

# CS 6604: Quantum Information & Semidefinite Programming

*Instructor:* Jamie Sikora

*Editor:* Alice Zheng

*Taught in:* Spring 2024, Fall 2026

|          |                                                       |           |
|----------|-------------------------------------------------------|-----------|
| <b>1</b> | <b>Mathematical Background</b>                        | <b>1</b>  |
| 1.1      | Complex Euclidean spaces . . . . .                    | 1         |
| 1.2      | Operators . . . . .                                   | 5         |
| 1.3      | Superoperators . . . . .                              | 14        |
| 1.4      | Definite operators . . . . .                          | 15        |
| 1.5      | Real analysis . . . . .                               | 20        |
| 1.6      | Convex analysis . . . . .                             | 26        |
| <b>2</b> | <b>Semidefinite Programming</b>                       | <b>33</b> |
| 2.1      | Introduction to semidefinite programming . . . . .    | 33        |
| 2.2      | Standard form . . . . .                               | 37        |
| 2.3      | Duality theory . . . . .                              | 40        |
| 2.4      | Dual of the dual . . . . .                            | 44        |
| 2.5      | A simpler way to take the dual . . . . .              | 45        |
| 2.6      | Optimality conditions: Slater's theorem . . . . .     | 51        |
| 2.7      | Clark's theorem and ascent directions . . . . .       | 54        |
| <b>3</b> | <b>Quantum Information</b>                            | <b>59</b> |
| 3.1      | Quantum states and measurements . . . . .             | 59        |
| 3.2      | Distinguishing between two states . . . . .           | 64        |
| 3.3      | Distinguishing between many states . . . . .          | 68        |
| 3.4      | Duality of Uhlmann's and Alberti's theorems . . . . . | 70        |
| 3.5      | Channel norms and fidelity properties . . . . .       | 74        |
| 3.6      | Operator geometric means . . . . .                    | 77        |
| 3.7      | Quantum entropy functions . . . . .                   | 81        |

# Introduction

This course introduces the topic of semidefinite programming and its use in the study of quantum information concepts. It is designed to be a graduate-level seminar course. There are two primary parts: in Chapter 2 we cover the foundations of semidefinite programming from an optimization standpoint, and in Chapter 3 we focus on applying the former to topics and quantum information. Readers interested in only one of the two parts may nevertheless find this course informative; semidefinite programming is a useful toolset for analyzing quantum theory, which in turn provides operational meaning to the former. Quantum information is far from the only use case of semidefinite programming, the concepts presented in this course being applicable to a variety of problems in computer science and other fields.

Semidefinite programming is a powerful class of optimization problems with frequent uses both in analytical proofs and finding solutions via relaxations. We introduce semidefinite programming with a focus on the relationship between primal and dual problems, including results such as Slater’s and Clark’s theorems. Quantum mechanics emerged around the turn of the 20th century as a leading physical theory, since having broad implications even reaching algorithm design and information processing. We introduce various quantum information concepts, motivating the generality of semidefinite programming as analogous to the extension of classical objects to quantum, and hence fitting for modeling. We derive semidefinite programs for key concepts in quantum information, using them to present new ways to derive important theorems. A more detailed overview of the covered topics can be glimpsed from the preceding page.

These lecture notes are indented as an additional resource, containing a more polished version of the same information present in class lectures and slides. This document is a work in progress, with future versions intended to include additional lectures on quantum applications. We welcome and greatly appreciate any reader feedback and corrections.

**Prerequisites:** No formal prerequisites, except for some familiarity with linear algebra. Chapter 1 provides an overview of the necessary concepts, though it is meant mostly as a refresher with somewhat rapid pacing. Prior experience with an introductory quantum and/or optimization course is helpful, but optional.

**Additional resources:** This course and its lecture notes aim to be self-contained, covering the sufficient background and introductory content. For greater detail of relevant concepts, avid readers are encouraged to pursue the following supplementary materials:

- John Watrous, *The Theory of Quantum Information*, Cambridge University Press, 2018.
- Michael A. Nielsen and Isaac L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition*, Cambridge University Press, 2011.

**Acknowledgements:** These lecture notes would not be possible without students in the Spring 2024 offering, who have kindly provided an initial set of scribed notes:

|                    |              |                |
|--------------------|--------------|----------------|
| Akshay Bansal      | Sumin Kang   | Ankith Mohan   |
| Keshav Bhateja     | Luis Lazcano | Raseen Nirjhar |
| Turbasu Chatterjee | Hyunwoo Lee  | Alice Zheng    |
| Myles Dixon        | Paul Maynard |                |

## Notation

| Symbol                                                                     | Meaning                                                                             | Page        |
|----------------------------------------------------------------------------|-------------------------------------------------------------------------------------|-------------|
| $\mathcal{X}, \mathcal{Y}, \mathbb{C}^n$                                   | Complex Euclidean space ( $n$ -dimensional)                                         | 1, 2        |
| $A, B \in \mathcal{L}(\mathcal{X}, \mathcal{Y}); \mathcal{L}(\mathcal{X})$ | Operators from $\mathcal{X}$ to $\mathcal{Y}$ ; from $\mathcal{X}$ to $\mathcal{X}$ | 5           |
| $\rho, \sigma \in \mathcal{D}(\mathcal{X})$                                | Density operators (quantum states) on $\mathcal{X}$                                 | 59          |
| $E \in \text{Eff}(\mathcal{X})$                                            | Effects (2-outcome measurements) on $\mathcal{X}$                                   | 63          |
| $\Phi, \Psi \in \mathcal{T}(\mathcal{X}, \mathcal{Y})$                     | Linear maps from $\mathcal{L}(\mathcal{X})$ to $\mathcal{L}(\mathcal{Y})$           | 14          |
| $\Phi, \Psi \in \mathcal{C}(\mathcal{X}, \mathcal{Y})$                     | CPTP maps (channels) from $\mathcal{L}(\mathcal{X})$ to $\mathcal{L}(\mathcal{Y})$  | 74          |
| $\mathbb{1}$                                                               | Identity                                                                            | 9, 14       |
| $\text{Herm}(\mathcal{X})$                                                 | Hermitian operators                                                                 | 8           |
| $\mathcal{U}(\mathcal{X}, \mathcal{Y}), \mathcal{U}(\mathcal{X})$          | Isometries from $\mathcal{X}$ to $\mathcal{Y}$ , Unitaries on $\mathcal{X}$         | 9           |
| $A \in \text{Pos}(\mathcal{X}), A \succcurlyeq 0$                          | Positive semidefinite                                                               | 16          |
| $A \in \text{Pd}(\mathcal{X}), A \succ 0$                                  | Positive definite                                                                   | 16          |
| $A \preccurlyeq 0$                                                         | Negative semidefinite                                                               | 16          |
| $A \prec 0$                                                                | Negative definite                                                                   | 16          |
| $\mathcal{X} \times \mathcal{Y}$                                           | Direct product                                                                      | 2           |
| $\mathcal{X} \oplus \mathcal{Y}$                                           | Direct sum                                                                          | 3           |
| $\mathcal{X} \otimes \mathcal{Y}, A \otimes B, \Phi \otimes \Psi$          | Tensor product                                                                      | 3, 7, 14    |
| $\langle \cdot, \cdot \rangle$                                             | Inner product                                                                       | 2, 12       |
| $\ \cdot\ $                                                                | Norm                                                                                | 3, 12       |
| $\text{Tr}(\cdot), \text{Tr}_{\mathcal{X}}(\cdot)$                         | Trace, Partial trace                                                                | 11, 14, 14  |
| $\bar{x}, \bar{A}$                                                         | Complex conjugate                                                                   | 2, 6        |
| $x^T, A^T$                                                                 | Transpose                                                                           | 2, 7        |
| $x^*, A^*, \ \cdot\ ^*, S^*$                                               | Adjoint (conjugate transpose), Dual norm, Dual cone                                 | 2, 7, 4, 29 |
| $S^\circ$                                                                  | Polar set                                                                           | 29          |
| $S^c$                                                                      | Set complement                                                                      | 21          |
| $x(i)$                                                                     | $i$ th entry of $x \in \mathbb{C}^n$                                                | 2           |
| $e_i, E_{i,j}$                                                             | Basis vectors, Basis operators                                                      | 4, 14       |
| $\sigma[A], \lambda[A]$                                                    | Singular values of $A$ , Eigenvalues of $A$                                         | 10, 11      |
| $B_r(x)$                                                                   | Ball of radius $r$ around $x$                                                       | 21          |

# Chapter 1

## Mathematical Background

The first few lectures are meant as an introduction to various mathematical concepts relevant to semidefinite programming. We begin with a crash course in linear algebra in the first three lectures, subsequently advancing to topics such as positive semidefiniteness, analysis, and convexity.

These lectures are meant as a part refresher, part notation reference, and a much larger part future resource. While some familiarity with general linear algebra concepts is assumed, the reader is not expected to already know, memorize, or even entirely comprehend things right away.

The content that follows is not the most general, most rigorous, or most detailed discussion of linear algebra and related topics. Concepts appear here due to their relevance to the rest of the course, with many more omitted for conciseness or clarity.

### 1.1 Complex Euclidean spaces

While linear algebra may be largely viewed as an interplay of vectors and matrices, we take the more general approach of treating vectors as belonging to some **vector space**: a non-empty set  $\mathcal{V}$  imbued with vector addition and scalar multiplication, where all scalars belong to some field. This has the advantage of not requiring such vectors to take any particular form, making many of the relevant notions generalize to higher orders of objects.

In this course, we primarily work with the following class of vector spaces.

**Definition 1.1.1.** A **complex Euclidean space** is a vector space over the field of *complex* numbers  $\mathbb{C}$  that is *finite-dimensional* and defines an *inner product* (explained in Definition 1.1.2).

We denote such spaces by calligraphic letters near the end of the Latin alphabet, like  $\mathcal{X}$  or  $\mathcal{Y}$ .

*Remark.* These are occasionally referred to as *Hilbert spaces*, which is a generalization (with appropriate restrictions) to possibly infinite dimensions. In this course, we instead use the term “space” to refer to complex Euclidean spaces in particular.

In a general sense, vectors in a complex Euclidean space are functions mapping some set of indices to complex values. Vectors with  $n$  possible index values can be thought of as ordered collections of  $n$  complex numbers. Without losing generality, we can label such indices 1 through

$n$ . A common notation for such complex Euclidean spaces and vectors is

$$\mathbb{C}^n = \left\{ \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} : x_1, \dots, x_n \in \mathbb{C} \right\}. \quad (1.1)$$

To avoid notational conflicts, we generally use  $x(i)$  instead of  $x_i$  to denote the  $i$ th entry of  $x \in \mathbb{C}^n$ .

When the entries of a vector  $x \in \mathbb{C}^n$  are stacked vertically as in (1.1), it is called a *column vector*. To obtain a *row vector*, **transpose**  $x$  like  $x^T = (x_1 \dots x_n)$ . The **complex conjugate**  $\bar{x}$  is obtained from  $x$  by taking the complex conjugate (i.e., negating the imaginary part) of each entry. Doing both operations, regardless of order, yields the *complex conjugate transpose* or the **adjoint**  $x^* = \bar{x}^T$ .

**Definition 1.1.2.** An **inner product** of a complex vector space  $\mathcal{V}$  is a binary function  $\langle \cdot, \cdot \rangle : \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{C}$  which satisfies

$$\langle u, v \rangle = \overline{\langle v, u \rangle}, \quad (1.2)$$

$$\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle \quad \text{and} \quad (1.3a)$$

$$\langle u, v + w \rangle = \langle u, v \rangle + \langle u, w \rangle, \quad (1.3b)$$

$$\langle ku, v \rangle = \bar{k} \langle u, v \rangle \quad \text{and} \quad (1.4a)$$

$$\langle u, kv \rangle = k \langle u, v \rangle, \quad (1.4b)$$

$$\langle u, u \rangle \geq 0 \quad \text{with} \quad (1.5a)$$

$$\langle u, u \rangle = 0 \quad \text{if and only if} \quad u = 0, \quad (1.5b)$$

for all  $u, v, w \in \mathcal{V}$  and  $k \in \mathbb{C}$ . Note that some of the above properties are implied by others.

The standard inner product on  $\mathbb{C}^n$ , for any  $u, v \in \mathbb{C}^n$ , is

$$\langle u, v \rangle := \sum_{i=1}^n \bar{u}(i) v(i) = u^* v. \quad (1.6)$$

Note that the latter form is a product of a column and a row vector. The intuition behind this notation should become clear once we define matrix products.

**Example.** Let  $u = \begin{pmatrix} i \\ 1 \end{pmatrix}$  and  $v = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ . Their inner product is then  $\langle u, v \rangle = \bar{i} \cdot 1 + \bar{1} \cdot 2 = -i + 2$ .

### 1.1.1 Combinations of spaces

For complex Euclidean spaces  $\mathcal{X} = \mathbb{C}^m$  and  $\mathcal{Y} = \mathbb{C}^n$ , we consider the following combinations.

**Direct product:**  $(x, y) \in \mathcal{X} \times \mathcal{Y}$  if and only if  $x \in \mathcal{X}$  and  $y \in \mathcal{Y}$ . This also works for arbitrary sets.

**Remark.** The direct product is technically not associative. Given an element of a third space  $z \in \mathcal{Z}$ ,  $(\mathcal{X} \times \mathcal{Y}) \times \mathcal{Z}$  and  $\mathcal{X} \times (\mathcal{Y} \times \mathcal{Z})$  contain elements of the form  $((x, y), z)$  and  $(x, (y, z))$ , respectively. They may nevertheless be viewed as equivalent by removing parentheses, and the direct product may thus be treated as being associative.

**Direct sum:**  $\mathcal{X} \oplus \mathcal{Y} = \mathbb{C}^{m+n}$ . When  $x \in \mathcal{X}$  and  $y \in \mathcal{Y}$  are column vectors, we can stack them vertically to obtain  $x \oplus y = \begin{bmatrix} x \\ y \end{bmatrix}$ . Vectors of this form cover the entirety of  $\mathcal{X} \oplus \mathcal{Y}$ .

**Tensor product:**  $\mathcal{X} \otimes \mathcal{Y} = \mathbb{C}^{mn}$ , implying that  $x \otimes y$  has an entry for each pair of those in  $x \in \mathcal{X}$  and  $y \in \mathcal{Y}$ . Vectors of this form span the tensor product space  $\mathcal{X} \otimes \mathcal{Y}$ , but not every element of the space can be written as a tensor product. Tensor products of vectors are defined as follows.

**Definition 1.1.3.** For any  $x \in \mathbb{C}^m$  and  $y \in \mathbb{C}^n$ , the vector  $x \otimes y \in \mathbb{C}^{mn}$  uniquely satisfies

$$(x \otimes y)(i, j) = x(i) \cdot y(j) \quad (1.7)$$

for all  $i \in \{1, \dots, m\}$  and  $j \in \{1, \dots, n\}$ .

Each of the above notions extends to combinations of larger numbers of complex Euclidean spaces, so long as the number of combined spaces is finite and countable. We use the shorthands  $\mathcal{X}^{\times n}$ ,  $\mathcal{X}^{\oplus n}$ , and  $\mathcal{X}^{\otimes n}$  to denote the respective combinations of  $n$  copies of the same space  $\mathcal{X}$ .

### 1.1.2 Norms

**Definition 1.1.4.** A **norm** of a vector space  $\mathcal{V}$  is a unary function  $\|\cdot\| : \mathcal{V} \rightarrow \mathbb{R}$  which satisfies

$$\|u + v\| \leq \|u\| + \|v\| \quad (\text{triangle inequality}), \quad (1.8)$$

$$\|ku\| = |k| \cdot \|u\| \quad (\text{so } \|0\| = 0), \quad (1.9)$$

$$\|u\| \geq 0, \quad (1.10)$$

$$\|u\| = 0 \quad \text{only if} \quad u = 0, \quad (1.11)$$

for all  $u, v \in \mathcal{V}$  and  $k \in \mathbb{C}$ .

The following are common vector norms for the vector space  $\mathbb{C}^n$ . For all  $x \in \mathbb{C}^n$ ,

$$\|x\|_1 := \sum_{i=1}^n |x(i)|, \quad (1.12)$$

$$\|x\|_2 := \sqrt{\sum_{i=1}^n |x(i)|^2}, \quad (1.13)$$

$$\|x\|_\infty := \max_{i \in \{1, \dots, n\}} |x(i)|. \quad (1.14)$$

Each of the above is a special case of a more general  $p$ -norm; for any  $p \in [1, \infty]$ , we define

$$\|x\|_p := \left( \sum_{i=1}^n |x(i)|^p \right)^{1/p}. \quad (1.15)$$

Inner products are additionally convenient due to being able to *induce* a norm. Specifically, having an inner product allows defining  $\|x\|_2 = \sqrt{\langle x, x \rangle}$ .

**Example.** Let  $x = \begin{pmatrix} 1 \\ 2 \\ -3i \end{pmatrix}$ . Note that  $|x(1)| = 1$ ,  $|x(2)| = 2$ , and  $|x(3)| = \sqrt{0^2 + (-3)^2} = 3$ . Then,

$$\|x\|_1 = 1 + 2 + 3 = 6, \quad (1.16)$$

$$\|x\|_2 = \sqrt{1^2 + 2^2 + 3^2} = \sqrt{1 + 4 + 9} = \sqrt{14}, \quad (1.17)$$

$$\|x\|_\infty = \max\{1, 2, 3\} = 3. \quad (1.18)$$

Note that  $\|x\|_\infty \leq \|x\|_2 \leq \|x\|_1$  in the above example. In general,  $\|x\|_q \leq \|x\|_p$  for  $q \geq p$ .

