Chapter 1

Tensor Product Space

The purpose of this chapter is to provide some basic background information.

- Hilbert Space
- Representation Theory
- Tensor Algebra
- Schmidt Decomposition and Singular Value Decomposition

1.1 Hilbert Space

- When quantum theory was developed in the 1920s, it had a strong influence on functional analysis. See [261] for an overview of the history.
- The quantum information theory appeared much later and the matrices are very important in this subject.
- The basic postulate of quantum mechanics is about the Hilbert space formalism.
  - ◇ To each quantum mechanical system is associated a complex Hilbert space.
  - ◇ The (pure) physical states of the system correspond to unit vectors of the Hilbert space.
- As the basic preparation, we shall describe the general notion of Hilbert space first before moving into the specifics for quantum mechanics.

Inner Product

- Given a vector space X over the complex field $\mathbb{C}$ (or other field), we say that it is equipped with an *inner product*

$$\langle \cdot, \cdot \rangle : X \times X \rightarrow \mathbb{C}$$

if the inner product satisfies the following axioms:

1. $\langle \mathbf{x}, \mathbf{y} \rangle = \overline{\langle \mathbf{y}, \mathbf{x} \rangle}$.
2. $\langle \mathbf{x} + \mathbf{y}, \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle$.
3. $\langle \alpha \mathbf{x}, \mathbf{y} \rangle = \overline{\alpha} \langle \mathbf{x}, \mathbf{y} \rangle$. (Conjugate linear in $\mathbf{x}$)[1]
4. $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$ and $\langle \mathbf{x}, \mathbf{x} \rangle = 0$ if and only if $\mathbf{x} = \mathbf{0}$.

◇ The quantity

$$\|\mathbf{x}\| := \langle \mathbf{x}, \mathbf{x} \rangle^{\frac{1}{2}} \tag{1.1}$$

naturally defines a vector norm on X.

◇ Two vectors $\mathbf{x}$ and $\mathbf{y}$ are said to be *orthogonal*, denoted by $\mathbf{x} \perp \mathbf{y}$, if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$.

[1] Such an arrangement is somewhat unconventional, but is easier for transition to the notion of density matrix in quantum mechanics. Note that the inner product is linear in $\mathbf{y}$.

- Some basic facts:
  - ◇ (Cauchy-Schwarz Inequality) For all $\mathbf{x}, \mathbf{y}$ in an inner product space,
$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \|\mathbf{y}\|. \tag{1.2}$$
Equality holds if and only if $\mathbf{x} = \alpha \mathbf{y}$ or $\mathbf{y} = \mathbf{0}$.
  - ◇ (Parallelogram Law) With the induced inner product norm,
$$\|\mathbf{x} + \mathbf{y}\|^2 + \|\mathbf{x} - \mathbf{y}\|^2 = 2\|\mathbf{x}\|^2 + 2\|\mathbf{y}\|^2. \tag{1.3}$$
  - ◇ (Continuity) Suppose that $\mathbf{x}_n \to \mathbf{x}$ and $\mathbf{y}_n \to \mathbf{y}$. Then $\langle \mathbf{x}_n, \mathbf{y}_n \rangle \to \langle \mathbf{x}, \mathbf{y} \rangle$.
  - ◇ Note that these geometric properties follow from the algebraic definition of inner product.
- A complete inner product space is called a *Hilbert space*.
  - ◇ A metric space X is *complete* if every Cauchy sequence in X converges to a point in X.
  - ◇ The set Q of rational numbers is a vector space over itself. The sequence $x_{n+1} = \frac{x_n}{2} + \frac{1}{x_n}$ is Cauchy in Q. However, it does not converge towards any rational limit. Why and what does this mean?

Orthogonal Complements

- Given any subset S in an inner product space X, the set
$$S^{\perp} := \{\mathbf{y} \in X | \mathbf{y} \perp \mathbf{x} \text{ for every } \mathbf{x} \in S\} \tag{1.4}$$
is called the *orthogonal complement* of S.
  - ◇ $S^{\perp}$ is a closed subspace.
  - ◇ $S \subset (S^{\perp})^{\perp}$.
  - ◇ $(S^{\perp})^{\perp}$ is the smallest closed subspace containing S.
- We say that X is the *direct sum* of two subspaces M and N, denoted by $M = M \bigoplus N$, if every $\mathbf{x} \in X$ has a unique representation of the form $\mathbf{x} = \mathbf{m} + \mathbf{n}$ where $\mathbf{m} \in M$ and $\mathbf{n} \in N$.
  - ◇ The Classical Projection Theorem: Let M be a closed subspace of a Hilbert space $\mathscr{H}$. Corresponding to every $\mathbf{x} \in \mathscr{H}$,
    - ▷ There exists a unique $\mathbf{m}_0 \in M$ such that $\|\mathbf{x} - \mathbf{m}_0\| \leq \|\mathbf{x} - \mathbf{m}\|$ for all $\mathbf{m} \in M$.
    - ▷ A necessary and sufficient condition that $\mathbf{m}_0$ be the unique best approximation of $\mathbf{x}$ in M is that $\mathbf{x} - \mathbf{m}_0 \perp M$.
  - ◇ If M is a closed linear subspace of a Hilbert space $\mathscr{H}$, then $\mathscr{H} = M \bigoplus M^{\perp}$.
- Given a vector $\mathbf{x}_0 \in \mathscr{H}$ and a closed subspace M, the unique vector $\mathbf{m}_0 \in M$ such that $\mathbf{x}_0 - \mathbf{m}_0 \in M^{\perp}$ is called the *orthogonal projection* of $\mathbf{x}_0$ onto M.

