n+1)$-qubit gate (with first n-qubit as control qubits) as follows.

§3.8 Implementation of Multi-Controlled Rotation Gates

Let $\boldsymbol{a} = (0, a_y, a_z)$ be a unit vector in $\mathbb{R}^3$ and $N = 2^n$.

1. The goal is to write the matrix representation of a multi-controlled gate in the form

$$\mathrm{blkdiag}(R_{\boldsymbol{a}}(\alpha_1), R_{\boldsymbol{a}}(\alpha_2), \cdots, R_{\boldsymbol{a}}(\alpha_N)) = C_N R_N C_{N-1} R_{N-1} \cdots C_1 R_1,$$

where $C_k \in \{\textbf{CNOT}_{1,n+1}, \textbf{CNOT}_{2,n+1}, \cdots, \textbf{CNOT}_{n,n+1}\}$ and $R_k = \mathrm{blkdiag}\big(R_{\boldsymbol{a}}(\theta_k), \cdots, R_a(\theta_k)\big)$ for all $1 \leqslant k \leqslant N$.

2. Using the property that $C_j C_j = \mathrm{I}_N$ and $C_j C_k = C_k C_j$ for all $1 \leqslant j, k \leqslant N$, the right-hand side of the decomposition sequence above can be rewritten as

$$C_N R_N C_{N-1} R_{N-1} \cdots C_1 R_1 = \widetilde{R}_N \widetilde{R}_{N-1} \cdots \widetilde{R}_1 \cdot (C_1 C_2 \cdots C_N),$$

where $\widetilde{R}_k = (C_N C_{N-1} \cdots C_k) R_k (C_k C_{k+1} \cdots C_N)$.

§3.8 Implementation of Multi-Controlled Rotation Gates

Let $\boldsymbol{a} = (0, a_y, a_z)$ be a unit vector in $\mathbb{R}^3$ and $N = 2^n$.

1. The goal is to write the matrix representation of a multi-controlled gate in the form

$$\mathrm{blkdiag}(R_{\boldsymbol{a}}(\alpha_1), R_{\boldsymbol{a}}(\alpha_2), \cdots, R_{\boldsymbol{a}}(\alpha_N)) = C_N R_N C_{N-1} R_{N-1} \cdots C_1 R_1,$$

where $C_k \in \big\{\mathbf{CNOT}_{1,n+1}, \mathbf{CNOT}_{2,n+1}, \cdots, \mathbf{CNOT}_{n,n+1}\big\}$ and $R_k = \mathrm{blkdiag}\big(R_{\boldsymbol{a}}(\theta_k), \cdots, R_a(\theta_k)\big)$ for all $1 \leqslant k \leqslant N$.

2. Using the property that $C_j C_j = \mathrm{I}_N$ and $C_j C_k = C_k C_j$ for all $1 \leqslant j, k \leqslant N$, the right-hand side of the decomposition sequence above can be rewritten as

$$C_N R_N C_{N-1} R_{N-1} \cdots C_1 R_1 = \widetilde{R}_N \widetilde{R}_{N-1} \cdots \widetilde{R}_1 \cdot (C_1 C_2 \cdots C_N),$$

where $\widetilde{R}_k = (C_N C_{N-1} \cdots C_k) R_k (C_k C_{k+1} \cdots C_N)$.

§3.8 Implementation of Multi-Controlled Rotation Gates

3. The effect of $C_k \cdots C_N$ on R_k leads to the result

$$\widetilde{R}_k = \mathrm{blkdiag}\big(R_{\boldsymbol{a}}(b_{k1}\theta_k), R_{\boldsymbol{a}}(b_{k2}\,\theta_k), \cdots, R_{\boldsymbol{a}}(b_{kN}\,\theta_k)\big)\,,$$

where $\boldsymbol{b}_k = [b_{k1}, b_{k2}, \cdots, b_{kN}]$ is given by

$$\boldsymbol{b}_k \equiv \big[\, b_{k1}, b_{k2}, \cdots, b_{kN} \big] = \boldsymbol{r}_k \,.*\, \cdots \,.*\, \boldsymbol{r}_{N-1} \,.*\, \boldsymbol{r}_N\,, \qquad (10)$$

where $\boldsymbol{r}_j$ are symbol of C_j.

4. Let S_k denote the $(k+1)$-row of the $N \times N$ Hadamard matrix M. We choose $\boldsymbol{r}_1, \cdots, \boldsymbol{r}_N$ properly from $\big\{S_{2^k} \,\big|\, 0 \leqslant k \leqslant n-1\big\}$ so that the corresponding $\boldsymbol{b}_k$ satisfies

   (i) $\boldsymbol{b}_1 = \mathbf{ones}(1, N)$;

   (ii) the collection $\{\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_N\}$ is linearly independent.

Note that the collection $\{\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_N\}$ is linearly independent means that $\{\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_N\}$ is a permutation of rows of M.