**Definition 1.1.5.** Let  $\|\cdot\|$  be a norm on a complex Euclidean space  $\mathcal{X}$ . Its **dual norm**, denoted by  $\|\cdot\|^*$ , is defined as

$$\|u\|^* := \sup_{v \in \mathcal{X}} \{|\langle u, v \rangle| : \|v\| = 1\}. \quad (1.19)$$

Some of the norms we have already seen are dual to each other. For instance,  $\|\cdot\|_p^* = \|\cdot\|_q$ , where  $\frac{1}{p} + \frac{1}{q} = 1$ . Knowing the formula of a dual norm is useful for the following theorem.

**Theorem 1.1.6** (Hölder's inequality). *For all  $u, v$  in a complex Euclidean space  $\mathcal{X}$ ,*

$$|\langle u, v \rangle| \leq \|u\|^* \|v\|. \quad (1.20)$$

The special case of Hölder's inequality corresponding to 2-norms (i.e.,  $p = q = 2$ ) is called the Cauchy-Schwarz inequality.

### 1.1.3 Bases

Replacing  $x \in \mathcal{X}$  with  $x/\|x\|$  (when  $x \neq 0$ ) is referred to as *normalizing*  $x$ . Note that a normalized  $x$  is one such that  $\|x\| = 1$ . Such vectors are also known as **unit vectors**.

**Definition 1.1.7.** A set of vectors  $\{v_1, \dots, v_m\} \subseteq \mathbb{C}^n$  is an **orthonormal set** if each vector is a unit vector and each pair of distinct vectors is orthogonal. In other words, assuming  $1 \leq i, j \leq m$ ,  $\|v_i\| = 1$  for all  $i$  and  $\langle v_i, v_j \rangle = 0$  for all  $i \neq j$ , or equivalently,

$$\langle v_i, v_j \rangle = \delta(i, j) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}. \quad (1.21)$$

An **orthonormal basis** is a basis that is also an orthonormal set.

**Example.** The following sets are orthonormal bases for their respective vector spaces.

$$\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\} \subset \mathbb{C}^3, \quad \left\{ \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}, \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix} \right\} \subset \mathbb{C}^2. \quad (1.22)$$

The basis for  $\mathbb{C}^3$  shown in (1.22) is an example of a **standard basis** or *computational basis*. We use  $e_i$  to denote such vectors, in which the  $i$ th entry is 1 and the rest are 0. The example above can then be rewritten as  $\{e_1, e_2, e_3\} \subset \mathbb{C}^3$ .

## 1.2 Operators

Up until this point, we have only considered elements of complex Euclidean spaces to be vectors. It is now time to move up a level and introduce matrices, which we do in a somewhat more general way. In doing so, many of the notions relevant for vectors should be extremely similar or directly applicable to matrices, though proper definitions of them may require some work.

**Definition 1.2.1.** A linear function  $A : \mathcal{X} \rightarrow \mathcal{Y}$  is one that satisfies

$$A(u + v) = A(u) + A(v) , \quad (1.23)$$

$$A(ku) = k A(u) , \quad (1.24)$$

for all  $u, v \in \mathcal{X}$  and  $k \in \mathbb{C}$ .

We denote the set of all such linear maps from  $\mathcal{X}$  to  $\mathcal{Y}$  by  $L(\mathcal{X}, \mathcal{Y})$ , and call them **operators**. If the spaces are the same, we can simplify the notation to  $L(\mathcal{X}) := L(\mathcal{X}, \mathcal{X})$ .

The property of mapping vectors to vectors makes it important to consider operators as more than just a larger collection of complex numbers. By the end of this lecture, we take this further up the levels of abstraction to map operators to operators (and beyond).

In writing down the explicit representation of a linear operator, it becomes important to establish the equivalence between operators and **matrices** in finite dimensions. In particular, we denote the set of  $m$ -by- $n$  complex matrices as

$$\mathbb{C}^{m \times n} = \left\{ \begin{pmatrix} x_{1,1} & \cdots & x_{1,n} \\ \vdots & \ddots & \vdots \\ x_{m,1} & \cdots & x_{m,n} \end{pmatrix} : x_{i,j} \in \mathbb{C} \right\} . \quad (1.25)$$

This also implies that any such matrix can be written down as an  $m$ -by- $n$  array of complex numbers. When  $M \in \mathbb{C}^{m \times n}$ , we denote the entry in its  $i$ th row and  $j$ th column by  $M(i, j)$ , which corresponds to  $x_{i,j}$  above. We can use the matrix  $M$  to define an operator  $A \in L(\mathbb{C}^n, \mathbb{C}^m)$  via its action on any  $u \in \mathbb{C}^n$  as

$$(Au)(i) = \sum_{j=1}^n M(i, j) u(j) \quad (1.26)$$

for any  $i \in \{1, \dots, m\}$ . Conversely, given an operator  $A$ , we can define  $M$  via

$$M(i, j) = e_i^* A e_j = \langle e_i, A e_j \rangle \quad (1.27)$$

for any  $i \in \{1, \dots, m\}$  and  $j \in \{1, \dots, n\}$ . Thus, the spaces  $\mathbb{C}^{m \times n}$  and  $L(\mathbb{C}^n, \mathbb{C}^m)$  may be viewed as being equivalent.

**Example.** Let  $A = \begin{pmatrix} 1 & 2 \\ 2 & 2 \\ 2 & 3 \end{pmatrix} \in L(\mathbb{C}^2, \mathbb{C}^3)$ , and let  $u = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \in \mathbb{C}^2$ . Then,  $Au = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} \in \mathbb{C}^3$ .

### 1.2.1 Matrix operations

We now consider a few notable ways to transform operators to operators.

As alluded to before, a benefit of considering general vector spaces is that operators/matrices likewise form a vector space. To show this, we need to properly define addition and scalar multiplication (much like we did for vectors), which we do for both operators and matrices.

For operators  $A, B \in L(\mathcal{X}, \mathcal{Y})$ ,  $u \in \mathcal{X}$ , and  $k \in \mathbb{C}$ , define addition and scalar multiplication as

$$(A + B)u = Au + Bu, \quad (1.28)$$

$$(kA)u = kAu, \quad (1.29)$$

where both of  $A + B$  and  $kA$  are in  $L(\mathcal{X}, \mathcal{Y})$ . If instead  $A \in L(\mathcal{Z}, \mathcal{Y})$ ,  $B \in L(\mathcal{X}, \mathcal{Z})$ , and  $u \in \mathcal{X}$ , we can compose  $B$  and  $A$  to obtain  $AB \in L(\mathcal{X}, \mathcal{Y})$  via

$$(AB)u = A(Bu). \quad (1.30)$$

While (1.28) and (1.29) show that operators form a vector space, they do not directly tell us how to explicitly get any particular entry of the above operators. For matrices  $A, B \in \mathbb{C}^{m \times n}$  and  $k \in \mathbb{C}$ , the entries in row  $i \in \{1, \dots, m\}$  and column  $j \in \{1, \dots, n\}$  of  $A + B, kA \in \mathbb{C}^{m \times n}$  are

$$(A + B)(i, j) = A(i, j) + B(i, j), \quad (1.31)$$

$$(kA)(i, j) = kA(i, j). \quad (1.32)$$

In this context, composition is known as *matrix multiplication*. For any  $A \in \mathbb{C}^{m \times l}$  and  $B \in \mathbb{C}^{l \times n}$ ,

$$(AB)(i, j) = \sum_{k=1}^l A(i, k) B(k, j), \quad (1.33)$$

where  $AB \in \mathbb{C}^{m \times n}$  and thus  $i \in \{1, \dots, m\}, j \in \{1, \dots, n\}$ . Note that the two sets of equations, (1.28) to (1.30) and (1.31) to (1.33), are equivalent since they can be deduced from each other via (1.26) and (1.27).

Let us now make an observation about vector multiplication. Column vectors  $u, v \in \mathbb{C}^n$  can also be thought of as  $n$ -by-1 matrices. By matrix multiplication,  $u^*v$  is a scalar, which is why the last part of (1.6) makes sense. On the other hand,  $uv^*$  is an  $n$ -by- $n$  matrix.

**Example.** Let  $u = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$  and  $v = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$ . Then,  $u^*v = 1 \cdot 3 + 2 \cdot 4 = 11$  and  $uv^* = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \begin{pmatrix} 3 & 4 \end{pmatrix} = \begin{pmatrix} 3 & 4 \\ 6 & 8 \end{pmatrix}$ .

The matrix  $e_i e_j^*$  specifically has a 1 entry in the  $i$ th row and  $j$ th column, and 0 everywhere else. We revisit this idea near the end of this lecture, but for now it allows us to write any  $A \in \mathbb{C}^{m \times n}$  as

$$A = \sum_{i=1}^m \sum_{j=1}^n A(i, j) e_i e_j^*. \quad (1.34)$$

This in turn allows us to extend several operations to matrices. The complex conjugate is perhaps the simplest; as in the vector case, we conjugate each entry separately.

**Definition 1.2.2.** The **complex conjugate** of  $A \in \mathbb{C}^{m \times n}$  is denoted by  $\bar{A} \in \mathbb{C}^{m \times n}$  and defined as

$$\bar{A} = \sum_{i=1}^m \sum_{j=1}^n \overline{A(i, j)} e_i e_j^*. \quad (1.35)$$

As the transpose turns column vectors into row vectors, it flips the dimensions of matrices.

**Definition 1.2.3.** The **transpose** of  $A \in \mathbb{C}^{m \times n}$  is denoted by  $A^T \in \mathbb{C}^{n \times m}$  and defined as

$$A^T = \sum_{i=1}^m \sum_{j=1}^n A(i, j) e_j e_i^* = \sum_{i=1}^n \sum_{j=1}^m A(j, i) e_i e_j^* . \quad (1.36)$$

The same change of dimensions is done by the adjoint, which we define for linear operators.

**Definition 1.2.4.** The **adjoint** of  $A \in L(\mathcal{X}, \mathcal{Y})$  is the unique linear operator  $A^* \in L(\mathcal{Y}, \mathcal{X})$  satisfying

$$\langle Ax, y \rangle = \langle x, A^* y \rangle \quad (1.37)$$

for all  $x \in \mathcal{X}$  and  $y \in \mathcal{Y}$ . Similar to the vector case, we may alternatively define  $A^* = \overline{A}^T$ .

*Remark.* We have picked a particular choice of basis (the computational basis) when defining the complex conjugate and the transpose, though the definition would work just as well with any other orthonormal basis. The complex conjugate and the transpose are thus *basis-dependent*. As the adjoint was defined without a basis, it is *basis-independent*.

We previously introduced the idea of tensor products, defining them for complex Euclidean spaces and vectors (Definition 1.1.3). We now extend them to operators, and then to matrices.

**Definition 1.2.5.** Let  $\mathcal{X}_A, \mathcal{X}_B, \mathcal{Y}_A$ , and  $\mathcal{Y}_B$  be complex Euclidean spaces. The **tensor product** of  $A \in L(\mathcal{X}_A, \mathcal{Y}_A)$  and  $B \in L(\mathcal{X}_B, \mathcal{Y}_B)$  is the unique linear operator

$$A \otimes B \in L(\mathcal{X}_A \otimes \mathcal{X}_B, \mathcal{Y}_A \otimes \mathcal{Y}_B) , \quad (1.38)$$

satisfying for all  $u \in \mathcal{X}_A$  and  $v \in \mathcal{X}_B$

$$(A \otimes B)(u \otimes v) = (Au) \otimes (Bv) . \quad (1.39)$$

We often consider a particular specification of tensor products, defined for matrices as follows.

**Definition 1.2.6.** The **Kronecker product** of two matrices

$$A = \begin{pmatrix} A(1,1) & \dots & A(1,n) \\ \vdots & \ddots & \vdots \\ A(m,1) & \dots & A(m,n) \end{pmatrix} \in \mathbb{C}^{m \times n} , \quad B \in \mathbb{C}^{k \times l} \quad (1.40)$$

is a tensor product, defined as the block matrix

$$A \otimes B = \begin{bmatrix} A(1,1) B & \dots & A(1,n) B \\ \vdots & \ddots & \vdots \\ A(m,1) B & \dots & A(m,n) B \end{bmatrix} \in \mathbb{C}^{mk \times nl} . \quad (1.41)$$

*Remark.* In practice, the terms *Kronecker product* and *tensor product* are often interchangeable.

**Example.** Let  $A = \begin{pmatrix} 2 & -1 \end{pmatrix} \in \mathbb{C}^{1 \times 2}$  and  $B = \begin{pmatrix} 1 & 2 & 2 \\ 2 & 2 & 3 \end{pmatrix} \in \mathbb{C}^{2 \times 3}$ . Their Kronecker product is

$$A \otimes B = [2B \quad -1B] = \begin{pmatrix} 2 & 4 & 4 & -1 & -2 & -2 \\ 4 & 4 & 6 & -2 & -2 & -3 \end{pmatrix} \in \mathbb{C}^{1 \cdot 2 \times 2 \cdot 3} = \mathbb{C}^{2 \times 6}. \quad (1.42)$$

**Lemma 1.2.7** (Properties of tensor products). *For any  $k \in \mathbb{C}$  and operators  $A \in L(\mathcal{X}_A, \mathcal{Y}_A)$ ,  $B \in L(\mathcal{X}_B, \mathcal{Y}_B)$ ,  $C \in L(\mathcal{X}_C, \mathcal{Y}_C)$ , and  $D \in L(\mathcal{X}_D, \mathcal{Y}_D)$ , the following identities hold.*

$$A \otimes B \neq B \otimes A \quad \text{in general,} \quad (1.43)$$

$$A \otimes (B + C) = A \otimes B + A \otimes C \quad \text{when } \mathcal{X}_B = \mathcal{X}_C \text{ and } \mathcal{Y}_B = \mathcal{Y}_C, \quad (1.44)$$

$$(A + B) \otimes C = A \otimes C + B \otimes C \quad \text{when } \mathcal{X}_A = \mathcal{X}_B \text{ and } \mathcal{Y}_A = \mathcal{Y}_B, \quad (1.45)$$

$$(kA) \otimes B = A \otimes (kB) = k(A \otimes B), \quad (1.46)$$

$$(A \otimes B) \otimes C = A \otimes (B \otimes C), \quad (1.47)$$

$$(A \otimes B)(C \otimes D) = AC \otimes BD \quad \text{when } \mathcal{X}_A = \mathcal{Y}_C \text{ and } \mathcal{X}_B = \mathcal{Y}_D, \quad (1.48)$$

$$(A \otimes \mathbb{1}_{\mathcal{Y}_D})(\mathbb{1}_{\mathcal{X}_A} \otimes D) = A \otimes D, \quad (1.49)$$

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1} \quad \text{when } A \text{ and } B \text{ are invertible,} \quad (1.50)$$

$$(A \otimes B)^* = A^* \otimes B^*. \quad (1.51)$$

The conditions in (1.44) and (1.45) are such that the constituents of  $B + C$  and  $A + B$  have equal dimensions, and in (1.48) such that  $AC$  and  $BD$  make sense. For  $A$  and  $B$  to be invertible in (1.50) they must be square operators, meaning  $\mathcal{X}_A = \mathcal{Y}_A$  and  $\mathcal{X}_B = \mathcal{Y}_B$  (see Definition 1.2.12).

## 1.2.2 Classes of operators

We now define several important classes of operators, starting with a general one.

**Definition 1.2.8.** A **normal** operator  $A \in L(\mathcal{X})$  acting on a complex Euclidean space  $\mathcal{X}$  is one that commutes with its adjoint, meaning  $AA^* = A^*A$ .

This definition restricts  $A$  to be square. Most classes below are subclasses of normal operators, and so they are square as well. Normal operators carry some useful implications, one being an important decomposition that we define shortly. In the absence of other restrictions, in this course we assume operators to be normal. The class of normal operators thus receives no special notation.

**Example.** Let  $A = \begin{pmatrix} 2 & i \\ 1 & 2 \end{pmatrix} \in \mathbb{C}^{2 \times 2}$ . Then,  $A^* = \begin{pmatrix} 2 & 1 \\ -i & 2 \end{pmatrix}$  and  $A$  is normal since

$$AA^* = \begin{pmatrix} 2 \cdot 2 + i \cdot (-i) & 2 \cdot 1 + i \cdot 2 \\ 1 \cdot 2 + 2 \cdot (-i) & 1 \cdot 1 + 2 \cdot 2 \end{pmatrix} = \begin{pmatrix} 5 & 2 + 2i \\ 2 - 2i & 5 \end{pmatrix}, \quad (1.52a)$$

$$A^*A = \begin{pmatrix} 2 \cdot 2 + 1 \cdot 1 & 2 \cdot i + 1 \cdot 2 \\ -i \cdot 2 + 2 \cdot 1 & -i \cdot i + 2 \cdot 2 \end{pmatrix} = \begin{pmatrix} 5 & 2 + 2i \\ 2 - 2i & 5 \end{pmatrix}. \quad (1.52b)$$

**Definition 1.2.9.** A **Hermitian**, or *self-adjoint*, operator  $A \in L(\mathcal{X})$  acting on a complex Euclidean space  $\mathcal{X}$  is one that equals its adjoint, meaning  $A = A^*$ . The set of all such

Hermitian operators on  $\mathcal{X}$  is denoted by  $\text{Herm}(\mathcal{X})$ .