Gram-Schmidt Orthogalization Process

- Any given sequence (finite or infinite) $\{\mathbf{x}_n\}$ of linearly independent vectors in an inner product space X can be orthogonalized in the following sense.
  - ◇ There exists an orthonormal sequence $\{\mathbf{z}_n\}$ in X such that for each finite positive integer ℓ,
  $$\operatorname{span}\{\mathbf{x}_1, \ldots, \mathbf{x}_\ell\} = \operatorname{span}\{\mathbf{z}_1, \ldots, \mathbf{z}_\ell\}. \tag{1.5}$$
  - ◇ Indeed, the sequence $\{\mathbf{z}_n\}$ can be generated via the Gram-Schmidt process.
  $$\begin{aligned} \mathbf{z}_1 &:= \frac{\mathbf{x}_1}{\|\mathbf{x}_1\|}, \\ \mathbf{w}_n &:= \mathbf{x}_n - \sum_{i=1}^{n-1} \langle \mathbf{z}_i, \mathbf{x}_n \rangle \mathbf{z}_i, \quad n > 1, \\ \mathbf{z}_n &:= \frac{\mathbf{w}_n}{\|\mathbf{w}_n\|}, \quad n > 1. \end{aligned}$$
- Additional insights:
  - ◇ What if the sequence cannot be enumerated?
  - ◇ Show that in finite dimensional space, the Gram-Schmidt process is equivalent to that any full column rank matrix $A \in \mathbb{R}^{n\times m}$ can be decomposed as
  $$A = QR \tag{1.6}$$
  where $Q \in \mathbb{R}^{n\times m}$, $Q^T Q = I_m$, and $R \in \mathbb{R}^{m\times m}$, R is upper triangular.

Orthogonal Polynomials

- One can apply the Gram-Schmidt procedure with respect to a variety of inner products in the space $C^1[a,b]$ on the sequence of polynomials $\{1, x^1, x^2, \ldots\}$ to produce families of orthogonal polynomials.
- Six popular orthogonal polynomial families:
  - ◇ Gegenbauer: $[-1,1]$, $\omega(x) = (1-x^2)^{\lambda-\frac{1}{2}}$, λ is any number $\geq -\frac{1}{2}$.
  - ◇ Hermite: $(-\infty,\infty)$, $\omega(x) = e^{-x^2}$.
  - ◇ Laguerre: $[0,\infty)$, $\omega(x) = e^{-x}x^{\alpha}$, α is number ≥ -1.
  - ◇ Legendre: Needed in Gaussian quadrature.
  - ◇ Jacobi: $[-1,1]$, $\omega(x) = (1-x)^a(1+x)^b$, a any irrational or ration ≥ -1.
  - ◇ Chebyshev: $(-1,1)$, $\omega(x) = (1-x^2)^{\pm\frac{1}{2}}$.
- Show that the resulting orthonormal polynomials $\{z_n\}$ satisfy a recursive relationship of the form

$$z_n(x) = (a_n x + b_n) z_{n-1}(x) - c_n z_{n-2}(x), \quad n = 2, 3, \ldots \quad (1.7)$$

  where the coefficients a_n, b_n, c_n can be explicitly determined.
- Show that the zeros of the orthonormal polynomials are real, simple, and located in the interior of $[a,b]$.

1.2 Representation Theory

- Depending on the applications, a Hilbert space can have two kinds of representations.
  - ⋄ Sometimes, due to issues such as feasibility, it is necessary to approximate a desired target vector.
    - ▷ This leads to the Fourier series.
    - ▷ In a Hilbert space, the most fundamental notion of approximation is the projection.
  - ⋄ Any bounded linear functional $f : \mathscr{H} \to \mathbb{F}$, where $\mathbb{F}$ is the field over which the vector space is defined, can uniquely be represented by an element in $\mathscr{H}$.
    - ▷ This is the so called Riesz representation theorem.
    - ▷ One can thus use the bounder linear functional to represent a Hilbert space $\mathscr{H}$ and vise versa.

Approximation

One of the most fundamental optimization question is as follows:

- Let $\mathbf{x}_0$ represent a target vector in a Hilbert space $\mathscr{H}$.
  - ◇ The target vector could mean a true solution that is hard to get in the abstract space $\mathscr{H}$.
- Let M denote a subspace of $\mathscr{H}$.
  - ◇ Consider M as the set of all computable, constructible, or reachable (by some reasonable means) vectors in $\mathscr{H}$.
  - ◇ M shall be called a feasible set.
- Want to solve the optimization problem

$$\min_{\mathbf{m}\in M} \|\mathbf{x}_0 - \mathbf{m}\|. \tag{1.8}$$

Projection Theorem

- A necessary and sufficient condition that $\mathbf{m}_0 \in M$ solves the above optimization problem is that

$$\mathbf{x}_0 - \mathbf{m}_0 \perp M. \tag{1.9}$$

  - ◇ The minimizer $\mathbf{m}_0$ is unique, if it exists.
  - ◇ The existence is guaranteed only if M is closed.