§3.8 Implementation of Multi-Controlled Rotation Gates

3. The effect of $C_k \cdots C_N$ on R_k leads to the result

$$\widetilde{R}_k = \mathrm{blkdiag}\big(R_{\boldsymbol{a}}(b_{k1}\theta_k), R_{\boldsymbol{a}}(b_{k2}\,\theta_k), \cdots, R_{\boldsymbol{a}}(b_{kN}\,\theta_k)\big)\,,$$

where $\boldsymbol{b}_k = [b_{k1}, b_{k2}, \cdots, b_{kN}]$ is given by

$$\boldsymbol{b}_k \equiv \Big[\, b_{k1}, b_{k2}, \cdots, b_{kN} \Big] = \boldsymbol{r}_k \,.\!* \,\cdots\, .\!*\, \boldsymbol{r}_{N-1} \,.\!*\, \boldsymbol{r}_N\,, \qquad (10)$$

where $\boldsymbol{r}_j$ are symbol of C_j.

4. Let S_k denote the $(k+1)$-row of the $N \times N$ Hadamard matrix M. We choose $\boldsymbol{r}_1, \cdots, \boldsymbol{r}_N$ properly from $\big\{S_{2^k} \,\big|\, 0 \leqslant k \leqslant n-1\big\}$ so that the corresponding $\boldsymbol{b}_k$ satisfies

   (i) $\boldsymbol{b}_1 = \mathbf{ones}(1, N)$;

   (ii) the collection $\{\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_N\}$ is linearly independent.

   Note that the collection $\{\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_N\}$ is linearly independent means that $\{\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_N\}$ is a permutation of rows of M.

§3.8 Implementation of Multi-Controlled Rotation Gates

Once we have these $\boldsymbol{b}_k$'s, we then solve

$$\begin{bmatrix} b_{11} & b_{21} & \cdots & b_{N1} \\ b_{12} & b_{22} & \cdots & b_{N2} \\ \vdots & & & \vdots \\ b_{1N} & b_{2N} & \cdots & b_{NN} \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_8 \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_8 \end{bmatrix}$$

to obtain $\theta_1, \cdots, \theta_N$.

5. Let $\{x_1, x_2, \cdots, x_N\}$ be a reflected binary code (with $x_1 = 0$) for the list of numbers $\{0, 1, \cdots, N-1\}$, and $f : \{1, \cdots, N\} \to \{1, \cdots n\}$ be defined by

   $f(j)$ is the location where the bit expression of x_j and x_{j+1} differ $(x_{N+1} \equiv 0)$

   A choice of $C_1, C_2, \cdots, C_N$ and $\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_N$ are given by

   $$C_j = \textbf{CNOT}_{f(j),n+1}\,, \quad \boldsymbol{b}_j = \text{ the binary expression of } x_j.$$

§3.8 Implementation of Multi-Controlled Rotation Gates

Once we have these $\boldsymbol{b}_k$'s, we then solve

$$\begin{bmatrix} b_{11} & b_{21} & \cdots & b_{N1} \\ b_{12} & b_{22} & \cdots & b_{N2} \\ \vdots & & & \vdots \\ b_{1N} & b_{2N} & \cdots & b_{NN} \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_8 \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_8 \end{bmatrix}$$

to obtain $\theta_1, \cdots, \theta_N$.

5. Let $\{x_1, x_2, \cdots, x_N\}$ be a reflected binary code (with $x_1 = 0$) for the list of numbers $\{0, 1, \cdots, N-1\}$, and $f\colon \{1, \cdots, N\} \to \{1, \cdots n\}$ be defined by

$f(j)$ is the location where the bit expression of x_j and x_{j+1} differ $(x_{N+1} \equiv 0)$

A choice of $C_1, C_2, \cdots, C_N$ and $\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_N$ are given by

$$C_j = \textbf{CNOT}_{f(j), n+1}\,, \qquad \boldsymbol{b}_j = \text{ the binary expression of } x_j.$$

§3.8 Implementation of Multi-Controlled Rotation Gates

Remark: A way to obtain a reflected binary code for the numbers $\{0, 1, 2, \cdots, 2^n - 1\}$ is given as follows: 假設原始的值從 0 開始, 格雷碼產生的規律是:

1. 第一步, 改變最右邊的位元值;
2. 第二步，改變右邊起第一個為 1 的位元的左邊的位元;
3. 重複第一步和第二步，直到所有的格雷碼產生完畢。

Example

A Gray code for the case $n = 3$ is given by

$$000 \to 001 \to 011 \to 010 \to 110 \to 111 \to 101 \to 100 \, .$$

§3.8 Implementation of Multi-Controlled Rotation Gates

- **Algorithm of multi-control rotation gates decomposition**:

Suppose that we are given matrix

$$R = \mathrm{blkdiag}\big(R_{\boldsymbol{a}}(\alpha_1), \cdots, R_{\boldsymbol{a}}(\alpha_N)\big)$$

for some unit vector $\boldsymbol{a} = (0, a_y, a_z)$, where $N = 2^n$.