It directly follows from the definition that every Hermitian operator is normal.

*Example.*  $A = \begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix} = A^*$ . Thus,  $A \in \text{Herm}(\mathbb{C}^2)$ .

Before covering other classes of operators, let us define the notions of identity and inverses.

**Definition 1.2.10.** The **identity** operator  $\mathbb{1}_{\mathcal{X}} \in L(\mathcal{X})$  acting on a complex Euclidean space  $\mathcal{X}$  is one that uniquely satisfies  $\mathbb{1}_{\mathcal{X}}(u) = u$  for all  $u \in \mathcal{X}$ .

When  $\mathcal{X} = \mathbb{C}^n$ , we may alternatively write  $\mathbb{1}_n$ . When the space the identity operator acts on is clear from context, we may omit the subscript entirely.

In matrix form, the identity operator has ones on its diagonal and zeroes elsewhere, that is,  $\mathbb{1}(i, j) = e_i^* e_j = \delta(i, j)$ . Remarkably, this representation is independent of a particular basis choice.

**Lemma 1.2.11** (Resolution of the identity). *Let  $\mathcal{X}$  be a complex Euclidean space with  $|\mathcal{X}| = n$  and let  $\{x_1, \dots, x_n\} \subseteq \mathcal{X}$  be an orthonormal basis. The identity operator  $\mathbb{1}_{\mathcal{X}}$  may be written as*

$$\mathbb{1}_{\mathcal{X}} = \sum_{i=1}^n x_i x_i^* . \quad (1.53)$$

We leave the proof of Lemma 1.2.11 to the reader as Exercise 1.2.

**Definition 1.2.12.** Let  $\mathcal{X}$  be a complex Euclidean space. A square operator  $A \in L(\mathcal{X})$  is **invertible** when there exists an operator  $B \in L(\mathcal{X})$  satisfying

$$AB = BA = \mathbb{1}_{\mathcal{X}} . \quad (1.54)$$

*Remark.* The statements  $AB = \mathbb{1}_{\mathcal{X}}$  and  $BA = \mathbb{1}_{\mathcal{X}}$  are equivalent, so in practice just one suffices in demonstrating that  $A$  is invertible. The equivalence can be seen by noting that  $AB = \mathbb{1}_{\mathcal{X}}$  implies  $B = B(AB) = (BA)B$  and  $(\mathbb{1}_{\mathcal{X}} - BA)B = 0$ , and thus  $\mathbb{1}_{\mathcal{X}} - BA = 0$  since  $B$  is full rank.

We demonstrate several additional properties of invertibility in Lecture 1.4.

**Definition 1.2.13.** An **isometry** between complex Euclidean spaces  $\mathcal{X}$  and  $\mathcal{Y}$  is an operator  $A \in L(\mathcal{X}, \mathcal{Y})$  satisfying  $A^* A = \mathbb{1}_{\mathcal{X}}$ . The set of all such isometries is denoted by  $U(\mathcal{X}, \mathcal{Y})$ .

Isometries preserve Euclidean norms and inner products, meaning that for all  $u, v \in \mathcal{X}$ , an isometry  $A \in L(\mathcal{X}, \mathcal{Y})$  satisfies  $\|Au\|_2 = \|u\|_2$  and  $\langle Au, Av \rangle = \langle u, v \rangle$ . Either of these conditions may be used as equivalent definitions of isometries.

**Example.** Let  $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$ , meaning  $A \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} a \\ b \\ 0 \end{pmatrix}$  for all  $a, b \in \mathbb{C}$ .  $A$  is an isometry since

$$A^* = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad \text{and} \quad A^*A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbb{1}_2. \quad (1.55)$$

Note that  $A \in \mathbb{C}^{m \times n}$  may only be an isometry when  $m \geq n$ , meaning  $A$  is no wider than it is tall.

When restricting isometries to square operators via  $\mathcal{X} = \mathcal{Y}$ , we get the following class.

**Definition 1.2.14.** A **unitary** operator  $U \in L(\mathcal{X})$  is one that satisfies  $UU^* = U^*U = \mathbb{1}_{\mathcal{X}}$ . The set of all such unitaries acting on  $\mathcal{X}$  is denoted by  $U(\mathcal{X})$ .

Note that this implies any unitary  $U$  is invertible as in Definition 1.2.12, with  $U^{-1} = U^*$ . The definition additionally implies that unitary matrices are a subclass of normal matrices.

**Example.**  $U_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ ,  $U_2 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ , and  $U_3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$  are self-adjoint, and we may check that  $U_1^*U_1 = U_2^*U_2 = U_3^*U_3 = \mathbb{1}_2$ . Thus, each of  $U_1, U_2, U_3 \in U(\mathbb{C}^2)$  is unitary.

Some classes of operators have been omitted for now, despite being central to this course. They are the various types of definite operators, and we discuss them in Lecture 1.4.

### 1.2.3 Decompositions

Recall that if  $u \in \mathbb{C}^{m \times 1}$  is a column vector and  $v^* \in \mathbb{C}^{1 \times n}$  a row vector, then the product  $uv^*$  is an  $m$ -by- $n$  matrix. This motivates the following decomposition.

**Definition 1.2.15.** Let  $A \in L(\mathcal{X}, \mathcal{Y})$ . We write  $A$  in its **singular value decomposition** as

$$A = \sum_{i=1}^r \sigma_i u_i v_i^*, \quad (1.56)$$

where  $r = \min\{\dim(\mathcal{X}), \dim(\mathcal{Y})\}$ ,  $\sigma_i \geq 0$  for all  $i \in \{1, \dots, r\}$ , and  $\{u_1, \dots, u_r\} \subset \mathcal{Y}$ ,  $\{v_1, \dots, v_r\} \subset \mathcal{X}$  are orthonormal sets. The scalars  $\sigma_i$  are called **singular values**, and vectors  $u_i$  and  $v_i$  are called left and right **singular vectors**, respectively.

To emphasize that the singular values come from  $A$ , we may use the notation  $\sigma_i[A]$ . For convenience, we may arrange the singular values into a column vector  $\sigma[A] = (\sigma_1[A] \ \dots \ \sigma_r[A])^T$ .

**Definition 1.2.16.** Given  $A \in L(\mathcal{X})$ , if there exists  $\lambda \in \mathbb{C}$  and nonzero  $x \in \mathcal{X} \setminus \{0\}$  satisfying  $Ax = \lambda x$ , then  $\lambda$  is an **eigenvalue** of  $A$ , and  $x$  is an **eigenvector** of  $A$ . Without loss of generality, we can assume  $x$  is normalized.  $(\lambda, x)$  is an eigenvalue-eigenvector pair, or **eigenpair**, and therefore associated with each other.

If  $A$  is normal, then it has an orthonormal basis of eigenvectors. If  $A$  is Hermitian (which implies  $A$  is normal), then it has real eigenvalues, a fact that we prove below.

*Proof.* Consider an eigenpair  $(\lambda, x)$  of  $A \in \text{Herm}(\mathcal{X})$ . We have  $x^*Ax = x^*(\lambda x) = \lambda x^*x$  and  $x^*A^*x = (Ax)^*x = \bar{\lambda}x^*x$ . Thus,  $A = A^*$  implies  $\lambda = \bar{\lambda}$  (as  $x^*x = \|x\|_2^2 \neq 0$ ), and  $\lambda \in \mathbb{R}$ .  $\square$

To emphasize that the eigenvalues come from  $A$ , we may use the notation  $\lambda_i[A]$ . For convenience, we may arrange the eigenvalues into a column vector  $\lambda[A] = (\lambda_1[A] \ \dots \ \lambda_n[A])^T$ .

**Definition 1.2.17.** If  $A \in \mathbb{C}^{n \times n}$  is normal, then it has a **spectral decomposition** of the form

$$A = \sum_{i=1}^n \lambda_i x_i x_i^*, \quad (1.57)$$

where  $\{x_1, \dots, x_n\}$  is an orthonormal basis for  $\mathbb{C}^n$ . We can alternatively write  $A = UDU^*$ , where  $U$  is unitary and  $D$  is a diagonal matrix containing the eigenvalues of  $A$ .

When  $A$  is normal, we have  $\sigma_i[A] = |\lambda_i[A]|$  for all singular values.

*Example.* Let  $A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ . Recall that  $e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$  and  $e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ .

$$Ae_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = e_1 \quad (\text{so } \lambda_1 = +1), \quad (1.58a)$$

$$Ae_2 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = -\begin{pmatrix} 0 \\ 1 \end{pmatrix} = -e_2 \quad (\text{so } \lambda_2 = -1). \quad (1.58b)$$

We then have that the spectral decomposition of  $A$  is  $A = (+1)e_1e_1^* + (-1)e_2e_2^* = e_1e_1^* - e_2e_2^*$ .

## 1.2.4 Operator functions

The spectral decomposition provides a way of relating operators to scalars via their eigenvalues. One consequence of this is the ability to extend functions of the form  $f : S \subseteq \mathbb{C} \rightarrow \mathbb{C}$  to normal operators  $A \in \mathbb{C}^{n \times n}$  having  $\lambda_i[A] \in S$  for all  $i \in \{1, \dots, n\}$ . Specifically,

$$f(A) = \sum_{i=1}^n f(\lambda_i) x_i x_i^*, \quad (1.59)$$

meaning that we apply  $f$  to each eigenvalue separately while keeping the remaining structure of the spectral decomposition (1.57) intact. Notable examples include the exponent  $\exp(A)$ , logarithm  $\ln(A)$ , and various powers such as the inverse  $A^{-1}$ , square root  $A^{1/2} = \sqrt{A}$ , and square  $A^2$ .

Next, we consider functions that map operators to scalars.

**Definition 1.2.18.** The **trace** of a square matrix  $A \in \mathbb{C}^{n \times n}$  is defined in all the following ways.

$$\text{Tr}(A) = \sum_{i=1}^n A(i, i) = \sum_{i=1}^n x_i^* A x_i = \sum_{i=1}^n \lambda_i[A], \quad (1.60)$$

for any orthonormal basis  $\{x_1, \dots, x_n\}$ . That is, the trace is the sum of the diagonal entries (with respect to any basis), and also the sum of the eigenvalues.

We may similarly define other functions that aggregate the eigenvalues of  $A$ . For instance, the determinant  $\det(A) = \prod_{i=1}^n \lambda_i[A]$  is the product of eigenvalues, and  $\text{rank}(A)$  is the number of nonzero eigenvalues (counting multiplicities). The latter plays a role in the singular value decomposition, Definition 1.2.15; we may remove any terms where  $\sigma_i = 0$  to instead obtain  $r = \text{rank}(A)$  and  $\sigma_i > 0$  for all  $i \in \{1, \dots, r\}$ .

**Lemma 1.2.19** (Cyclic property). *Suppose  $A \in L(\mathcal{Z}, \mathcal{X})$ ,  $B \in L(\mathcal{Y}, \mathcal{Z})$ , and  $C \in L(\mathcal{X}, \mathcal{Y})$ , which guarantees  $ABC$  is a square operator. Then,  $\text{Tr}(ABC) = \text{Tr}(CAB)$ .*

*Example.*  $\text{Tr}(Axx^*) = \text{Tr}(x^*Ax)$ . Since the quantity inside the trace is a scalar, this equals  $x^*Ax$ .

Note that  $\text{Tr}(ABC)$  may not necessarily equal  $\text{Tr}(CBA)$ .

*Example.* Let  $A, B, C \in \mathbb{C}^{2 \times 2}$  with  $A = e_1 e_2^*$  and  $C = e_2 e_1^*$ . Then,

$$\text{Tr}(ABC) = \text{Tr}(e_1 e_2^* B e_2 e_1^*) = \text{Tr}((e_1^* e_1) e_2^* B e_2) = \text{Tr}(e_2^* B e_2) = B(2, 2), \quad (1.61a)$$

$$\text{Tr}(CBA) = \text{Tr}(e_2 e_1^* B e_1 e_2^*) = \text{Tr}((e_2^* e_2) e_1^* B e_1) = \text{Tr}(e_1^* B e_1) = B(1, 1). \quad (1.61b)$$

As in general  $B(1, 1) \neq B(2, 2)$ , it follows that  $\text{Tr}(ABC) \neq \text{Tr}(CBA)$ .

We now define an inner product on matrices. This is the last step necessary to claim that the space of operators  $L(\mathcal{X}, \mathcal{Y})$  is a complex Euclidean space, and it allows us to induce a matrix norm and define matrix bases shortly.

**Definition 1.2.20.** Let  $A, B \in \mathbb{C}^{m \times n}$ . The **Hilbert-Schmidt inner product** can be defined via

$$\langle A, B \rangle = \sum_{i=1}^m \sum_{j=1}^n \overline{A(i, j)} B(i, j) = \text{Tr}(A^* B). \quad (1.62)$$

This extends to any  $A, B \in L(\mathcal{X}, \mathcal{Y})$ .

*Example.* Define the following matrices:

$$\mathbb{1}_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \in \mathbb{C}^{3 \times 3}, \quad A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 1 & 1 & 1 \end{pmatrix} \in \mathbb{C}^{3 \times 3}. \quad (1.63)$$

Their Hilbert-Schmidt inner product is then

$$\langle \mathbb{1}_3, A \rangle = \sum_{i=1}^3 \sum_{j=1}^3 \overline{\mathbb{1}_3(i, j)} A(i, j) = \overline{1} \cdot 1 + \overline{1} \cdot 3 + \overline{1} \cdot 1 = 1 + 3 + 1 = 5 = \text{Tr}(A). \quad (1.64)$$

Similarly,  $\langle \mathbb{1}_n, A \rangle = \text{Tr}(\mathbb{1}_n^* A) = \text{Tr}(A)$  for any  $A \in \mathbb{C}^{n \times n}$ , which is often used often in this course.

In Exercise 1.3, we ask the reader to show that the Hilbert-Schmidt inner product of Hermitian matrices is real.

Recall that for a matrix  $A \in \mathbb{C}^{m \times n}$ , the singular value decomposition yielded a vector of singular values  $\sigma \in \mathbb{R}_+^r$  with  $r = \min\{m, n\}$ . One possible approach in extending vector norms to matrices

is then simply by applying those norms on the vector of singular values. These are called **Schatten norms**, and some receive special names and notations:

$$\|A\|_1 = \|\sigma[A]\|_1 = \sum_{i=1}^r \sigma_i[A], \quad \text{Trace norm } \|\cdot\|_{\text{Tr}}, \quad (1.65)$$

$$\|A\|_2 = \|\sigma[A]\|_2 = \sqrt{\sum_{i=1}^r \sigma_i[A]^2}, \quad \text{Frobenius norm } \|\cdot\|_{\text{F}}, \quad (1.66)$$

$$\|A\|_\infty = \|\sigma[A]\|_\infty = \max_{i \in \{1, \dots, r\}} \sigma_i[A], \quad \text{Operator norm } \|\cdot\|_{\text{op}}. \quad (1.67)$$

For these norms in particular, we have alternate expressions that do not rely explicitly on  $\sigma_i[A]$ .

$$\|A\|_1 = \text{Tr}(\sqrt{A^*A}), \quad \|A\|_2 = \sqrt{\langle A, A \rangle}, \quad \|A\|_\infty = \max\{\|Au\|_2 : u \in \mathcal{X}, \|u\|_2 \leq 1\}. \quad (1.68)$$

All of the above are again special cases of the  $p$ -norm. For any  $p \in [1, \infty]$ , we have

$$\|A\|_p = \|\sigma[A]\|_p = \left( \sum_{i=1}^r \sigma_i[A]^p \right)^{1/p}. \quad (1.69)$$

Recall that normal operators have  $\sigma_i[A] = |\lambda_i[A]|$ . For normal operators, we may alternatively define the above norms in terms of the eigenvalues of  $A$ . This involves replacing  $\sigma[A]$  with  $\lambda[A]$ ,  $\sigma_i[A]$  with  $|\lambda_i[A]|$ , and  $r$  with  $n$  in (1.65) to (1.67) and (1.69). For any normal  $A \in \mathbb{C}^{n \times n}$  and  $p \in [1, \infty]$ ,

$$\|A\|_1 = \|\lambda[A]\|_1 = \sum_{i=1}^n |\lambda_i[A]|, \quad \|A\|_\infty = \|\lambda[A]\|_\infty = \max_{i \in \{1, \dots, n\}} |\lambda_i[A]|, \quad (1.70)$$

$$\|A\|_2 = \|\lambda[A]\|_2 = \sqrt{\sum_{i=1}^n |\lambda_i[A]|^2}, \quad \|A\|_p = \|\lambda[A]\|_p = \left( \sum_{i=1}^n |\lambda_i[A]|^p \right)^{1/p}. \quad (1.71)$$

**Example.** Let  $A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ , which is normal and has eigenvalues  $\pm 1$ . Thus,

$$\sigma[A] = \begin{pmatrix} \sigma_1[A] \\ \sigma_2[A] \end{pmatrix} = \begin{pmatrix} |\lambda_1[A]| \\ |\lambda_2[A]| \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \quad (1.72)$$

We may then compute several Schatten norms of  $A$  as follows:

$$\|A\|_1 = \|\sigma\|_1 = 1 + 1 = 2, \quad (1.73)$$

$$\|A\|_2 = \|\sigma\|_2 = \sqrt{1^2 + 1^2} = \sqrt{2}, \quad (1.74)$$

$$\|A\|_\infty = \|\sigma\|_\infty = \max\{1, 1\} = 1. \quad (1.75)$$

Note that we have  $\|A\|_\infty \leq \|A\|_2 \leq \|A\|_1$  again. In general,  $\|A\|_q \leq \|A\|_p$  for  $q \geq p$ .