- How to find this minimizer?
  - ◇ Assume that M is of finite dimension and has a basis

$$M = \mathrm{span}\{\mathbf{y}_1, \ldots, \mathbf{y}_n\}. \tag{1.10}$$

Write

$$\mathbf{m}_0 = \sum_{i=1}^{n} \alpha_i \mathbf{y}_i. \tag{1.11}$$

Then $\alpha_1, \ldots, \alpha_n$ satisfy the *normal equation*

$$\langle \mathbf{y}_j, \mathbf{x}_0 - \sum_{i=1}^{n} \alpha_i \mathbf{y}_i \rangle = 0, \quad j = 1, \ldots n. \tag{1.12}$$

  - ◇ In matrix form, the linear system can be written as

$$\begin{pmatrix} \langle \mathbf{y}_1, \mathbf{y}_1 \rangle & \langle \mathbf{y}_1, \mathbf{y}_2 \rangle & \ldots & \langle \mathbf{y}_1, \mathbf{y}_n \rangle \\ \langle \mathbf{y}_2, \mathbf{y}_1 \rangle & & & \\ \vdots & & \ddots & \vdots \\ \langle \mathbf{y}_n, \mathbf{y}_1 \rangle & & & \langle \mathbf{y}_n, \mathbf{y}_n \rangle \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix} = \begin{pmatrix} \langle \mathbf{y}_1, \mathbf{x}_0 \rangle \\ \langle \mathbf{y}_2, \mathbf{x}_0 \rangle \\ \vdots \\ \langle \mathbf{y}_n, \mathbf{x}_0 \rangle \end{pmatrix}. \tag{1.13}$$

    - ▷ If the basis $\{\mathbf{y}_1, \ldots, \mathbf{y}_n\}$ are orthonormal, then trivially

$$\alpha_i = \langle \mathbf{y}_i, \mathbf{x}_0 \rangle. \tag{1.14}$$

- ◇ What can be done if M is not of finite dimension?
- ◇ If M is of codimension n, i.e., if the orthogonal complement of M has dimension n, then a dual approximation problem can be formulated.

Schauder Basis versus Hamel Basis

- Conventionally, it is known that every finitely generated vector space has a basis. But what about vector spaces that are not finitely generated, such as $C[0,1]$?
- A vector space V is said to be spanned by an infinite set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots\}$, if each vector $\mathbf{v} \in V$ is a *finite linear combination* $a_{i_1}\mathbf{v}_{i_1} + \ldots + a_{i_n}\mathbf{v}_{i_n}$ of the $\mathbf{v}_i$'s.
- A set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots\}$ is a basis for V if and only if every element of V can be be written *in a unique way* as a finite linear combination of elements from the set.
- Actually, the notation $\{\mathbf{v}_1, \mathbf{v}_2, \ldots\}$ for an infinite set is also misleading because it seems to indicate that the set is *countable*.
- Want to allow the possibility that a vector space may have an uncountable basis.
  - ⋄ It is more sensible to use notation such as $\{\mathbf{v}_\lambda | \lambda \in \Lambda\}$, where Λ is some unspecified indexing set.
  - ⋄ What is the basis of $\mathbb{R}^\infty$? What is the basis of $\ell^2 \subset \mathbb{R}^\infty$?
- Every vector space has a basis.
  - ⋄ Need to employ Zorn's Lemma, i.e., any nonempty partially ordered set in which every chain has an upper bound contains at least one maximal element.
  - ⋄ Such a basis is called a *Hamel basis*.

Fourier Series

- Over a Hilbert space, there is a confusing but more useful notion of"spanning the space", called the *Schauder basis*.
  - ◇ A maximal orthonormal subset S of a Hilbert space $\mathscr{H}$ is called a *complete orthonormal basis*.
    - ▷ S may not be countable, but its existence is guarantee by Zorn's Lemma.
- If S is any orthonormal subset (not necessarily maximal) of $\mathscr{H}$ and $\mathbf{x} \in \mathscr{H}$, then
  - ◇ $\langle \mathbf{z}, \mathbf{x} \rangle$ is nonzero for at most a countable number of $\mathbf{z} \in S$.
  - ◇ (Bessel's Inequality)

$$\sum_{\mathbf{z} \in S} |\langle \mathbf{z}, \mathbf{x} \rangle|^2 \leq \|\mathbf{x}\|^2. \tag{1.15}$$

- The convergent series

$$\mathcal{F}(\mathbf{x}) := \sum_{\mathbf{z} \in S} \langle \mathbf{z}, \mathbf{x} \rangle \mathbf{z} \tag{1.16}$$

guaranteed by Bessle's inequality is called the *Fourier series* of $\mathbf{x}$ in $\mathscr{H}$.

  - ◇ $\mathcal{F}(\mathbf{x})$ belongs to the closure of span$\{S\}$.
  - ◇ $\mathbf{x} - \mathcal{F}(\mathbf{x}) \perp$ span$\{S\}$ (and hence its closure).
  - ◇ When will an orthonormal subset S generate a Hilbert space?