1. Let $\{x_1, x_2, \cdots, x_N\}$, where $x_1 = 0$, be a reflected binary code (Gray code) for the list of numbers $\{0, 1, \cdots, N-1\}$. Define $x_{N+1} = 0$ and $f \colon \{1, \cdots, N\} \to \{1, \cdots n\}$ by

   $f(j)$ is the location where the bit expression of x_j and x_{j+1} differ

   which can be implemented in matlab® by

   $$f(j) = \mathbf{find}(\mathbf{double}(\mathbf{xor}(\mathbf{flip}(\mathbf{de2bi}(x_j, n)), \mathbf{flip}(\mathbf{de2bi}(x_{j+1}, n)))) == 1)$$

   Set $C_j = \mathbf{CNOT}_{f(j), n+1}$ for $1 \leqslant j \leqslant N$.

§3.8 Implementation of Multi-Controlled Rotation Gates

- **Algorithm of multi-control rotation gates decomposition**:

Suppose that we are given matrix

$$R = \mathrm{blkdiag}\big(R_{\boldsymbol{a}}(\alpha_1), \cdots, R_{\boldsymbol{a}}(\alpha_N)\big)$$

for some unit vector $\boldsymbol{a} = (0, a_y, a_z)$, where $N = 2^n$.

1. Let $\{x_1, x_2, \cdots, x_N\}$, where $x_1 = 0$, be a reflected binary code (Gray code) for the list of numbers $\{0, 1, \cdots, N-1\}$. Define $x_{N+1} = 0$ and $f\colon \{1, \cdots, N\} \to \{1, \cdots n\}$ by

   $f(j)$ is the location where the bit expression of x_j and x_{j+1} differ

   which can be implemented in matlab® by

   $f(j) = \mathbf{find}(\mathbf{double}(\mathbf{xor}(\mathbf{flip}(\mathbf{de2bi}(x_j, n)), \mathbf{flip}(\mathbf{de2bi}(x_{j+1}, n)))) == 1)$

   Set $C_j = \mathbf{CNOT}_{f(j),n+1}$ for $1 \leqslant j \leqslant N$.

§3.8 Implementation of Multi-Controlled Rotation Gates

2. Define a $2^n \times 2^n$ matrix $B = [b_{ij}]$ by
$$b_{ij} = (-1)^{(i-1)\bullet x_j}\,,$$
where the exponent $(i-1)\bullet(x_j)$ is the bitwise dot product of $(i-1)$ and x_j (which can be implemented in matlab® by $\mathbf{de2bi}(i-1,n) * \mathbf{de2bi}(x_j,n)'$). Solve
$$B\begin{bmatrix}\theta_1\\ \theta_2\\ \vdots\\ \theta_N\end{bmatrix} = \begin{bmatrix}\alpha_1\\ \alpha_2\\ \vdots\\ \alpha_N\end{bmatrix}.$$

3. Define $R_k = \mathrm{blkdiag}\big(\underbrace{R_{\mathbf{a}}(\theta_k),\cdots,R_{\mathbf{a}}(\theta_k)}_{N \text{ copies of } R_{\mathbf{a}}(\theta_k)}\big)$ which can be implemented in matlab® by $R_k = \mathbf{kron}(\mathbf{eye}(N), R_{\mathbf{a}}(\theta_k))$. Then $R = C_N R_N C_{N-1} R_{N-1} \cdots C_1 R_1$.

§3.8 Implementation of Multi-Controlled Rotation Gates

2. Define a $2^n \times 2^n$ matrix $B = [b_{ij}]$ by
$$b_{ij} = (-1)^{(i-1)\bullet x_j},$$
where the exponent $(i-1)\bullet(x_j)$ is the bitwise dot product of $(i-1)$ and x_j (which can be implemented in matlab® by $\mathbf{de2bi}(i-1, n) * \mathbf{de2bi}(x_j, n)'$). Solve
$$B\begin{bmatrix}\theta_1\\ \theta_2\\ \vdots\\ \theta_N\end{bmatrix} = \begin{bmatrix}\alpha_1\\ \alpha_2\\ \vdots\\ \alpha_N\end{bmatrix}.$$

3. Define $R_k = \mathrm{blkdiag}\Big(\underbrace{R_{\boldsymbol{a}}(\theta_k), \cdots, R_{\boldsymbol{a}}(\theta_k)}_{N \text{ copies of } R_{\boldsymbol{a}}(\theta_k)}\Big)$ which can be implemented in matlab® by $R_k = \mathbf{kron}(\mathbf{eye}(N), R_{\boldsymbol{a}}(\theta_k))$. Then $R = C_N R_N C_{N-1} R_{N-1} \cdots C_1 R_1$.