The definition for **matrix bases** is nearly identical to that of vector bases in Definition 1.1.7. Aside from the set of vectors being a subset of  $\mathbb{C}^{m \times n}$ , the biggest change is that now we use the Hilbert-Schmidt inner product from Definition 1.2.20.

*Example.* The following sets are both bases for  $\mathbb{C}^{2 \times 2}$ :

$$\left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right\}, \quad (1.76a)$$

$$\left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1/\sqrt{2} \\ 1/\sqrt{2} & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1/\sqrt{2} \\ -1/\sqrt{2} & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right\}. \quad (1.76b)$$

The former is an example of the **standard basis** on  $\mathbb{C}^{2 \times 2}$ . We use  $E_{i,j} := e_i e_j^*$  to denote the matrix having a 1 entry in the  $i$ th row and  $j$ th column and 0 elsewhere. The first example may then be written as  $\{E_{1,1}, E_{1,2}, E_{2,1}, E_{2,2}\} \subset \mathbb{C}^{2 \times 2}$ .

### 1.3 Superoperators

By now, we have established the space of operators  $L(\mathcal{X})$  as a complex Euclidean space. Specifically, they have proper notions of addition and scalar multiplication, and inner products. Instead of mapping vectors to vectors as we did with operators, we may take a step higher and map operators to operators. Mappings of this sort are also called *superoperators*.

We start by considering mappings of the form  $\Phi : L(\mathcal{X}) \rightarrow L(\mathcal{Y})$ , that are *linear* in the same sense of Definition 1.2.1. The space of all such linear maps is denoted by  $T(\mathcal{X}, \mathcal{Y})$ , or simply  $T(\mathcal{X})$  when  $\mathcal{X} = \mathcal{Y}$ . Much like operators, superoperators form a complex Euclidean space, with addition and scalar multiplication defined as follows:

$$(\Phi + \Psi)(X) = \Phi(X) + \Psi(X), \quad (1.77)$$

$$(k\Phi)(X) = k\Phi(X). \quad (1.78)$$

Here,  $k \in \mathbb{C}$ ,  $X \in L(\mathcal{X})$ , and  $\Phi, \Psi, \Phi + \Psi, k\Phi \in T(\mathcal{X}, \mathcal{Y})$ . While there exists a valid notion of inner products over superoperators, we omit defining it here.

Combining superoperators via a tensor product yields a larger superoperator, with its constituents acting on their respective disjoint parts of the input. The definition essentially identical to the tensor product of operators (Definition 1.2.5), elevated by one level of abstraction.

**Definition 1.3.1.** Let  $\mathcal{X}_\Phi, \mathcal{X}_\Psi, \mathcal{Y}_\Phi$ , and  $\mathcal{Y}_\Psi$  be complex Euclidean spaces. The **tensor product** of  $\Phi \in T(\mathcal{X}_\Phi, \mathcal{Y}_\Phi)$  and  $\Psi \in T(\mathcal{X}_\Psi, \mathcal{Y}_\Psi)$  is the unique linear operator

$$\Phi \otimes \Psi \in L(\mathcal{X}_\Phi \otimes \mathcal{X}_\Psi, \mathcal{Y}_\Phi \otimes \mathcal{Y}_\Psi), \quad (1.79)$$

satisfying for all  $A \in L(\mathcal{X}_\Phi)$  and  $B \in L(\mathcal{X}_\Psi)$

$$(\Phi \otimes \Psi)(A \otimes B) = \Phi(A) \otimes \Psi(B). \quad (1.80)$$

Similar to the identity operator, we define the **identity map**  $\mathbb{1}_{L(\mathcal{X})} \in T(\mathcal{X})$  as the unique map satisfying  $\mathbb{1}_{L(\mathcal{X})}(X) = X$  for all  $X \in L(\mathcal{X})$ . This map leaves its input completely unchanged, and combining it with other maps yields operations that act on only part of their input.

For instance, consider the trace operation  $\text{Tr} : L(\mathcal{X}) \rightarrow \mathbb{C}$ . Identifying  $1 \times 1$  operators as scalars, i.e.,  $L(\mathbb{C}) = \mathbb{C}$ , lets us write  $\text{Tr} \in T(\mathcal{X}, \mathbb{C})$ . If we only want to trace out the  $\mathcal{X}$  part of operators over a larger space  $\mathcal{X} \otimes \mathcal{Y}$ , we can use the **partial trace**

$$\text{Tr}_{\mathcal{X}} := \text{Tr} \otimes \mathbb{1}_{L(\mathcal{Y})}, \quad (1.81)$$

which is the unique map satisfying

$$\text{Tr}_{\mathcal{X}}(X \otimes Y) = \text{Tr}(X) \cdot Y \quad (1.82)$$

for all  $X \in \mathcal{X}$  and  $Y \in \mathcal{Y}$ .

*Example.* Partial traces of arbitrary matrices may be computed by splitting them into blocks.

$$\text{Tr}_{\mathcal{X}} \left( \begin{pmatrix} \boxed{1 \ 2} & \boxed{3 \ 4} \\ \boxed{5 \ 6} & \boxed{7 \ 3} \\ \boxed{8 \ 9} & \boxed{2 \ 5} \\ \boxed{0 \ 1} & \boxed{4 \ 6} \end{pmatrix} \right) = \begin{pmatrix} \boxed{3 \ 7} \\ \boxed{9 \ 12} \end{pmatrix}, \quad \text{Tr}_{\mathcal{Y}} \left( \begin{pmatrix} \boxed{1 \ 2} & \boxed{3 \ 4} \\ \boxed{5 \ 6} & \boxed{7 \ 3} \\ \boxed{8 \ 9} & \boxed{2 \ 5} \\ \boxed{0 \ 1} & \boxed{4 \ 6} \end{pmatrix} \right) = \begin{pmatrix} \boxed{7 \ 6} \\ \boxed{9 \ 8} \end{pmatrix}. \quad (1.83)$$

Note that the traces of all matrices match;  $\text{Tr}(\text{Tr}_{\mathcal{X}}(A)) = \text{Tr}(\text{Tr}_{\mathcal{Y}}(A)) = \text{Tr}(A)$  for all  $A \in L(\mathcal{X} \otimes \mathcal{Y})$ .

Sometimes it becomes necessary to consider subsets of maps that preserve certain properties. For instance, a map  $\Phi \in T(\mathcal{X}, \mathcal{Y})$  is **Hermiticity-preserving** if  $\Phi : \text{Herm}(\mathcal{X}) \rightarrow \text{Herm}(\mathcal{Y})$ . That is, given any Hermitian  $X \in \text{Herm}(\mathcal{X})$ , the output  $\Phi(X) \in \text{Herm}(\mathcal{Y})$  is also Hermitian.

*Example.* The following linear maps  $\Phi$  are Hermiticity-preserving:

$$\begin{aligned} \Phi(X) &= \langle A, X \rangle, & \Phi \left( \begin{pmatrix} a & b \\ c & d \end{pmatrix} \right) &= \begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix}; & \Phi(X) &= \text{Tr}(X); & \Phi(X) &= BXB^*. \end{aligned} \quad (1.84)$$

We almost always work with Hermitian operators, so we often require the maps we consider to be Hermiticity-preserving. Similarly,  $\Phi \in T(\mathcal{X}, \mathcal{Y})$  is **trace-preserving** if the trace of its outputs always matches the trace of its inputs, i.e.,  $\text{Tr}(X) = \text{Tr}(\Phi(X))$  for all  $X \in \text{Herm}(\mathcal{X})$ .

### 1.3.1 Representations

Similar to how operators are equivalent to matrices, superoperators can be represented as rank-four tensors. Since this is more dimensions than an ordinary piece of paper is ought to have, we typically look for ways to relate superoperators back to operators so that we can write them down. Methods of this sort enlarge or split up the operators, one possible approach being as follows.

**Definition 1.3.2.** Let  $\mathcal{X}$  and  $\mathcal{Y}$  be complex Euclidean spaces. A map  $\Phi \in T(\mathcal{X}, \mathcal{Y})$  may be defined in its **Stinespring representation** by a third complex Euclidean space  $\mathcal{Z}$  and operators  $A, B \in L(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Z})$ . When acting on any  $X \in L(\mathcal{X})$ , the map  $\Phi$  then produces

$$\Phi(X) = \text{Tr}_{\mathcal{Z}}(AXB^*). \quad (1.85)$$

Every map  $\Phi \in T(\mathcal{X}, \mathcal{Y})$  has a Stinespring representation, which is not necessarily unique.

## 1.4 Definite operators

In defining various classes of operators in Lecture 1.2, we have skipped over several classes that are fundamental to semidefinite programming. In this lecture, we aim to define these classes and prove a few of their properties.

A goal of these classes is in part to enable comparing operators via defining notions of positive, negative, nonpositive, and nonnegative operators. For a Hermitian operator  $A \in \text{Herm}(\mathcal{X})$  this can be done via its *quadratic form*  $x^*Ax$ , though we must additionally ensure that the sign of this form is consistent. Operators that satisfy this are known as *(semi)definite* and have a bounded quadratic form, a property that makes them useful for certain optimization problems.

**Definition 1.4.1.** A **positive semidefinite (PSD)** operator acting on a complex Euclidean space  $\mathcal{X}$  is a Hermitian operator  $A \in \text{Herm}(\mathcal{X})$  that satisfies  $x^*Ax \geq 0$  for all  $x \in \mathcal{X}$ . The set of all such operators is denoted by  $\text{Pos}(\mathcal{X})$ , and we may write  $A \in \text{Pos}(\mathcal{X})$  as  $A \succcurlyeq 0$ .

*Example.*  $A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \succcurlyeq 0$  since it is Hermitian and  $x^*Ax = |x(1)|^2 + 2|x(2)|^2 \geq 0$  for all  $x \in \mathbb{C}^2$ .

*Example.*  $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \not\succcurlyeq 0$  since  $x^*Ax = -2 \not\geq 0$  for  $x = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ .

The prefix *semi-* above indicates that the quadratic form is permitted to be zero. We now define a class for which this is not the case, taking care to avoid evaluating with the zero vector.

**Definition 1.4.2.** A **positive definite (PD)** operator acting on a complex Euclidean space  $\mathcal{X}$  is a Hermitian operator  $A \in \text{Herm}(\mathcal{X})$  that satisfies  $x^*Ax > 0$  for all  $x \in \mathcal{X} \setminus \{0\}$ . The set of all such operators is denoted by  $\text{Pd}(\mathcal{X})$ , and we may write  $A \in \text{Pd}(\mathcal{X})$  as  $A \succ 0$ .

*Example.*  $A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \succ 0$  since it is Hermitian and  $x^*Ax = |x(1)|^2 + 2|x(2)|^2 > 0$  for all  $x \neq 0$ .

*Example.*  $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \not\succ 0$  since  $x^*Ax = 0 \not> 0$  for  $x = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ .

Note that in the last example,  $A \not\succ 0$  but  $A \succcurlyeq 0$ . This implies  $\text{Pos}(\mathcal{X}) \not\subseteq \text{Pd}(\mathcal{X})$ . On the other hand, we have  $\text{Pd}(\mathcal{X}) \subseteq \text{Pos}(\mathcal{X})$ , which can be seen from the above definitions. In particular,  $x^*Ax > 0$  implies  $x^*Ax \geq 0$ , and when  $x = 0$  we have  $x^*Ax = 0$ .

*Remark.* We have restricted positive semidefinite and positive definite operators in the above definitions to necessarily be Hermitian, which ensures that  $x^*Ax$  is real. There exist extensions of these notions to non-Hermitian operators, but we do not consider them in this course.

The notation  $A \succcurlyeq 0$  and  $A \succ 0$  suggests treating such notions as inequalities. We may add operators to both sides, for instance by writing  $A - B \succcurlyeq 0$  as  $A \succcurlyeq B$ , or  $-A \succcurlyeq 0$  as  $0 \succcurlyeq A$ . The latter form provides a means of defining the negative versions of the above classes as follows.

**Definition 1.4.3.** A **negative semidefinite** operator is the negative of some positive semidefinite operator. We may write  $-A \in \text{Pos}(\mathcal{X})$  as  $A \preccurlyeq 0$ .

**Definition 1.4.4.** A **negative definite** operator is the negative of some positive definite operator. We may write  $-A \in \text{Pd}(\mathcal{X})$  as  $A \prec 0$ .

### 1.4.1 Properties of definite operators

We next establish some conditions as equivalent to an operator being positive semidefinite. This list is not exhaustive, with many other conditions omitted for brevity.

**Lemma 1.4.5** (Properties of positive semidefinite operators). *Let  $A \in L(\mathcal{X})$  for some complex Euclidean space  $\mathcal{X}$ . The following statements are equivalent.*

1.  $A$  is positive semidefinite, i.e.,  $A \in \text{Pos}(\mathcal{X})$ .
2.  $A$  is Hermitian and the eigenvalues of  $A$  are nonnegative, i.e.,  $A \in \text{Herm}(\mathcal{X})$  and  $\lambda_i[A] \geq 0$  for all  $i \in \{1, \dots, n\}$  with  $n = |\mathcal{X}|$ .
3. There exists  $B \in L(\mathcal{X}, \mathcal{Y})$  for some choice of a complex Euclidean space  $\mathcal{Y}$  such that  $A = B^*B$ .
4. There exists a collection of vectors  $\{u_1, \dots, u_n\} \subseteq \mathcal{Y}$  with  $n = |\mathcal{X}|$  for some choice of a complex Euclidean space  $\mathcal{Y}$  such that  $A(i, j) = \langle u_i, u_j \rangle$ .

*Proof.* We show equivalence of the above statements via the following cycle of implications.

$$(1) \Rightarrow (2) \Rightarrow (3) \Rightarrow (4) \Rightarrow (1) \quad (1.86)$$

(1)  $\Rightarrow$  (2): Assume  $A \in \text{Pos}(\mathcal{X})$ , which implies  $A \in \text{Herm}(\mathcal{X})$ . Let  $(\lambda_i, v_i)$  be an eigenpair of  $A$  for any  $i \in \{1, \dots, n\}$ , with  $n = |\mathcal{X}|$  and  $v_i$  normalized. We then have

$$v_i^* A v_i = v_i^* (\lambda_i v_i) = \lambda_i v_i^* v_i = \lambda_i \|v_i\|_2^2 = \lambda_i. \quad (1.87)$$

$A \in \text{Pos}(\mathcal{X})$  implies  $v_i^* A v_i \geq 0$  and hence  $\lambda_i \geq 0$ . Thus, statement 1 implies statement 2.

(2)  $\Rightarrow$  (3): Consider the spectral decomposition of  $A$  and define  $B \in L(\mathcal{X})$  as

$$A = \sum_{i=1}^n \lambda_i v_i v_i^*, \quad B = \sum_{i=1}^n \sqrt{\lambda_i} v_i v_i^*, \quad (1.88)$$

where we may take  $\sqrt{\lambda_i}$  since by statement 2 we have  $\lambda_i \geq 0$ . We may then compute

$$B^* B = \left( \sum_{i=1}^n \sqrt{\lambda_i} v_i v_i^* \right) \left( \sum_{j=1}^n \sqrt{\lambda_j} v_j v_j^* \right) = \sum_{i=1}^n \sum_{j=1}^n \sqrt{\lambda_i} \sqrt{\lambda_j} v_i v_i^* v_j v_j^* = \sum_{i=1}^n \lambda_i v_i v_i^* = A, \quad (1.89)$$

where we have  $\overline{\sqrt{\lambda_i}} = \sqrt{\lambda_i}$  since  $A \in \text{Herm}(\mathcal{X})$  and  $v_i^* v_j = \delta(i, j)$  since eigenvectors of normal operators are orthonormal. Thus, statement 2 implies statement 3.

(3)  $\Rightarrow$  (4): Assume there exists  $B \in L(\mathcal{X}, \mathcal{Y})$  such that  $A = B^* B$ . Recall that  $\{e_1, \dots, e_n\} \subseteq \mathcal{X}$  with  $n = |\mathcal{X}|$  is the standard basis on  $\mathcal{X}$ . Let  $\{u_1, \dots, u_n\} \subseteq \mathcal{Y}$  be the columns of  $B$ , i.e.,

$$B = \sum_{i=1}^n u_i e_i^*. \quad (1.90)$$

We may then compute the formula for the element  $A(i, j)$  as follows.

$$A(i, j) = e_i^* A e_j = e_i^* B^* B e_j = e_i^* \left( \sum_{k=1}^n e_k u_k^* \right) \left( \sum_{l=1}^n u_l e_l^* \right) e_j = u_i^* u_j = \langle u_i, u_j \rangle, \quad (1.91)$$

as  $e_i^* e_k$  and  $e_l^* e_j$  are nonzero only if  $k = i$  and  $l = j$ . Thus, statement 3 implies statement 4.