- If S is a maximal orthonormal subset of $\mathscr{H}$, then any $\mathbf{x} \in \mathscr{H}$ can be identified via a square-summable sum

$$\mathbf{x} = \sum_{\mathbf{z} \in S} \langle \mathbf{z}, \mathbf{x} \rangle \mathbf{z}. \tag{1.17}$$

  ◇ The Fourier coefficients uniquely characterize $\mathbf{x}$.

- A Hilbert space $\mathscr{H}$ has a countable orthonormal basis if and only if $\mathscr{H}$ has a countable dense subset if and only if every orthonormal basis for H is countable.

  ◇ Such a space is said to be *separable*.

- Examples:

  ◇ The subspace of polynomials is dense in $\mathcal{L}_2[-1, 1]$.

  ◇ The Legendre polynomials, i.e., the orthonormal sequence

$$\mathbf{z}_n(x) = \sqrt{\frac{2n+1}{2}} \frac{(-1)^n}{2^n n!} \frac{d^n}{dx^n} (1 - x^2)^n \tag{1.18}$$

  is complete in $\mathcal{L}_2[-1, 1]$.

  ◇ The trigonometric system

$$\{1, e^{ix}, e^{-ix}, e^{2ix}, e^{-2ix}, e^{3ix}, e^{-3ix}, \ldots\}$$

  is a Schauder basis for the space $\mathcal{L}_p([0, 2])$ for any $1 < p < \infty$, but is not a Schauder basis for $\mathcal{L}_1([0, 2])$.

- So, what is the fundamental difference a Hamel basis and a Schauder basis?

1.3 Tensor Algebra

- Before any physical mechanism is applied, tensor product is a mathematical way to describe the interaction of every elements in one vector to every elements in another vector.
  - ◇ Such a simultaneous interaction is the basic notion of a bipartite quantum system whose subsystems are entangled.
- The interaction is in the simplest form of a bilinear map, but can be interpreted in many different ways.
- You could plug in numbers to see what these bilinear map does if the "actions" are multiplications, but the theory applies to any kind of bilinear maps. Do abstract thinking, instead.
- Do not confuse the tensor product with the Kronecker product.
  - ◇ Although they carry the same amount of information, they appear in different forms.

Tensor Product

- Given two Hilbert spaces $\mathscr{H}_1$ and $\mathscr{H}_2$ with complete orthonormal basis states $\{\mathbf{e}_i\}$ and $\{\mathbf{f}_j\}$, respectively, the tensor product space $\mathscr{H}_1 \otimes \mathscr{H}_2$ is defined by

$$\mathscr{H}_1 \otimes \mathscr{H}_2 := \{\sum_{i,j} \mathbf{u}_i \otimes \mathbf{v}_j | \mathbf{u}_i \in \mathscr{H}_1, \mathbf{v}_j \in \mathscr{H}_2\}. \tag{1.19}$$

  - ◇ The tensor product $\otimes$ represents only an *bilinear* relationship applied to the Cartesian product $\mathscr{H}_1 \times \mathscr{H}_2$.
    - ▷ No other strings, physically or mathematically, are assumed.
  - ◇ The summation is formal over any index subset with finite support[2].
- An algebra can be defined on $\mathscr{H}_1 \otimes \mathscr{H}_2$ to make it a vector space.
  - ◇ What is an algebra?
- An inner product can also be defined via the relationship

$$\langle \mathbf{x} \otimes \mathbf{y}, \mathbf{z} \otimes \mathbf{w} \rangle := \langle \mathbf{x}, \mathbf{z} \rangle \langle \mathbf{y}, \mathbf{w} \rangle. \tag{1.20}$$

  - ◇ It might be needed to carry out a topological completion[3] to make $\mathscr{H}_1 \otimes \mathscr{H}_2$ a Hilbert space with $\{\mathbf{e}_i \otimes \mathbf{f}_j\}$ as the basis.
- What is the geometry of a unit circle in $\mathscr{H}_1 \otimes \mathscr{H}_2$?

[2]For practical reason, we are assuming the Hilbert spaces are separable, but the concept can be generalized.
[3]Not needed for finite dimensional Hilbert spaces.

Outer Product

- Suppose that an element in a Hilbert space is represented by a *column* vector whose entries are the coordinates with respect to a given basis. Then we can think of the tensor product $\mathbf{a} \otimes \mathbf{b}$ as the *outer product* of column vectors $\mathbf{a}$ and $\mathbf{b}$, that is[4], $\mathbf{a} \otimes \mathbf{b} = \mathbf{a}\mathbf{b}^\top$ which is a matrix of size $m \times n$.
  - ◇ The notion of outer product can readily be generalized to higher-order tensors by simply adding the list of indices.
- Conventionally, entries of a matrix are enumerated by stacking columns together as a 1-D array. Thus, it is suggested to enumerate the composite basis in the order as

$$\mathbf{e}_1 \otimes \mathbf{f}_1, \mathbf{e}_2 \otimes \mathbf{f}_1, \ldots, \mathbf{e}_1 \otimes \mathbf{f}_2, \mathbf{e}_2 \otimes \mathbf{f}_2, \ldots, \mathbf{e}_1 \otimes \mathbf{f}_3, \ldots. \tag{1.21}$$

  - ◇ This ordering is consistent with the vectorization of the outer product.
- Any element $\boldsymbol{\psi} \in \mathscr{H}_1 \otimes \mathscr{H}_2$ can be written as

$$\boldsymbol{\psi} = \sum_{i,j} c_{ji} \mathbf{e}_i \otimes \mathbf{f}_j. \tag{1.22}$$

  That is, any element $\boldsymbol{\psi}$ in a bipartite system can be represented by a matrix $C = [c_{ji}]$.
  - ◇ $\|\boldsymbol{\psi}\| := \sum_{i,j} |c_{ij}|^2$.
  - ◇ We can also write $\boldsymbol{\psi} = \mathbf{vec}(C)$.

[4]It is confusing, but sometimes the outer product over complex space is defined to be $\mathbf{a} \otimes \mathbf{b} = \mathbf{a}\overline{\mathbf{b}}^\top$. By doing so, we lose the bilinearity of $\otimes$. To avoid the confusion, we shall use Dirac's bra-ket notation $|\mathbf{x}\rangle\langle\mathbf{y}|$ to denote the conjugated outer product.

Linear Operators

- Any linear operator

$$\mathcal{T} : \mathscr{H}_1 \otimes \mathscr{H}_2 \rightarrow \widehat{\mathscr{H}_1} \otimes \widehat{\mathscr{H}_2}$$

is necessarily an order-4 tensor.

  - ◇ First, its action on the basis element $\mathbf{e}_i \otimes \mathbf{f}_j$ is given by

$$\mathcal{T}(\mathbf{e}_i \otimes \mathbf{f}_j) = \sum_{s,t} \tau_{stij} \widehat{\mathbf{e}}_s \otimes \widehat{\mathbf{f}}_t. \tag{1.23}$$

  - ◇ Employ the multi-index notation $\mathcal{I} = (s,t)$ and $\mathcal{J} = (i,j)$. Identify $\widehat{E}_{\mathcal{I}} := \widehat{\mathbf{e}}_s \otimes \widehat{\mathbf{f}}_t$ and $E_{\mathcal{J}} := \mathbf{e}_i \otimes \mathbf{f}_j$. Then (1.23) can be expressed as

$$\mathcal{T}(E_{\mathcal{J}}) = \sum_{\mathcal{I}} \tau_{\mathcal{I}\mathcal{J}} \widehat{E}_{\mathcal{I}}, \tag{1.24}$$

    which is precisely the conventional matrix representation of a linear transformation.

- If $C = [c_{ij}] = \sum_{\mathcal{J}} c_{\mathcal{J}} E_{\mathcal{J}}$, then

$$\mathcal{T}(C) = \sum_{\mathcal{J}} c_{\mathcal{J}} \mathcal{T}(E_{\mathcal{J}}) = \sum_{\mathcal{I}} (\sum_{\mathcal{J}} \tau_{\mathcal{I}\mathcal{J}} c_{\mathcal{J}}) \widehat{E}_{\mathcal{I}}. \tag{1.25}$$

  - ◇ This operation defines a *tensor-to-matrix multiplication* which generalizes the conventional matrix-to-vector multiplication.

Tensor Product of Operators

- Given two linear operators $\mathcal{A}$ and $\mathcal{B}$ on $\mathscr{H}_1$ and $\mathscr{H}_2$, respectively, want a special linear operator $\mathcal{A} \otimes \mathcal{B}$ on $\mathscr{H}_1 \otimes \mathscr{H}_2$ such that

$$(\mathcal{A} \otimes \mathcal{B})(\mathbf{e}_i \otimes \mathbf{f}_j) := (\mathcal{A}\mathbf{e}_i) \otimes (\mathcal{B}\mathbf{f}_j), \tag{1.26}$$

  for all basis states $\{\mathbf{e}_i\}$ and $\{\mathbf{f}_j\}$.

- If $\mathcal{A}$ and $\mathcal{B}$ are represented by matrices A and B, respectively, then the matrix representation of $\mathcal{A} \otimes \mathcal{B}$ is precisely the Kronecker product $A \otimes B$ of A and B.
  - ◇ Write $\mathcal{A}\mathbf{e}_i = \sum_s a_{si}\widehat{\mathbf{e}}_s$ and $\mathcal{B}\mathbf{f}_j = \sum_t b_{tj}\widehat{\mathbf{f}}_t$. Then, by the condition (1.26), we have

$$(\mathcal{A}\mathbf{e}_i) \otimes (\mathcal{B}\mathbf{f}_j) = \sum_s \sum_t a_{si} b_{tj} (\widehat{\mathbf{e}}_s \otimes \widehat{\mathbf{f}}_t). \tag{1.27}$$

  - ◇ With respect to the outer product ordering (1.21), the matrix representation of the order-4 tensor with $\tau_{stij} := [a_{si}b_{tj}]$ is a block matrix whose (i, j) block is made of

$$\begin{bmatrix} a_{1i}b_{1j} & a_{1i}b_{2j} & a_{1i}b_{3j} \dots \\ a_{2i}b_{1j} & a_{2i}b_{2j} & a_{2i}b_{3j} \dots \\ a_{3i}b_{1j} & a_{3i}b_{2j} & a_{3i}b_{3j} \dots \\ \vdots & & \vdots \end{bmatrix}$$

    - ▷ If A and B are expressed in columns, then the block is precisely $\mathbf{a}_i \otimes \mathbf{b}_j$.
    - ▷ Block size = (row# of A) × (row# of B).
    - ▷ There are (column# of A) × (column# of B) blocks.