(4)  $\Rightarrow$  (1): Assume there exists a set of vectors  $\{u_1, \dots, u_n\} \subseteq \mathcal{Y}$  with  $n = |\mathcal{X}|$  such that  $A(i, j) = \langle u_i, u_j \rangle = u_i^* u_j$  for all  $i, j \in \{1, \dots, n\}$ . Let  $x \in \mathcal{X}$ , we may then compute

$$x^* A x = \sum_{i=1}^n \sum_{j=1}^n x^*(i) A(i, j) x(j) = \sum_{i=1}^n \sum_{j=1}^n x^*(i) u_i^* u_j x(j) \quad (1.92a)$$

$$= \left( \sum_{i=1}^n u_i x(i) \right)^* \left( \sum_{j=1}^n u_j x(j) \right) = \left| \sum_{i=1}^n u_i x(i) \right|^2 \geq 0, \quad (1.92b)$$

and hence  $A \in \text{Pos}(\mathcal{X})$ . Thus, statement 4 implies statement 1.

The implication cycle (1.86) is now complete, and hence statements 1 to 4 are equivalent.  $\square$

Note that  $B$  in (1.88) is obtained from  $A$  by applying the square root to its eigenvalues. This process is thus an operator function as shown in (1.59), and we may alternatively write  $B = \sqrt{A}$ .  $B$  is then a square root of  $A$ , with  $B^2 = A$  implied either by reasoning similar to (1.89) or by noting that  $f(x) = (\sqrt{x})^2 = x$ . We make use of this fact to considerably simplify the proof of the analogous implication in Lemma 1.4.8, which is otherwise nearly identical.

While an operator  $A \in \text{Pos}(\mathcal{X})$  may have many square roots, the form given in (1.88) and denoted by  $\sqrt{A}$  is the unique square root that is also positive semidefinite.

**Example.** Consider the following positive definite operator  $A$  and its spectral decomposition

$$A = \frac{1}{2} \begin{bmatrix} 5 & 3 \\ 3 & 5 \end{bmatrix} = 4v_1 v_1^* + 1v_2 v_2^*, \quad \text{where} \quad v_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad v_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}. \quad (1.93)$$

The unique positive (semi)definite square root of  $A$  is as follows.

$$\sqrt{A} = \sqrt{4}v_1 v_1^* + \sqrt{1}v_2 v_2^* = 2v_1 v_1^* + 1v_2 v_2^* = \frac{1}{2} \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}. \quad (1.94)$$

We next analyze a similar set of statements for positive definite matrices. Several of them now in addition require an operator to be invertible, a notion first defined in Definition 1.2.12. To complete the relevant proofs, we first relate invertibility to operator decompositions.

**Lemma 1.4.6** (Invertibility from singular values).  *$A \in L(\mathbb{C}^n)$  being invertible is equivalent to its every singular value being positive. The singular value decompositions of  $A$  and its inverse are*

$$A = \sum_{i=1}^n \sigma_i x_i y_i^*, \quad A^{-1} = \sum_{i=1}^n \sigma_i^{-1} y_i x_i^*. \quad (1.95)$$

**Lemma 1.4.7** (Invertibility from eigenvalues). *A normal  $A \in L(\mathbb{C}^n)$  being invertible is equivalent to its every eigenvalue being non-zero. The spectral decompositions of  $A$  and its inverse are*

$$A = \sum_{i=1}^n \lambda_i v_i v_i^* , \quad A^{-1} = \sum_{i=1}^n \lambda_i^{-1} v_i v_i^* . \quad (1.96)$$

The proofs of Lemmas 1.4.6 and 1.4.7 are left to the reader as Exercise 1.4.

We now present a non-exhaustive list of properties of positive definite operators.

**Lemma 1.4.8** (Properties of positive definite operators). *Let  $A \in L(\mathcal{X})$  for some complex Euclidean space  $\mathcal{X}$ . The following statements are equivalent.*

1.  *$A$  is positive definite, i.e.,  $A \in \text{Pd}(\mathcal{X})$ .*
2.  *$A$  is Hermitian and the eigenvalues of  $A$  are positive, i.e.,  $A \in \text{Herm}(\mathcal{X})$  and  $\lambda_i[A] > 0$  for all  $i \in \{1, \dots, n\}$  with  $n = |\mathcal{X}|$ .*
3. *There exists an invertible  $B \in L(\mathcal{X})$  such that  $A = B^*B$ .*
4.  *$A \in \text{Pos}(\mathcal{X})$  and  $A$  is invertible.*

*Proof.* We show equivalence of the above statements via the following cycle of implications.

$$(1) \implies (2) \implies (3) \implies (4) \implies (2) \implies (1) \quad (1.97)$$

(1)  $\Rightarrow$  (2): Assume  $A \in \text{Pd}(\mathcal{X})$ , which implies  $A \in \text{Herm}(\mathcal{X})$ . Since  $\text{Pd}(\mathcal{X}) \subseteq \text{Pos}(\mathcal{X})$ , we have  $A \in \text{Pos}(\mathcal{X})$  and thus  $\lambda_i[A] \geq 0$  by Lemma 1.4.5. To prove  $\lambda_i[A] \neq 0$ , we assume there exists an eigenpair  $(\lambda_i, v_i)$  for some  $i \in \{1, \dots, n\}$  with  $n = |\mathcal{X}|$  such that  $\lambda_i = 0$ , and then aim to arrive at a contradiction. We then have

$$v_i^* A v_i = v_i^* (\lambda_i v_i) = v_i^* (0 \cdot v_i) = 0 . \quad (1.98)$$

This is a contradiction since  $A \in \text{Pd}(\mathcal{X})$  implies  $x^* A x > 0$  for any nonzero  $x \in \mathcal{X} \setminus \{0\}$ , and we just showed that  $v_i^* A v_i = 0$  for  $v_i \neq 0$ . Thus, statement 1 implies statement 2.

(2)  $\Rightarrow$  (3): Assume that  $\lambda_i > 0$  for all  $i \in \{1, \dots, n\}$  with  $n = |\mathcal{X}|$ . Define  $B = \sqrt{A}$ , which is then invertible by Lemma 1.4.7, and Hermitian assuming  $A \in \text{Herm}(\mathcal{X})$ . We have

$$A = B^2 = B^* B , \quad (1.99)$$

and thus statement 2 implies statement 3.

(3)  $\Rightarrow$  (4): Assume there exists an invertible  $B \in L(\mathcal{X})$  such that  $A = B^*B$ . By Lemma 1.4.5, we have  $A \in \text{Pos}(\mathcal{X})$ . It remains to prove that  $A$  is invertible. Since  $B$  is invertible, there exists  $C \in L(\mathcal{X})$  such that  $BC = CB = \mathbb{1}_{\mathcal{X}}$ . Similarly, we have

$$C^* B^* = B^* C^* = \mathbb{1}_{\mathcal{X}}^* = \mathbb{1}_{\mathcal{X}} . \quad (1.100)$$

We may now show that  $CC^*$  is the inverse of  $A$  by taking their product.

$$A(CC^*) = B^*(BC)C^* = B^*C^* = \mathbb{1}_{\mathcal{X}}. \quad (1.101)$$

This implies  $A$  is invertible, and thus statement 3 implies statement 4.

(4)  $\Rightarrow$  (2): By Lemma 1.4.5,  $A \in \text{Pos}(\mathcal{X})$  implies  $\lambda_i[A] \geq 0$ . By Lemma 1.4.7,  $A$  being invertible implies  $\lambda_i[A] \neq 0$ . Thus,  $A \in \text{Herm}(\mathcal{X})$  and  $\lambda_i[A] > 0$ , and hence statement 4 implies statement 2.

(2)  $\Rightarrow$  (1): Fix an arbitrary nonzero vector  $x \in \mathcal{X} \setminus \{0\}$ . Let  $(\lambda_i, v_i)$  for  $i \in \{1, \dots, n\}$  with  $n = |\mathcal{X}|$  be the eigenpairs of  $A$ . Since  $\{v_1, \dots, v_n\} \subseteq \mathcal{X}$  forms a basis, we may write

$$x = \sum_{i=1}^n \alpha_i v_i, \quad \exists k \in \{1, \dots, n\}, \alpha_k \neq 0. \quad (1.102)$$

Noting that statement 2 gives  $\lambda_i[A] > 0$  and setting  $k$  as in the above, we have

$$x^*Ax = \left( \sum_{i=1}^n \bar{\alpha}_i v_i^* \right) A \left( \sum_{j=1}^n \alpha_j v_j \right) = \sum_{i=1}^n \sum_{j=1}^n \bar{\alpha}_i \alpha_j v_i^* (A v_j) = \sum_{i=1}^n \sum_{j=1}^n \lambda_j \bar{\alpha}_i \alpha_j v_i^* v_j \quad (1.103a)$$

$$= \sum_{i=1}^n \lambda_i |\alpha_i|^2 \geq \lambda_k |\alpha_k|^2 > 0, \quad (1.103b)$$

since  $A \in \text{Herm}(\mathcal{X})$  implies it is normal and hence its eigenvectors are orthonormal. Thus, statement 2 implies statement 1.

The implication cycle (1.97) is now complete, and hence statements 1 to 4 are equivalent.  $\square$

We make frequent use of Lemmas 1.4.5 and 1.4.8 in the remainder of this course due to the significance of positive (semi)definite operators to semidefinite programming.

## 1.5 Real analysis

Optimization problems involve finding the best value of a function over a set. To better understand properties of these two, we now move onto relevant ideas from analysis such as openness, boundedness, limits of sequences, and continuity. We explore these in a general way that is mostly distinct from the linear algebra covered thus far, later intertwining them by making the sets contain objects such as operators and superoperators.

### 1.5.1 Properties of sets

A set is an unordered collection of distinct things. We use the symbol ' $\in$ ' to indicate that something is an element of a set. For a set  $A$ , we have for every  $x$  either  $x \in A$  or  $x \notin A$ . We typically use curly braces to indicate sets; for instance, a set containing only  $x$ ,  $y$ , and  $z$  as its elements is  $\{x, y, z\}$ . To select elements from a set  $A$  that satisfy some property  $P$ , we use the notation  $\{x \in A : P(x)\}$ . Every element of a set  $B$  also being an element of a set  $A$  means that  $B$  is a subset of  $A$ , i.e.,  $B \subseteq A$ . For any two sets  $A$  and  $B$ , elements contained in both sets form their intersection  $A \cap B$ , and elements contained in at least one form their union  $A \cup B$ .

Recall that any complex Euclidean space  $\mathcal{X}$  defines an inner product, which induces a norm. This gives a notion of distance, thus letting us define the set of all points in  $\mathcal{X}$  “close” to some  $x \in \mathcal{X}$ .

**Definition 1.5.1.** A **ball** of radius  $r > 0$  around  $x \in \mathcal{X}$  is defined as

$$B_r(x) := \{y \in \mathcal{X} : \|x - y\| < r\}. \quad (1.104)$$

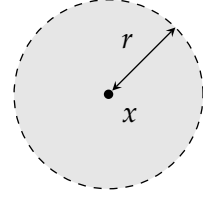


Figure 1.1:  $B_r(x)$ .

A visual rendition of a ball in  $\mathbb{C}^2$  can be seen in Figure 1.1. The dashed line at the edge of the set indicates that points with  $\|x - y\| = r$  do not belong to the ball. This is just one example of what is called an open set, which we define below.

**Definition 1.5.2.** For all  $x \in S$  in an **open** set  $S$ , there exists  $r > 0$  such that  $B_r(x) \subseteq S$ .

That is, every point  $x$  has a small neighborhood around it that falls within the set.

*Example.*  $(a, b)$  is open. For every  $x$  such that  $a < x < b$ , let  $r = \min\{x - a, b - x\}$ .

What the definition of an open set prohibits are points on its “edge”, where any step away from the set—no matter how small—lands you outside of it. As the above example shows, the set  $(a, b)$  has no such points; for any  $x$  that we pick, there are always points closer to the boundary than  $x$ . We can similarly apply this principle to sets in higher dimensions.

*Example.*  $\{(x, y) : x^2 + y^2 < 1\}$  is open. For any element  $(x, y)$  of this set, let  $r = 1 - \sqrt{x^2 + y^2}$ .

If we switch around the elements that are in and not in a set, we get the following definition.

**Definition 1.5.3.** The **complement** of a set  $S \subseteq \mathcal{X}$  is  $S^c := \{x \in \mathcal{X} : x \notin S\}$ .

This lets us define closed sets through the openness of their complement.

**Definition 1.5.4.** A **closed** set  $S$  is one that has an open complement  $S^c$ .

Closed sets must thus contain their boundary, as their open set complements exclude it.

*Example.*  $[a, b]$  is closed since its complement  $(-\infty, a) \cup (b, +\infty)$  is open. We let  $r = a - x$  when  $x < a$ , and  $r = x - b$  when  $x > b$ .

Note that  $[a, b]$  is also not open, since the points  $a$  and  $b$  have no balls around them that are in the set. Yet, the notions of openness and closeness are not opposites, as sets can be neither or both.

*Example.*  $[a, b)$  is neither open nor closed. There exists no ball centered at  $a$  contained in  $[a, b)$ , and there exists no ball centered at  $b$  contained in  $(-\infty, a) \cup [b, +\infty)$ .

Note that this set’s boundary,  $\{a, b\}$ , is neither completely included nor excluded from it.

*Example.*  $\mathbb{C}^n$  and  $\emptyset$  (the empty set) are both open and closed. The former contains all balls centered at any points, and the latter is open vacuously. In addition, they are each other’s complements.

Note also that these sets have no boundary.

*Remark.* Recall that we defined a ball to be a subset of a complex Euclidean space. Due to our definition of open sets using this notion of a ball, Definition 1.5.2 and Definition 1.5.4 are thus specialized to finite-dimensional vector spaces equipped with a norm. While there exist alternative notions of open sets for other types of spaces, we do not cover them here.

We now consider sequences of points, defining a notion of them “approaching” some point.

**Definition 1.5.5.** A sequence  $x_1, x_2, x_3, \dots$  has a **limit point**  $\lim_{k \rightarrow \infty} x_k = x$  when for all  $\epsilon > 0$ , there exists a threshold  $n$  such that for all  $k \geq n$ ,  $\|x_k - x\| \leq \epsilon$ .

*Example.* The limit point of  $x_k = 2^{-k}$  is  $x = 0$ . Observe that  $\|x_k - x\| = \|2^{-k} - 0\| = 2^{-k}$ . For any  $\epsilon > 0$ , set  $n = \log_2(\epsilon^{-1})$ . When  $k \geq n$ , we then have  $2^{-k} \leq 2^{-n} = 2^{\log_2(\epsilon)} = \epsilon$ , as needed.

*Example.* Let  $x_k$  be the decimal expansion of  $\pi$  up to  $k$  decimals (e.g.,  $x_2 = 3.14$ ). The limit point of  $x_k$  is  $x = \pi$ . The proof is similar to the above example, with  $\|x_k - x\| \leq 10^{-k}$  and  $n = \log_{10}(\epsilon^{-1})$ .

While in both of the above examples the values of  $\|x_k - x\|$  decrease with  $k$ , note that Definition 1.5.5 only requires that they are  $\epsilon$ -close past some threshold  $n$ . Limit points reveal the following property of closed sets, which serves as an alternative definition.

**Lemma 1.5.6.** A set  $S$  is closed if and only if it contains the limit points of all convergent sequences in  $S$ .

Let us confirm this property with an earlier example.

*Example.* The set  $[a, b)$  is not closed, since  $x_k = b - 2^{-k}(b - a)$  is in this set but converges to  $b$ .

As a last remark on the notions of open and closed sets, we note how these properties behave under intersections of sets.

**Lemma 1.5.7.** The intersection of finitely many open sets is open.

*Proof.* It suffices to prove this for two open sets  $A$  and  $B$  (via Definition 1.5.2), and the rest follows by induction. Letting  $x \in A \cap B$ , we have  $x \in A$  and  $x \in B$ , and so there exist  $a, b > 0$  such that  $B_a(x) \subseteq A$  and  $B_b(x) \subseteq B$ . Then,  $B_{\min\{a,b\}}(x) \subseteq A \cap B$ .  $\square$

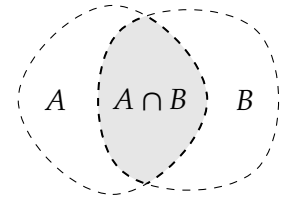


Figure 1.2: Intersection of two open sets.

A visualization of this property is seen in Figure 1.2; the boundary of  $A \cup B$  is formed from parts of the boundaries of  $A$  and  $B$ , and is thus not part of  $A \cup B$ .

On the other hand, the intersection of infinitely many sets may or may not be open.

*Example.* Consider the following intersections of infinitely many sets of the form  $(-\epsilon, \epsilon)$ .

$$\bigcap_{\epsilon \geq 1} (-\epsilon, \epsilon) = (-1, 1), \quad \bigcap_{\epsilon > 0} (-\epsilon, \epsilon) = \{0\}. \quad (1.105)$$

Note that the former is open, while the latter is not.

*Remark.* Similar examples exist even for intersections of a *countably* infinite number of sets. For instance, the intersection of  $S_k = (-2^k, 2^k)$  is open when  $k \in \mathbb{Z}_{\geq 0}$ , but closed when  $k \in \mathbb{Z}$ .

When instead dealing with closed sets, closeness is preserved under intersections even with infinitely many constituents.

**Lemma 1.5.8.** *The intersection of (finitely or infinitely many) closed sets is closed.*

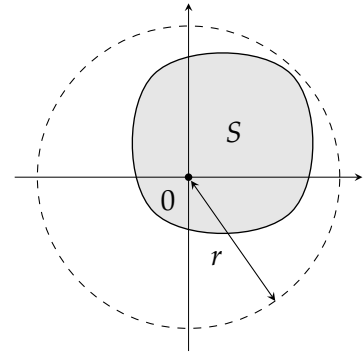
*Example.* Consider the following intersections of infinitely many sets of the form  $[-\epsilon, \epsilon]$ .

$$\bigcap_{\epsilon \geq 1} [-\epsilon, \epsilon] = [-1, 1], \quad \bigcap_{\epsilon > 0} [-\epsilon, \epsilon] = \{0\}. \quad (1.106)$$

Note that both are closed.