Tensor Product versus Kronecker Product

- The same notation $\otimes$ has been used for different meanings:
  - ◇ Physicists use it to denote the tensor product. That is,

$$\mathbf{a} \otimes \mathbf{b} := \mathbf{a}\mathbf{b}^\top, \tag{1.28}$$

    which is a matrix. (Without the complex conjugation!)
  - ◇ Mathematicians use it to denote the Kronecker product. That is,

$$\mathbf{a} \otimes \mathbf{b} := [a_1 b_1, a_1 b_2, \ldots, a_2 b_1, a_2 b_2, \ldots]^\top, \tag{1.29}$$

    which a long vector.
- What if in higher-dimensional spaces?
  - ◇ The tensor product can be generalized naturally to higher order tensor products by simply adding indices to the right.
  - ◇ The generalization of the Kronecker product will involve nasty multiplications.
- Mathematicians sometime prefer to keep the Kronecker notation and denote tensor/outer product by $\circ$. See [399].
  - ◇ Thus

$$\begin{aligned} \mathbf{a} \circ \mathbf{b} &:= \mathbf{a}\mathbf{b}^\top, \\ \mathbf{vec}(\mathbf{a} \circ \mathbf{b}) &= \mathbf{b} \otimes \mathbf{a}, \\ \mathbf{a} \otimes \mathbf{b} &= \mathbf{vec}(\mathbf{b}\mathbf{a}^\top). \end{aligned} \tag{1.30}$$

- In quantum mechanics, there is this complex outer product:

$$|\mathbf{a}\rangle \langle \mathbf{b}| := \mathbf{a}\mathbf{b}^*. \tag{1.31}$$

Lexicographic Ordering

- Sometimes it is easier to systematically enumerate the basis in the *lexicographic ordering*:

$$\mathbf{e}_1 \otimes \mathbf{f}_1, \mathbf{e}_1 \otimes \mathbf{f}_2, \ldots, \mathbf{e}_2 \otimes \mathbf{f}_1, \mathbf{e}_2 \otimes \mathbf{f}_2, \ldots. \tag{1.32}$$

  If an element $\boldsymbol{\psi} \in \mathscr{H}_1 \otimes \mathscr{H}_2$ is expressed in this ordering

$$\boldsymbol{\psi} = \sum_{i,j} d_{ij} \mathbf{e}_i \otimes \mathbf{f}_j, \tag{1.33}$$

  then $D = C^\top$. (Only a flip of indices. No practical meaning.)

- If the bases are ordered in this way, then the matrix representation of the order-4 tensor with $\tau_{stij} := [a_{si} b_{tj}]$ is precisely

$$\begin{bmatrix} a_{11}b_{11} & a_{11}b_{12} & \ldots & a_{12}b_{11} & \ldots \\ a_{11}b_{21} & a_{11}b_{22} & \ldots & a_{12}b_{21} & \\ \vdots & & & \vdots & \\ a_{21}b_{11} & a_{21}b_{12} & \ldots & & \\ a_{21}b_{21} & a_{21}b_{22} & \ldots & & \\ \vdots & & & & \end{bmatrix} = \begin{bmatrix} a_{11}B & a_{12}B & \ldots \\ a_{21}B & a_{22}B & \ldots \\ \vdots & & \end{bmatrix}$$

  which can be effectively written as the Kronecker product $A \otimes B$.

  - ◇ Check this out as an important exercise!

- We shall take advantage of the lexicographical ordering.

$\mathfrak{B}(\mathscr{H})$ and $\mathrm{Tr}(\mathcal{A})$

- The set of bounded linear operators $\mathscr{H} \to \mathscr{H}$ is denoted by $\mathfrak{B}(\mathscr{H})$.
  - ◇ For quantum computation, we are particularly interested in some special subsets of $\mathfrak{B}(\mathscr{H})$:
    - ▷ Density matrices for representing mixed states.
    - ▷ Unitary transformation for representing *reversible* computation.
- Given $\mathcal{A} \in \mathfrak{B}(\mathscr{H})$ and an orthonormal Schauder basis $\{\mathbf{e}_i\}$, define the *trace* of $\mathcal{A}$ by

$$\mathrm{Tr}(\mathcal{A}) := \sum_i < \mathbf{e}_i, \mathcal{A}\mathbf{e}_i > . \tag{1.34}$$

  - ◇ The summation could be divergent.
  - ◇ If $\mathscr{H}$ is finite-dimensional, then $\mathrm{Tr}(\mathcal{A})$ coincides with the definition of the trace of the matrix representation of A.