We next define the notion of boundedness for sets, which essentially means that the set does not trail off to infinity. In particular, we attempt to contain the set within a ball; doing so successfully ensures that elements within it can only grow so large.

**Definition 1.5.9.** A **bounded** set  $S$  is one for which there exists  $r > 0$  satisfying  $S \subseteq B_r(0)$ . In other words,  $\|x\| \leq r$  for all  $x \in S$ .



*Remark.* In the above definition, we make an arbitrary choice by centering the containing ball at 0. It turns out that this loses no generality, a fact that we leave to the reader as Exercise 1.7.

Figure 1.3:  $S$  is bounded since  $S \subseteq B_r(0)$ .

*Example.*  $[0, 1]$  is bounded. Let  $r = 2$ , then  $B_r(0) = (-2, 2)$  and  $[0, 1] \subseteq B_r(0)$ .

*Example.*  $\mathbb{C}^n$  is not bounded. For any  $r > 0$ , there exists an  $x \in \mathbb{C}^n$  such that  $\|x\| > r$ .

Sometimes we wish a set to be closed in addition to bounded. This earns it a special name.

**Definition 1.5.10.** A **compact** set is one which is closed and bounded.

As we observe later, this has nice consequences in the context of optimization problems.

Most of the definitions above—of limit points, the ball (and thus open set and bounded sets)—require a choice of a specific norm. The point of the following statement is to demonstrate that it does not matter which norm you choose, for with a simple adjustment of constants, all norms are equivalent (in finite dimensions).

**Lemma 1.5.11.** *Given norms  $\|\cdot\|_a$  and  $\|\cdot\|_b$  on a complex Euclidean space  $\mathcal{X}$ , there exist constants  $c, d > 0$  such that  $c\|x\|_a \leq \|x\|_b \leq d\|x\|_a$  for all  $x \in \mathcal{X}$ .*

*Example.* Let  $x \in \mathbb{C}^n$ . We have  $\|x\|_2 \leq \|x\|_1 \leq \sqrt{n}\|x\|_2$ .

In optimization, we often would like to pick the maximum or minimum number from a set. However, there could be situations where such an optimal value is not part of the set, or where it does not exist. To address this, we define slightly more general notions, called supremum and infimum. We make use of the extended real line, defined as  $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$ .

**Definition 1.5.12.** The **supremum** of  $S \subseteq \overline{\mathbb{R}}$  is the smallest  $b \in \overline{\mathbb{R}}$  such that  $x \leq b$  for all  $x \in S$ . We then denote  $\sup(S) = b$ . In other words, the supremum is the least upper bound. If  $\sup(S) \in S$ , then  $\max(S) = \sup(S)$  is the **maximum**.

*Example.* Let  $S_1 = [a, b]$  and  $S_2 = (a, b)$ . Then,  $\sup(S_1) = \sup(S_2) = b$ . Likewise  $\max(S_1) = b$ , but  $S_2$  has no maximum.

The supremum is thus conceptually similar to a maximum of a set, a key difference being that the maximum may not exist for some sets such as  $(a, b)$  above.

Note that we make the definition of the supremum in terms of  $\overline{\mathbb{R}}$  in place of  $\mathbb{R}$  for  $\sup(S)$  to always exist. Otherwise, we cannot guarantee that there is a least upper bound (as is with  $S = \emptyset$ , where any number is vacuously an upper bound) or that an upper bound exists in the first place (as is with  $S = \mathbb{R}$ , which has no upper bound). Instead, the above definition as-is gives  $\sup(\emptyset) = -\infty$  and  $\sup(\mathbb{R}) = \infty$ .

Similarly, we define the notion of a smallest value of a set.

**Definition 1.5.13.** The **infimum** of  $S \subseteq \overline{\mathbb{R}}$  is the largest  $a \in \overline{\mathbb{R}}$  such that  $x \geq a$  for all  $x \in S$ . We then denote  $\inf(S) = a$ . In other words, the infimum is the greatest lower bound. If  $\inf(S) \in S$ , then  $\min(S) = \inf(S)$  is the **minimum**.

*Example.* Let  $S_1 = [a, b]$  and  $S_2 = (a, b)$ . Then,  $\inf(S_1) = \inf(S_2) = a$ . Likewise  $\min(S_1) = a$ , but  $S_2$  has no minimum.

*Remark.* We could make the definitions of the supremum and infimum more general by defining them on partially ordered sets instead of  $\overline{\mathbb{R}}$ .

## 1.5.2 Properties of functions

Oftentimes we wish to do more than just to pick the best value from a set, and instead to pick one that is the largest or smallest after being somehow manipulated. That is, we wish to find an input to some process that gives a good output value, with this process modeled via a function.

Since we want to compare the outputs of these functions, we require their outputs to be real, or possibly infinite. Though, it becomes important to ask when such functions output finite values.

**Definition 1.5.14.** The **domain** of a real-valued function  $f : \mathcal{X} \rightarrow \overline{\mathbb{R}}$  is the set of all inputs that  $f$  maps to finite outputs.

$$\text{dom}(f) := \{x \in \mathcal{X} : -\infty < f(x) < +\infty\}. \quad (1.107)$$

**Example.** Consider a function  $f : \mathbb{R} \rightarrow \{-\infty\} \cup \mathbb{R}$  defined as

$$f(x) = \begin{cases} -\infty, & \text{if } x \leq 0, \\ \log(x), & \text{if } x > 0. \end{cases} \quad (1.108)$$

The domain of  $f$  is  $\text{dom}(f) = \{x \in \mathbb{R} : x > 0\}$ .

Several nice properties arise for functions that do not have sudden changes, defined as follows.

**Definition 1.5.15.** A function  $f : S \rightarrow \mathcal{Y}$  **continuous at**  $c \in S$  has  $f(c) = \lim_{i \rightarrow \infty} f(x_i)$  for any sequence  $x_1, x_2, \dots \in S$  with limit point  $c = \lim_{i \rightarrow \infty} x_i$ . A **continuous** function  $f$  is continuous at all  $c \in S$ .

That is, for a function  $f$  to be differentiable at  $c$ , any sequence convergent to  $c$  must converge when evaluated through  $f$ , with the value it converges to equal across sequences and to  $f(c)$ .

Due to invoking limits on  $x_i$  and  $f(x_i)$ , the definition of continuity above requires norms on both  $\mathcal{X}$  and  $\mathcal{Y}$ , where  $S \subseteq \mathcal{X}$ . Still, this is just one of several ways to define continuous functions, with one other approach instead leveraging the definition of open sets.

**Example.** The following function is not continuous on its domain  $\mathbb{R}$ .

$$f(x) = \begin{cases} \sin(x^{-1}) & \text{if } x \neq 0, \\ 0 & \text{otherwise.} \end{cases} \quad (1.109)$$

It is continuous at  $c = \pi^{-1}$  where  $f(c) = 0$ , but not continuous at  $c = 0$  since it evaluates to 0 at  $x_k = (k\pi)^{-1}$  and to 1 at  $x_k = ((2k + 1/2)\pi)^{-1}$ .

Similar to how we can evaluate the limit of a sequence when mapped through a function, we can find suprema and infima of sets mapped through functions. Specifically, this means we instead apply the supremum and infimum operators to the *image* of the function.

That is, if  $A \subseteq B$  and  $f : B \rightarrow \mathbb{R}$  is a function,

$$\sup_{x \in A} f(x) := \sup\{f(x) : x \in A\}, \quad \inf_{x \in A} f(x) := \inf\{f(x) : x \in A\}. \quad (1.110)$$

Earlier, we noted that compactness and continuity yield nice properties, one in particular arising from a combination of the two.

**Lemma 1.5.16** (Extreme value theorem). *If a function is continuous on a compact set, then it has a maximum and a minimum (i.e., it attains its supremum and infimum).*

The same does not necessarily hold over non-compact sets, as seen in the below example.

**Example.** Consider  $f(x) = x^2$  over  $x \in (0, 1)$ , which is not compact since it is not closed. Let  $S = \{x^2 : x \in (0, 1)\}$ . Then,  $\inf(S) = 0$  and  $\sup(S) = 1$ , but  $\min(S)$  and  $\max(S)$  do not exist.

Determining whether the maximum and minimum exist is important, as in certain cases we care about whether we can attain optimal solutions or only approach them.

## 1.6 Convex analysis

We continue with topics in analysis, now introducing additional structure via the concept of convexity. This is a vital concept for optimization problems as it simplifies the process of finding feasible solutions and, importantly, prohibits functions from having local minima.

While these topics are still independent from linear algebra, we now begin to progressively interlink the two by treating elements of sets as operators (correspondingly, we now denote elements of sets by uppercase letters such as  $X$  and  $Y$ , as we did for operators). This lets us introduce sets of operators important to quantum theory, which appear again in Chapter 3.

We begin with a means of combining objects that support addition, which forms the core idea behind concepts related to convexity.

**Definition 1.6.1.** A **convex combination** of points  $X_1, \dots, X_n$  is

$$\sum_{i=1}^n \lambda_i X_i, \quad \text{such that } \lambda_1, \dots, \lambda_n \geq 0 \text{ and } \sum_{i=1}^n \lambda_i = 1. \quad (1.111)$$

Note that the above assumptions imply  $\lambda_i \in [0, 1]$  for every  $\lambda_i$ . In some sense, convex combinations are “averages” of multiple points. Convex combinations of two points cover the line segment between them; of three points, a triangle with them as the vertices; and so on.

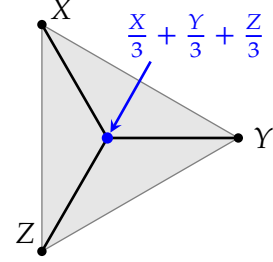


Figure 1.4: Convex combinations of three points.

### 1.6.1 Convex sets

We next define the notion of convexity for a set.

**Definition 1.6.2.** A **convex set**  $C$  has  $\lambda X + (1 - \lambda)Y \in C$  for all  $X, Y \in C$  and  $\lambda \in [0, 1]$ . That is, convex sets contain all of their line segments.

*Remark.* A convex set  $C$  contains all the convex combinations of vectors in  $C$ . This is equivalent to the above definition, despite convex combinations potentially consisting of three (or more) points. We leave this to the reader as Exercise 1.8.

Let us take a look at the following examples to illustrate the definition of convex sets.

**Example.** The set  $C = \{X \in \mathcal{X} : \|X\|_2 \leq 1\}$  is convex.

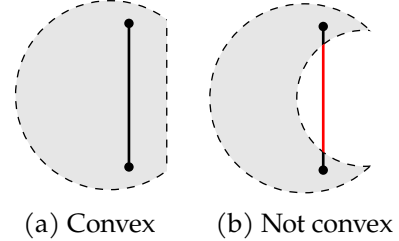


Figure 1.5: Convexity of sets.

*Proof.* Let  $X, Y \in C$  (so,  $\|X\|_2 \leq 1$  and  $\|Y\|_2 \leq 1$ ) and  $\lambda \in [0, 1]$ . We have the following.

$$\|\lambda X + (1 - \lambda)Y\|_2 \leq \|\lambda X\|_2 + \|(1 - \lambda)Y\|_2 \quad (1.112a)$$

$$= \lambda\|X\|_2 + (1 - \lambda)\|Y\|_2 \quad (1.112b)$$

$$\leq \lambda + (1 - \lambda) \quad (1.112c)$$

$$= 1. \quad (1.112d)$$

Thus,  $\lambda X + (1 - \lambda)Y \in C$ . □

**Example.** The set  $C = \{X \in \mathcal{X} : \|X\|_2 = 1\}$  is *not* convex.

*Proof.* Let  $X$  be any vector in  $C$ , so  $\|X\|_2 = 1$ . Define  $Y = -X$ , and note that  $Y \in C$  as well. Choose  $\lambda = 1/2$ . Notice that  $X/2 + Y/2 = 0 \notin C$ , so  $C$  cannot be convex.  $\square$

**Example.** The set of unitaries  $U(\mathcal{X})$  is *not* convex, which is left to the reader as Exercise 1.10.

Intersections of sets preserve convexity (see Figure 1.2 for a visualization).

**Lemma 1.6.3.** Let  $\{C_i : i \in S\}$  be a collection of convex sets. Then,  $\bigcap_{i \in S} C_i$  is convex.

*Proof.* Let  $X, Y \in \bigcap_{i \in S} C_i$ , meaning  $X \in C_i$  and  $Y \in C_i$  for all  $i \in S$ . Let  $\lambda \in [0, 1]$ . Then,  $\lambda X + (1 - \lambda)Y \in C_i$  for any  $i \in S$ , since  $C_i$  is convex. Therefore,  $\lambda X + (1 - \lambda)Y \in \bigcap_{i \in S} C_i$ .  $\square$

We take a slight detour to define a couple of notions relevant to convex sets.

**Definition 1.6.4.** A **hyperplane** in  $\mathcal{X}$  is specified by a nonzero vector  $A \in \mathcal{X} \setminus \{0\}$  and  $b \in \mathbb{R}$ , and is defined as  $H = \{X \in \mathcal{X} : \langle A, X \rangle = b\}$ .

In a  $d$ -dimensional space, a hyperplane has  $d - 1$  dimensions. In two dimensions it is a line, in three dimensions a plane, and so on. A hyperplane splits the whole space in two, which is relevant for the following.

**Definition 1.6.5.** A **halfspace** is the part of the vector space  $\mathcal{X}$  that lies on one side of a hyperplane; that is,  $\{X \in \mathcal{X} : \langle A, X \rangle \leq b\}$ .

We can harness the ability of hyperplanes to separate spaces specifically to separate points from sets. This has additional considerations relating to the set boundary.

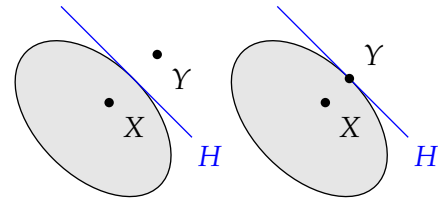
**Definition 1.6.6.** Hyperplane  $H = \{X \in \mathcal{X} : \langle A, X \rangle = b\}$  **separates** a point  $Y$  from a set  $C$  when

$$\langle A, X \rangle \leq b < \langle A, Y \rangle \quad \text{for all } X \in C, \quad (1.113)$$

implying  $Y \notin C$ . Hyperplane  $H = \{X \in \mathcal{X} : \langle A, X \rangle = b\}$  **weakly separates** a point  $Y$  from a set  $C$  when

$$\langle A, X \rangle \leq b \leq \langle A, Y \rangle \quad \text{for all } X \in C. \quad (1.114)$$

Note that  $Y$  may be on the boundary in this case.



(a) Separates (b) Weakly separates

Figure 1.6: Separating hyperplanes.

Hyperplane separation relates to convexity of sets via the following theorem.

**Theorem 1.6.7** (Separating hyperplane theorem). If  $C$  is a convex set and  $Y \notin C$ , then there exists a hyperplane that weakly separates  $Y$  from  $C$ . If  $C$  is additionally closed, then there exists a

hyperplane that separates  $Y$  from  $C$ .

*Remark.* While we do not use it here, a more general version of the above theorem separates two (disjoint, possibly closed) convex sets.

For points outside of non-convex sets, there may not exist a separating hyperplane.

Convex sets (that are also closed) may alternatively be defined by intersections of their tangent hyperplanes.

**Lemma 1.6.8.**  $C$  is closed and convex if and only if  $C$  is the intersection of all halfspaces that contain it.

*Proof.* Let  $D$  be the intersection of all halfspaces that contain  $C$ . We have  $C \subseteq D$ . Suppose  $Y \notin C$ . Then there exists a hyperplane  $H$  such that  $H$  separates  $Y$  from  $C$ . Thus  $Y \notin D$ . Taking the contrapositive yields  $D \subseteq C$ .  $\square$

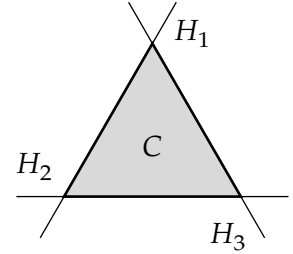


Figure 1.7:  $C$  is the intersection of halfspaces that contain it.

## 1.6.2 Convex cones

We now introduce the notion of cones, as well as their duals and polars. These structures are relevant for conic programming (which semidefinite programming resides in).

**Definition 1.6.9.** A set  $K$  is a **cone** if for all  $X \in K$  and  $\lambda > 0$ , we have  $\lambda X \in K$ .

That is, for every point in a cone  $K$ , the ray from the origin through that point and further must also be contained in  $K$ . As such, cones generally stretch out towards infinity.

*Remark.* A special case of the above rule is the point  $X = 0$ , for which the ray  $\lambda X$  is just  $\{0\}$ . Note that since we have defined cones via  $\lambda > 0$ , the point 0 may or may not be part of a valid cone.

*Example.* The sets  $\mathbb{C}^n$ ,  $\{0\}$ ,  $\text{Pos}(\mathcal{X})$ ,  $\text{Pd}(\mathcal{X})$ , and  $\mathcal{L}_n := \{(x, t) \in \mathbb{R}^{n+1} : \|x\|_2 \leq t\}$  are all cones.  $\mathcal{L}_n$  (also called the second-order cone) is additionally convex.

We say that  $K$  is a **convex cone** when it is convex and a cone. Defining the addition of sets  $A$  and  $B$  as  $A + B = \{a + b : a \in A, b \in B\}$ , we get an alternative definition of convexity for cones.

**Lemma 1.6.10.** A cone  $K$  is convex if and only if  $K + K \subseteq K$ .

*Proof.* Suppose  $K$  is a convex cone. Let  $X, Y \in K$ , which implies  $X/2 + Y/2 \in K$  by convexity. Since  $K$  is a cone, we scale by 2 to get  $X + Y \in K$ . Therefore,  $K + K \subseteq K$ . Conversely, suppose  $K + K \subseteq K$  and  $X, Y \in K$ . Since  $K$  is a cone,  $\lambda X \in K$  and  $(1 - \lambda)Y \in K$  for any  $\lambda \in [0, 1]$ . Thus,  $\lambda X + (1 - \lambda)Y \in K$ , meaning  $K$  is convex.  $\square$

*Example.* The sets  $\text{Pd}(\mathcal{X})$  and  $\text{Pos}(\mathcal{X})$  are convex cones. Note that  $\text{Pd}(\mathcal{X})$  does not contain its boundary and is open, whereas  $\text{Pos}(\mathcal{X})$  contains its boundary (including 0) and is closed.

The above example is the link connecting conic programming (optimizing over convex cones) to semidefinite programming (optimizing over  $\text{Pos}(\mathcal{X})$ , a specific kind of convex cone).

The set  $D(\mathcal{X})$  below has later relevance as the set of quantum states.

*Example.*  $D(\mathcal{X}) := \{X \in \text{Pos}(\mathcal{X}) : \text{Tr}(X) = 1\}$  is convex but not a cone.

In particular,  $D(\mathcal{X})$  is the intersection of a convex cone and a plane, both being convex.

*Example.*  $\text{Pos}(\mathcal{X}) \cup (-\text{Pos}(\mathcal{X}))$  is a cone but not convex.

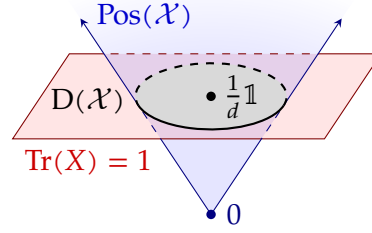


Figure 1.8: The set  $D(\mathcal{X})$ .

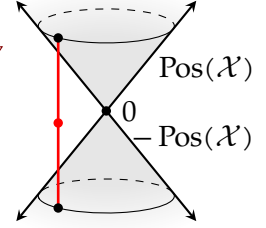


Figure 1.9: A cone that is not convex.

We can check this by combining elements from both parts of the cone. E.g.,  $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \in \text{Pos}(\mathcal{X})$  and  $\begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix} \in -\text{Pos}(\mathcal{X})$ , but  $\begin{pmatrix} 1/2 & 0 \\ 0 & -1/2 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix} \notin \text{Pos}(\mathcal{X}) \cup (-\text{Pos}(\mathcal{X}))$ .

Duality is a general concept in mathematics with versatile appearances. For our purposes, optimization over a convex cone is closely related to a different problem over a *dual* of said cone.

**Definition 1.6.11.** The **dual** of a set  $S \subseteq \mathcal{X}$  is  $S^* = \{Y \in \mathcal{X} : \langle X, Y \rangle \geq 0, \forall X \in S\}$ .

The dual  $S^*$  is closed and convex, since it is an intersection of halfspaces.  $S^*$  is always a cone, which is left as Exercise 1.11 to the reader. A convex cone  $K$  is *self-dual* when  $K = K^*$ .

*Example.* The following identities hold:

$$\text{Pos}(\mathcal{X})^* = \text{Pos}(\mathcal{X}), \quad (\mathbb{R}_+^n)^* = \mathbb{R}_+^n := \left\{ \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n : \forall i, x_i \geq 0 \right\}, \quad (1.115a)$$

$$\text{Pd}(\mathcal{X})^* = \text{Pos}(\mathcal{X}), \quad \mathcal{L}_n^* = \mathcal{L}_n := \{(x, t) \in \mathbb{R}^{n+1} : \|x\|_2 \leq t\}, \quad (1.115b)$$

meaning that  $\text{Pos}(\mathcal{X})$ ,  $\mathbb{R}_+^n$ , and  $\mathcal{L}_n$  are self-dual, while  $\text{Pd}(\mathcal{X})$  is not.

*Example.* The dual cone of  $D(\mathcal{X}) := \{X \in \text{Pos}(\mathcal{X}) : \text{Tr}(X) = 1\}$  is  $D(\mathcal{X})^* = \text{Pos}(\mathcal{X})$ .

*Proof.* From Definition 1.6.11,  $A \subseteq B$  implies  $B^* \subseteq A^*$  for any sets  $A, B$ . So,  $D(\mathcal{X}) \subseteq \text{Pos}(\mathcal{X})$  implies  $\text{Pos}(\mathcal{X})^* = \text{Pos}(\mathcal{X}) \subseteq D(\mathcal{X})^*$ . It remains to show  $D(\mathcal{X})^* \subseteq \text{Pos}(\mathcal{X})$ , to which end we fix an arbitrary  $E \in D(\mathcal{X})^*$  and aim to show  $E \in \text{Pos}(\mathcal{X})$ . We then have the following:

$$\langle E, X \rangle \geq 0, \quad \forall X \in D(\mathcal{X}) \quad (\text{dual cone definition}), \quad (1.116a)$$

$$\langle E, \lambda X \rangle \geq 0, \quad \forall X \in D(\mathcal{X}), \lambda \geq 0, \quad (1.116b)$$

$$\langle E, \tilde{X} \rangle \geq 0, \quad \forall \tilde{X} \in \text{Pos}(\mathcal{X}) \quad (\text{rescaling } \tilde{X} = \lambda X), \quad (1.116c)$$

which implies  $E \in \text{Pos}(\mathcal{X})^* = \text{Pos}(\mathcal{X})$  by definition of the dual cone, finishing the proof.  $\square$

The *polar* of a set is in some way an “opposite” of its dual cone, being defined similarly except with the inequality going the other way (and relative to 1 instead of 0).

**Definition 1.6.12.** The **polar** of a set  $S \subseteq \mathcal{X}$  is  $S^\circ = \{Y \in \mathcal{X} : \langle X, Y \rangle \leq 1, \forall X \in S\}$ .

*Remark.* Some sources instead define the polar cone via  $\langle X, Y \rangle \leq 0$ . The two definitions are equivalent when  $S$  is a cone, but conflict in general.

The polar  $S^\circ$  is closed and convex, since it is an intersection of halfspaces.  $S^\circ$  is not necessarily a cone.

*Example.* Suppose  $\langle X, Y \rangle = 1/2$  for some  $X \in S^\circ$  and  $Y \in S$ . Setting  $\lambda = 3$ , we obtain  $\langle \lambda X, Y \rangle = 3/2$ , meaning  $\lambda X \notin S^\circ$  and  $S^\circ$  is not a cone.

*Example.* The dual set of  $D(\mathcal{X}) := \{X \in \text{Pos}(\mathcal{X}) : \text{Tr}(X) = 1\}$  is  $D(\mathcal{X})^\circ = \{E : 1 - E \in \text{Pos}(\mathcal{X})\}$ .

*Proof.* For convenience, denote  $S := \{E : 1 - E \in \text{Pos}(\mathcal{X})\}$ . Let  $E \in S, X \in D(\mathcal{X})$ ; then,

$$0 \leq \langle X, 1 - E \rangle = \langle X, 1 \rangle - \langle X, E \rangle = 1 - \langle X, E \rangle, \quad (1.117)$$

meaning  $E \in D(\mathcal{X})^\circ$  and  $S \subseteq D(\mathcal{X})^\circ$ . To show  $D(\mathcal{X})^\circ \subseteq S$ , let  $E \in D(\mathcal{X})^\circ, X \in D(\mathcal{X})$ ; then,

$$0 \leq 1 - \langle X, E \rangle = \langle X, 1 \rangle - \langle X, E \rangle = \langle X, 1 - E \rangle, \quad (1.118)$$

implying  $1 - E \in D(\mathcal{X})^* = \text{Pos}(\mathcal{X})$ , so  $E \in S$  and  $D(\mathcal{X})^\circ \subseteq S$ , as needed.  $\square$

Recall that we can use  $A \succcurlyeq B$  to mean  $A - B \succcurlyeq 0$ , i.e.,  $A - B \in \text{Pos}(\mathcal{X})$ . With this notation,

$$D(\mathcal{X})^\circ = \{E : E \preccurlyeq 1\}, \quad (1.119a)$$

$$D(\mathcal{X})^* = \{E : E \succcurlyeq 0\} = \text{Pos}(\mathcal{X}), \quad (1.119b)$$

$$D(\mathcal{X})^\circ \cap D(\mathcal{X})^* = \{E : 0 \preccurlyeq E \preccurlyeq 1\} =: \text{Eff}(\mathcal{X}). \quad (1.119c)$$

$\text{Eff}(\mathcal{X})$  is called the set of *effects* (i.e., measurements) on  $\mathcal{X}$ . For  $X \in D(\mathcal{X})$  and  $E \in \text{Eff}(\mathcal{X})$ , note that  $0 \leq \langle X, E \rangle \leq 1$  can be interpreted as a probability. We revisit this idea in Chapter 3.

When  $K$  is a cone, its polar  $K^\circ$  is also a cone and we additionally have  $K^\circ = -K^*$ .

### 1.6.3 Convex functions

Lastly, we extend convexity to functions, noting their significance in optimization problems.

**Definition 1.6.13.** Let  $C$  be a convex set. A **convex** function  $f : C \rightarrow \mathbb{R}$  satisfies

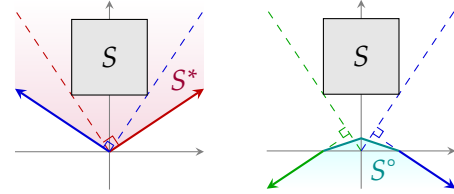
$$f(\lambda X + (1 - \lambda)Y) \leq \lambda f(X) + (1 - \lambda)f(Y), \quad \forall X, Y \in C, \lambda \in [0, 1]. \quad (1.120)$$

That is, convex functions lie below their secant lines. It is crucial that  $C$  is convex: otherwise,  $\lambda X + (1 - \lambda)Y$  may be outside of the domain of  $f$ , making (1.120) ill-defined.

*Example.* Let  $C = \mathbb{R}$ . The function  $f(x) = x^2$  is convex, but  $f(x) = x^3$  is not convex.

Convexity of functions can be visualized as in Figure 1.11; non-convex functions have at least one secant that crosses below the function, whereas all secants of convex functions stay above. Equivalently, the *epigraph* (area above) of a convex function is a convex set.

In Exercise 1.12, we ask the reader to show that convex functions form a convex cone.



(a) Dual cone

(b) Polar set

Figure 1.10: Dual and polar of a set.

**Definition 1.6.14.** Suppose  $C$  is a convex set and  $f : C \rightarrow \mathbb{R}$  is a convex function. A **convex optimization problem** has the form

$$\begin{array}{ll} \text{minimize:} & f(x) \\ \text{subject to:} & x \in C. \end{array} \quad (1.121)$$

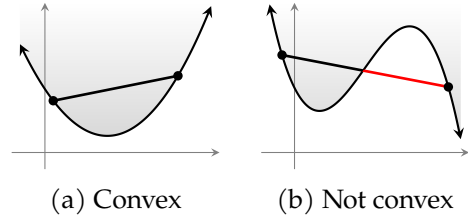


Figure 1.11: Convexity of functions.

We also refer to the above problem as minimizing a convex function  $f$  over a convex set  $C$ . Convex optimization problems have nice properties, many of which we do not cover in this course.

*Example.*  $C = [a, b]$  and  $f(x) = x^2$  are convex, and so is the following optimization problem:

$$\begin{array}{ll} \text{minimize:} & x^2 \\ \text{subject to:} & a \leq x \leq b. \end{array} \quad (1.122)$$

Semidefinite programming is a special case of convex optimization, when  $f$  is linear and  $C$  is related to positive semidefinite matrices. Most of the convex sets we consider in this course are subsets of Hermitian matrices (having  $A = A^*$ , as defined in Lecture 1.2). Why? Recall from Exercise 1.3 that their inner products are real-valued, so they can be used as functions in optimization problems. In contrast to this, complex numbers are not nicely ordered—which is also the reason why we restrict the codomain of  $f$  to  $\mathbb{R}$  in Definition 1.6.14.

## Chapter 1 exercises

**Exercise 1.1.** Prove Hölder's inequality (Theorem 1.1.6) using the definition of the dual norm (Definition 1.1.5). That is, show that

$$|\langle u, v \rangle| \leq \|u\|^* \|v\| \quad (1.123)$$

for all  $u, v$  in some complex Euclidean space  $\mathcal{X}$ , where  $\|\cdot\|^*$  is defined as

$$\|u\|^* := \sup_{w \in \mathcal{X}} \{|\langle u, w \rangle| : \|w\| = 1\} . \quad (1.124)$$

**Exercise 1.2.** Prove Lemma 1.2.11. Let  $\mathcal{X}$  be a complex Euclidean space with  $n = |\mathcal{X}|$ , and show that for any orthonormal basis  $\{x_1, \dots, x_n\} \subseteq \mathcal{X}$ , the identity operator may be written as

$$\mathbb{1}_{\mathcal{X}} = \sum_{i=1}^n x_i x_i^* . \quad (1.125)$$

**Exercise 1.3.** Use the definition of the Hilbert-Schmidt inner product (Definition 1.2.20) to show that the inner product of two Hermitian matrices is real. That is, let  $A, B \in \text{Herm}(\mathbb{C}^{n \times n})$  with spectral decompositions  $A = \sum_{i=1}^n \lambda_i x_i x_i^*$  and  $B = \sum_{j=1}^n \mu_j y_j y_j^*$ . Then, show the equality below:

$$\langle A, B \rangle = \left\langle \sum_{i=1}^n \lambda_i x_i x_i^*, \sum_{j=1}^n \mu_j y_j y_j^* \right\rangle \stackrel{?}{=} \sum_{i=1}^n \sum_{j=1}^n \lambda_i \mu_j |\langle x_i, y_j \rangle|^2 , \quad (1.126)$$

where the rightmost expression is real since  $\lambda_i, \mu_j \in \mathbb{R}$ .

**Exercise 1.4.** Prove Lemmas 1.4.6 and 1.4.7. Show that an operator  $A \in L(\mathcal{X})$  is invertible if and only if its singular values are positive. Given that  $A$  is normal, i.e.,  $AA^* = A^*A$ , show that it is invertible if and only if its eigenvalues are nonzero.

**Exercise 1.5.** Let  $A$  be a Hermitian operator. Prove that  $\langle u, v \rangle_A := u^* A v$  is a valid inner product (according to Definition 1.1.2) if and only if  $A > 0$ .

**Exercise 1.6.** Check whether or not the set of rational numbers  $\mathbb{Q}$  is closed. Here,

$$\mathbb{Q} := \{a/b : a, b \in \mathbb{Z} \wedge b \neq 0\} . \quad (1.127)$$

**Exercise 1.7.** In Definition 1.5.9, we centered the ball bounding a set at zero. Prove the following statement to show that this does not lose generality. If  $S \subseteq B_c(x)$  for some  $x \in S$  and  $c > 0$ , then there exists  $d > 0$  such that  $S \subseteq B_d(0)$ .

**Exercise 1.8.** Prove that a convex set containing the line segments between all pairs of its points (as in Definition 1.6.2) also contains all convex combinations (see Definition 1.6.1) of (any number of) its points.

**Exercise 1.9.** Prove that the set  $\mathbb{Q}$  of rational numbers is not convex, according to Definition 1.6.2.

**Exercise 1.10.** Prove that the set  $U(\mathcal{X})$  of unitaries on  $\mathcal{X}$  (see Definition 1.2.14) is not convex.

**Exercise 1.11.** Let  $S$  be a set. Show that the dual  $S^*$  is a cone (see Definitions 1.6.9 and 1.6.11).

**Exercise 1.12.** Prove that the set of convex functions (see Definition 1.6.13) is a convex cone.

*Hint:* You could use Definition 1.6.9, plus either Definition 1.6.2 or Lemma 1.6.10.

## Chapter 2

# Semidefinite Programming

Having reviewed the sufficient mathematical background in the preceding chapter, we are now ready for an overview of *semidefinite programming* (SDP). While we do not discuss this here, SDP has connections to (often being a generalization of) other types of convex optimization problems.

We aim to increase familiarity with SDPs; from their formal definition and standard form conversions, to the process and intuition behind taking duals. Duality theory in general is a common topic in these discussions—we consider the dual of nearly every optimization problem, with later in the chapter analyzing how they are related through their values and feasible regions.

For a more in-depth overview of semidefinite programming, refer to the following resources:

- Stephen Boyd, Lieven Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
- Levent Tunçel, *Polyhedral and Semidefinite Programming Methods in Combinatorial Optimization*, American Mathematical Society, 2016.

Developing semidefinite programming as a toolset aids later portions of the course, where we use it as a lens for presenting key concepts and results in quantum information.