k-fold Tensor Product

- Given a Hilbert space $\mathscr{H}$, the k-fold tensor product is denoted by
$$\mathscr{H}^{\otimes k} := \mathscr{H} \otimes \ldots \otimes \mathscr{H}. \tag{1.35}$$
- Similarly, if $\mathcal{A} \in \mathfrak{B}(\mathscr{H})$, then
$$\mathcal{A}^{\otimes k} := \mathcal{A} \otimes \ldots \otimes \mathcal{A} \in \mathfrak{B}(\mathscr{H}^{\otimes k}). \tag{1.36}$$

Adjoint Operator

- If $\mathcal{A} : \mathscr{H} \to \mathscr{K}$ is a linear operator, then there exists a unique operator, denoted by $\mathcal{A}^*$, such that the *adjoint equation*

$$\langle \mathcal{A}\mathbf{h}, \mathbf{k} \rangle_{\mathscr{K}} = \langle \mathbf{h}, \mathcal{A}^*\mathbf{k} \rangle_{\mathscr{H}} \tag{1.37}$$

  is satisfied for all $\mathbf{h} \in \mathscr{H}$ and $\mathbf{k} \in \mathscr{K}$.

  - ◇ The notion that $\mathcal{A}^*$ is the conjugate transpose of $\mathcal{A}$ is not always true.
  - ◇ How to find $\mathcal{A}^*$ in general?
  - ◇ See [287] for an interesting discussion on the adjoint for non-linear operators.

- An operator $\mathcal{A} \in \mathfrak{B}(\mathscr{H})$ is said to be *positive* (or positive semi-definite) if and only if $\langle \mathbf{h}, \mathcal{A}\mathbf{h} \rangle \geq 0$ for all $\mathbf{h} \in \mathscr{H}$.

  - ◇ A positive operator $\mathcal{A}$ over the complex field is automatically self-adjoint because

$$\langle \mathbf{h}, \mathcal{A}\mathbf{h} \rangle = \overline{\langle \mathbf{h}, \mathcal{A}\mathbf{h} \rangle} = \langle \mathcal{A}\mathbf{h}, \mathbf{h} \rangle = \langle \mathbf{h}, \mathcal{A}^*\mathbf{h} \rangle$$

  - ◇ In the real-valued case, it is required to add the condition of "symmetry" specifically.
  - ◇ Will use the notation $\mathcal{A} \geq 0$ to denote a positive operator.

Dirac Notation

- Will begin to interchange the ket notation $|\mathbf{x}\rangle$ with the column vector $\mathbf{x}$ in a Hilbert space $\mathscr{H}$.
- The bra notation $\langle\mathbf{x}|$ can be thought of as the row vector $\mathbf{x}^*$, but more precisely it should be considered as a linear functional acting on $\mathscr{H}$.
  - ◇ By the Riesz representation theorem, a bounded linear functional $f : \mathscr{H} \to \mathbb{C}$ acting on an element $\mathbf{h}$ can be denoted by

    $$f(\mathbf{h}) = \langle\mathbf{z}|\mathbf{h}\rangle \tag{1.38}$$

    for some $\mathbf{z} \in \mathscr{H}$.
  - ◇ It is common to suppress the vector or functional from the bra-ket notation and only use a label inside the typography for the bra or ket.

Finiteness versus Infiniteness

- In quantum mechanics, a quantum state is typically represented as an element of a complex Hilbert space, for example, the infinite-dimensional vector space of all possible wave functions.
- For quantum computation, we shall consider only finite dimensional spaces over complex field.
  - ◇ With respect to a given basis, such a space is isomorphic to $\mathbb{C}^n$.
  - ◇ Current technology even limits us to very low dimensions.

1.4 Schmidt Decomposition and Singular Value Decomposition

- The theory is applicable to bipartite system only.
- Both decompositions are equivalent in finite dimensional systems.
- Generalization to infinite dimensional space is not always possible.
  - ◇ Unlike in finite dimensions, it is possible that a bounded self-adjoint operator on a complex Hilbert space may have no eigenvalues.
- The proofs are fundamentally different.

Singular Value Decomposition

- The singular value decomposition (SVD) is a matrix factorization that serves both as the first computational step in many numerical algorithms and as the first conceptual step in many theoretical studies.
- Given $A \in \mathbb{R}^{m \times n}$ $(m \geq n)$, the image of the unit sphere in $\mathbb{R}^n$ under A is a hpyerellipse (of dimension n) in $\mathbb{R}^m$.
  - ◇ The unit vectors $\mathbf{u}_i \in \mathbb{R}^m$ in the directions of the principal semi-axes of the hyperellipse are called the *left singular vectors* of A.
  - ◇ The unit vectors $\mathbf{v}_i \in \mathbb{R}^n$ in the directions of the preimage of the principal semi-axes of the hyperellipse are called the *right singular vectors* of A.
  - ◇ The lengths σ_i of the principal semi-axes of the hyperellipse are called the *singular values* of A.
- The above geometric relationships are mathematically equivalent to:

$$A\mathbf{v}_i = \sigma_i \mathbf{u}_i. \tag{1.39}$$

  - ◇ It is clear that all $\mathbf{u}_i$, $i = 1, \ldots, m$, are mutually orthogonal.
  - ◇ It will be shown that all $\mathbf{v}_i$, $i = 1, \ldots n$, are also mutually orthogonal.
- The concept holds in the complex spaces, but the geometry is not obvious.
  - ◇ How does the unit circle in $\mathbb{C}^2$ look like?

- Recall that $A^*A \in \mathbb{C}^{n\times n}$ is hermitian and positive semi-definite.
  - ◇ A^*A has a complete set of orthonormal eigenvectors.
  - ◇ All eigenvalues of A^*A are nonnegative.
  - ◇ Denote positive eigenvalues of A^*A by $\sigma_1^2 \geq \ldots \geq \sigma_r^2 > 0$.
    - ▷ It can be proved $r = \text{rank}(A)$.
  - ◇ Denote

$$A^*A\mathbf{v}_i = \sigma_i^2\mathbf{v}_i, \quad i = 1, \ldots, n.$$

- Some important observations:
  - ◇ Matrices A^*A and AA^* have the same set of positive eigenvalues.

$$(AA^*)A\mathbf{v}_i = \sigma_i^2 A\mathbf{v}_i.$$

  - ◇ The vector $\mathbf{u}_i := A\mathbf{v}_i/\sigma_i$ is of unit length.
  - ◇ $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ are orthonormal eigenvector of AA^*.
  - ◇ Can choose $\{\mathbf{u}_{r+1}, \ldots, \mathbf{u}_m\}$ to be orthonormal eigenvectors of AA^* with eigenvalue 0. (How to achieve this?)

- The *singular value decomposition* of A:
  - ◇ Denote

$$\begin{cases} V & := & [\mathbf{v}_1, \ldots, \mathbf{v}_n] \in \mathbb{C}^{n\times n} \\ U & := & [\mathbf{u}_1, \ldots, \mathbf{u}_m] \in \mathbb{C}^{m\times m} \\ \Sigma & := & \mathrm{diag}\{\sigma_1, \ldots, \sigma_r\} \in \mathbb{R}^{r\times r}. \end{cases}$$

  - ◇ Any $A \in \mathbb{C}^{m\times n}$ can be decomposed as a triplet

$$A = U \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} V^*. \tag{1.40}$$

    - ▷ Columns of U and V are orthonormal.
    - ▷ Can also expressed as

$$A = \sum_i \sigma_i \mathbf{u}_i \mathbf{v}_i^*. \tag{1.41}$$

      ✓ How to compromise the notation issue, $\mathbf{u}_i\mathbf{v}_i^* = \mathbf{u}_i \otimes \mathbf{v}_i$?
  - ◇ Write $U = [U_1, U_2],\ V = [V_1, V_2]$, then

$$U^*AV = \begin{bmatrix} U_1^* \\ U_2^* \end{bmatrix} A[V_1, V_2] = \begin{bmatrix} U_1^*AV_1 & U_1^*AV_2 \\ U_2^*AV_1 & U_2^*AV_2 \end{bmatrix}.$$

    - ▷ Note that

$$\begin{cases} AV_2 & = & 0, \\ U_2^*AV_1 & = & U_2^*U_1\Sigma = 0, \\ U_1^*AV_1 & = & \Sigma, \end{cases}$$

    by the choice of U.

Schmidt Decomposition

- Given two Hilbert spaces $\mathscr{H}$ and $\mathscr{K}$, assume that the tensor product space $\mathscr{H} \otimes \mathscr{K}$ is complete. Then any *unit* vector $\boldsymbol{\psi} \in \mathscr{H} \otimes \mathscr{K}$ can be written in the form

$$\boldsymbol{\psi} = \sum_k \sigma_k \mathbf{h}_k \otimes \mathbf{k}_k \tag{1.42}$$

where the vectors $\mathbf{h}_k \in \mathscr{H}$ and $\mathbf{k}_k \in \mathscr{K}$ are orthonormal, $\sigma_k > 0$ and $\sum_k \sigma_k^2 = 1$.

  - ◇ The number of terms in the summation, called the Schmidt number, is finite and $\boldsymbol{\psi}$ dependent. (Rank of a matrix $\boldsymbol{\psi}$?)

- The proof is in the spirit as the SVD:
  - ◇ Think of the representation $\boldsymbol{\psi} = \sum_{i,j} c_{ij} \mathbf{e}_i \otimes \mathbf{f}_j$.
  - ◇ Let $C = U\Sigma V^*$ be the (full) SVD of C.
    - ▷ $\|C\|_F = 1$, so $\sum_i \sigma_i^2 = 1$.
  - ◇ Write $U = [u_{st}]$ and $V = [v_{\mu\nu}]$. Then

$$\boldsymbol{\psi} = \sum_{i,j} (\sum_{k=1}^{r} \sigma_k u_{ik} \overline{v}_{jk}) \mathbf{e}_i \otimes \mathbf{f}_j.$$

  - ◇ Suffices to define

$$\begin{cases} \mathbf{h}_k := \sum_i u_{ik} \mathbf{e}_i, \\ \mathbf{k}_k := \sum_j \overline{v}_{jk} \mathbf{f}_j. \end{cases}$$

- Is there a generalization of SVD to higher dimensional tensors?