### 2.1 Introduction to semidefinite programming

What is a semidefinite program? A semidefinite program (SDP) is an optimization problem of a linear function over a positive semidefinite variable subject to affine constraints. More formally:

**Definition 2.1.1.** Let  $\mathcal{X}$  and  $\mathcal{Y}$  be complex Euclidean spaces. A **semidefinite program** is defined by Hermitian operators  $A \in \text{Herm}(\mathcal{X})$  and  $B \in \text{Herm}(\mathcal{Y})$ , as well as a Hermiticity-preserving map  $\Phi$  (meaning  $\Phi$  is linear and maps  $\text{Herm}(\mathcal{X})$  to  $\text{Herm}(\mathcal{Y})$ ). The **standard form** of the SDP is

$$\begin{aligned} \alpha = \text{maximize:} \quad & \langle A, X \rangle \\ \text{subject to:} \quad & \Phi(X) = B \\ & X \in \text{Pos}(\mathcal{X}) . \end{aligned} \tag{2.1}$$

Table 2.1 summarizes common names for some parts of (2.1) above. These allow us to see how the formal definition captures its preceding description:  $\langle A, X \rangle$  is the linear function being optimized, the variable  $X$  is positive semidefinite, and affine constraints are enforced by  $\Phi(X) = B$ .

*Remark.* When  $B = 0$ , the SDP affine constraint is additionally linear.

| Symbol                                            | Name                                                                    |
|---------------------------------------------------|-------------------------------------------------------------------------|
| $(A, B, \Phi)$                                    | Data (defines the SDP)                                                  |
| $X$                                               | Variable                                                                |
| $\Phi(X) = B,$<br>$X \in \text{Pos}(\mathcal{X})$ | Constraints                                                             |
| $\langle A, \cdot \rangle$                        | Objective function                                                      |
| $\langle A, X \rangle$                            | Objective function value                                                |
| $\alpha$                                          | Optimal objective function<br>value (finite, $-\infty$ , or $+\infty$ ) |

Table 2.1: Names for various parts of an SDP.

Although an SDP is an optimization problem, we sometimes view it as a function with the data  $(A, B, \Phi)$  as an input and the optimal objective function value  $\alpha$  as an output. For this reason, we occasionally refer to  $\alpha$  as the SDP's *value*.

### 2.1.1 Feasibility

It is often important to consider which values of the variable  $X$  are *feasible*, meaning that they satisfy the constraints of the SDP. For instance, we can consider the set of all such values.

**Definition 2.1.2.** The **feasible region** of the SDP defined by  $(A, B, \Phi)$  is the set

$$\mathcal{A} := \{X \in \text{Pos}(\mathcal{X}) : \Phi(X) = B\}. \quad (2.2)$$

Any  $X \in \mathcal{A}$  is **feasible**, and  $X \in \mathcal{A} \cap \text{Pd}(\mathcal{X})$  is **strictly feasible**.

Feasibility extends to SDPs as a whole, dependent on whether the feasible region is empty.

**Definition 2.1.3.** An SDP with a feasible region  $\mathcal{A}$  is:

**Infeasible** if  $\mathcal{A} = \emptyset$  (equivalently,  $\alpha = -\infty$ ),  
**Feasible** if  $\mathcal{A} \neq \emptyset$  (equivalently,  $\alpha > -\infty$ ),  
**Unbounded** if  $\alpha = +\infty$  (implying  $\mathcal{A}$  is unbounded).

When  $\mathcal{A} \cap \text{Pd}(\mathcal{X}) \neq \emptyset$ , the SDP is **strictly feasible**.

*Remark.* The above implication is single-directional;  $\mathcal{A}$  may be unbounded with a finite  $\alpha$ .

The equivalent definition of feasibility in terms of  $\alpha$  above follows from treating “maximize” in (2.1) as taking the supremum. Alternatively, the SDP may thus be written as  $\alpha = \sup_{X \in \mathcal{A}} \langle A, X \rangle$ .

*Remark.* We may likewise find the smallest value of linear functions by replacing “maximize” in (2.1) by “minimize”. Consequently, the values of  $\alpha$  flip in Definition 2.1.3; the minimization SDP

is infeasible / feasible / unbounded when  $\alpha$  is  $+\infty$  /  $< +\infty$  /  $-\infty$ , respectively.

We aim to determine the value of  $\alpha$ , which sometimes may be obtained from a specific  $X$ .

**Definition 2.1.4.** An **optimal solution**  $X \in \mathcal{A}$  is one that satisfies  $\langle A, X \rangle = \alpha$ .

*Remark.* An optimal SDP solution may not exist, even when  $\alpha$  is finite (see Section 2.1.4).

## 2.1.2 Geometric interpretation

Figure 2.1 shows what an SDP may look like. The constraints  $X \in \text{Pos}(\mathcal{X})$  and  $\Phi(X) = B$  are represented by the blue cone of PSD matrices and the red affine plane. Both of these stretch out to infinity (upwards for the cone, in four directions for the plane). Any  $X$  belonging to both satisfies both constraints and is thus feasible; therefore, their intersection forms the feasible region  $\mathcal{A}$  (in this case, a parabola).  $\text{Pd}(\mathcal{X})$  is the same as the blue cone minus its surface, so any  $X$  away from the edge of the parabola is strictly feasible. Not pictured here is the objective function;  $A$  defines a direction, and moving  $X$  in it increases the objective function value. Thus, either the optimal  $X$  happens to be on the edge of the hyperbola, or  $A$  points upwards and  $X$  diverges.

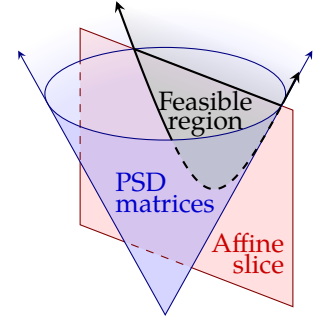


Figure 2.1: Visualization of an SDP.

## 2.1.3 Example SDPs

*Example.* This SDP has a hard affine constraint that makes the feasible region a set with one value.

$$\begin{aligned} \alpha = \text{maximize: } & \langle \mathbb{1}_2, X \rangle \\ \text{subject to: } & X = \mathbb{1}_2 \\ & X \in \text{Pos}(\mathbb{C}^2). \end{aligned} \quad (2.3)$$

From comparing with (2.1), we directly notice that  $A = B = \mathbb{1}_2$  and  $\mathcal{X} = \mathbb{C}^2$ . Since  $X$  takes place of  $\Phi(X)$  in the affine constraint, we have  $\Phi(X) = X$ , giving the identity map. Thus,  $\mathcal{Y} = \mathbb{C}^2$  as well. Notably, the objective function  $\langle \mathbb{1}_2, X \rangle$  may be written as  $\text{Tr}(X)$ .

$$\begin{aligned} \text{Vector spaces: } & \mathcal{X} = \mathbb{C}^2, \mathcal{Y} = \mathbb{C}^2 \\ \text{Data: } & A = \mathbb{1}_2, B = \mathbb{1}_2, \Phi = \mathbb{1}_{\text{L}(\mathbb{C}^2)} \\ \text{Constraints: } & X = \mathbb{1}_2, X \succcurlyeq 0 \\ \text{Feasible region: } & \mathcal{A} = \{\mathbb{1}_2\} \\ \text{Objective function: } & \langle \mathbb{1}_2, X \rangle = \text{Tr}(X) \\ \text{Feasibility: } & \text{Strictly feasible} \\ \text{Boundedness: } & \text{Bounded} \\ \text{Optimal solution? } & \text{Yes, } X = \mathbb{1}_2 \\ \text{Optimal value: } & \alpha = \text{Tr}(\mathbb{1}_2) = 2 \end{aligned}$$

The constraint  $X = \mathbb{1}_2$  forces this to be the only value  $X$  can take; since  $\mathbb{1}_2 \succcurlyeq 0$  it is feasible, and since  $\mathbb{1}_2 \succ 0$  it is also strictly feasible. Being the only possible value, it makes the feasible set  $\mathcal{A} = \{\mathbb{1}_2\}$  and is hence the optimal solution. Evaluating for  $\alpha$ , we get  $\text{Tr}(\mathbb{1}_2) = 2$ .

*Example.* This next SDP features a slightly different constraint that makes it infeasible.

$$\begin{aligned}
\alpha = \text{maximize: } & \text{Tr}(X) \\
\text{subject to: } & X = -\mathbb{1}_2 \\
& X \in \text{Pos}(\mathbb{C}^2) .
\end{aligned} \tag{2.4}$$

The difference with the previous example is that the affine constraint is now  $X = -\mathbb{1}_2$ . But, noting that  $-\mathbb{1}_2 \not\geq 0$ , this only value is ruled out. Thus,  $\mathcal{A} = \emptyset$  and the SDP is infeasible, meaning  $\alpha = -\infty$ . An infeasible SDP is not unbounded.

$$\begin{aligned}
\text{Vector spaces: } & \mathcal{X} = \mathbb{C}^2, \mathcal{Y} = \mathbb{C}^2 \\
\text{Data: } & A = \mathbb{1}_2, B = -\mathbb{1}_2, \Phi = \mathbb{1}_{L(\mathbb{C}^2)} \\
\text{Constraints: } & X = -\mathbb{1}_2, X \geq 0 \\
\text{Feasible region: } & \mathcal{A} = \emptyset \\
\text{Objective function: } & \text{Tr}(X) \\
\text{Feasibility: } & \text{Infeasible} \\
\text{Boundedness: } & \text{Bounded} \\
\text{Optimal solution? } & \text{No solution} \\
\text{Optimal value: } & \alpha = -\infty
\end{aligned}$$

*Example.* For completeness, we now show an SDP with an unbounded feasible region.

$$\begin{aligned}
\alpha = \text{maximize: } & \text{Tr}(X) \\
\text{subject to: } & \langle E_{1,1}, X \rangle = 1 \\
& X \in \text{Pos}(\mathbb{C}^2) .
\end{aligned} \tag{2.5}$$

$$\begin{aligned}
\text{Feasibility: } & \text{Strictly feasible} \\
\text{Boundedness: } & \text{Unbounded} \\
\text{Optimal solution? } & \text{No, diverges} \\
\text{Optimal value: } & \alpha = +\infty
\end{aligned}$$

Strict feasibility follows from  $X = \mathbb{1}_2 > 0$ .

We may explicitly construct the sequence of divergent solutions to show the SDP is unbounded. Consider  $X = E_{1,1} + tE_{2,2}$  with  $t \geq 0$ , which is feasible. We have  $\text{Tr}(X) = 1 + t \rightarrow \infty$  as  $t \rightarrow \infty$ .

## 2.1.4 Example without an optimal solution

In this section we show that optimal solutions do not always exist, even for feasible SDPs.

*Example.* The below problem is an SDP (even though it may not look like one; we come back to this in Lecture 2.2). It is feasible with  $\alpha = 0$ , but has no optimal solution.

$$\begin{aligned}
\alpha = \text{minimize: } & s \\
\text{subject to: } & \begin{pmatrix} t & 1 \\ 1 & s \end{pmatrix} \in \text{Pos}(\mathbb{C}^2) .
\end{aligned} \tag{2.6}$$

We use the remainder of this lecture to prove the above statement.

We start with a couple of technical results about  $2 \times 2$  positive semidefinite matrices.

**Lemma 2.1.5.**  $\begin{pmatrix} t & b \\ b^* & s \end{pmatrix} \in \text{Pos}(\mathbb{C}^2)$  if and only if  $s, t \geq 0$  and  $st \geq |b|^2$ .

*Proof.* Recall from Lemma 1.4.5 that a matrix being PSD is equivalent to it being Hermitian and having nonnegative eigenvalues. We show the two directions separately.

$\Rightarrow$ : Since  $A := \begin{pmatrix} t & b \\ b^* & s \end{pmatrix} \in \text{Pos}(\mathbb{C}^2)$ , we have  $e_1^* A e_1 = t \geq 0$  and  $e_2^* A e_2 = s \geq 0$  by definition.

$A \in \text{Pos}(\mathcal{X})$  also implies  $\lambda_1[A], \lambda_2[A] \geq 0$ , and thus  $\det(A) = st - |b|^2 = \lambda_1[A] \cdot \lambda_2[A] \geq 0$ .

$\Leftarrow$ : By the same determinant calculation,  $st \geq |b|^2$  implies  $\lambda_1[A] \cdot \lambda_2[A] \geq 0$ , meaning it is

not possible for one of  $\lambda_1[A]$  and  $\lambda_2[A]$  to be positive and the other to simultaneously be negative. Since their sum  $\lambda_1[A] + \lambda_2[A] = \text{Tr}(A) = s + t \geq 0$  is nonnegative, we conclude that  $\lambda_1[A], \lambda_2[A] \geq 0$ .  $\square$

*Remark.* Although we do not show it here, the same principle extends to larger matrices—for instance, the diagonal entries of any PSD matrix are nonnegative. The full statement requires the determinants of certain submatrices to be nonnegative (for more details, see *Sylvester's criterion*).

**Corollary 2.1.6.** *If a diagonal entry of a PSD matrix is 0, then the entire row and column containing that entry must be 0 as well.*

*Proof.* Let  $X \in \text{Pos}(\mathcal{X})$  with  $x_{i,i} = 0$ . Consider the submatrix  $\begin{pmatrix} x_{i,i} & x_{i,j} \\ x_{j,i} & x_{j,j} \end{pmatrix} \in \text{Pos}(\mathbb{C}^2)$  for any  $j \neq i$ . By Lemma 2.1.5 we have  $x_{i,i}x_{j,j} \geq |x_{i,j}|^2$ , and since  $x_{i,i} = 0$  we conclude  $x_{i,j} = 0$ .  $\square$

Now, consider again the example SDP in (2.6). The matrix in its PSD constraint is equivalent to that in Lemma 2.1.5 with  $b = 1$ . Thus, feasibility is equivalent to  $s, t, st - 1 \geq 0$ , and by Corollary 2.1.6 we know that  $s \neq 0$  since the opposite leads to a contradiction.

We first show that the SDP is feasible via  $(s, t) = (1, 1)$ , implying  $\alpha \leq 1$  is finite. Since any feasible solution satisfies  $s > 0$ , we additionally have  $\alpha \geq 0$ . Construct a parametrized solution: let  $\epsilon > 0$  and  $(s, t) = (\epsilon, 1/\epsilon)$ . We may verify its feasibility via  $\epsilon, 1/\epsilon \geq 0$  and  $\epsilon \cdot (1/\epsilon) - 1 = 0 \geq 0$ . Since  $s = \epsilon$  can be made arbitrarily close to 0, we have  $\alpha = 0$ . But, attaining  $\alpha = s = 0$  is impossible, so an optimal solution does not exist. In summary, the example SDP is feasible, bounded, and possesses a finite optimal value, but does not have an optimal solution (i.e., although an infimum exists, no minimum exists).

## 2.2 Standard form

Recall that we have defined a semidefinite program (SDP) as being expressible standard form (2.1). Some SDPs we encounter look quite different from this form, but can nevertheless be converted to it; the goal of this lecture is to demonstrate this by example.

Specifically, we aim to find the standard form of the SDP in (2.6). We begin with a change of notation, in the process separating the affine and positive semidefinite constraints:

$$\begin{aligned} \alpha = \text{minimize:} \quad & x_{2,2} \\ \text{subject to:} \quad & x_{1,2} = x_{2,1} = 1 \\ & X = \begin{pmatrix} x_{1,1} & x_{1,2} \\ x_{2,1} & x_{2,2} \end{pmatrix} \in \text{Pos}(\mathbb{C}^2). \end{aligned} \tag{2.7}$$

The next step is to express the variables in the objective and affine constraint in terms of some linear

functions / maps acting on  $X$ , making sure to preserve Hermiticity. Define the following:

$$\tilde{A} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad \text{implying } \langle \tilde{A}, X \rangle = x_{2,2} \quad (\text{the objective function}), \quad (2.8)$$

$$C = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \text{implying } \langle C, X \rangle = x_{1,2} + x_{2,1} = 2 \operatorname{Re}(x_{1,2}), \quad (2.9)$$

$$D = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \text{implying } \langle D, X \rangle = -ix_{1,2} + ix_{2,1} = 2 \operatorname{Im}(x_{1,2}). \quad (2.10)$$

The identities in (2.9) and (2.10) follow from  $X$  being Hermitian, which implies  $x_{1,2} = \overline{x_{2,1}}$ . Setting  $2 \operatorname{Re}(x_{1,2}) = 2$  and  $2 \operatorname{Im}(x_{1,2}) = 0$  is equivalent to  $x_{1,2} = x_{2,1} = 1$ , and the SDP now becomes

$$\begin{aligned} \alpha = \text{minimize: } & \langle \tilde{A}, X \rangle \\ \text{subject to: } & \langle C, X \rangle = 2 \\ & \langle D, X \rangle = 0 \\ & X \in \operatorname{Pos}(\mathbb{C}^2). \end{aligned} \quad (2.11)$$

The two affine constraints above may be combined into  $\Phi(X) = B$  by defining

$$\Phi(X) = \begin{pmatrix} \langle C, X \rangle & 0 \\ 0 & \langle D, X \rangle \end{pmatrix}, \quad \text{and} \quad B = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}. \quad (2.12)$$

Note also that  $\Phi$  is a linear map, and it is Hermiticity-preserving.

At this point the SDP nearly exactly matches the standard form, but still has “minimize” in place of “maximize”. To see how to get around this, consider the following example.

*Example.* Let  $S = \{-12, -7, 0, 5, 20\}$  with  $\min(S) = -12$  and  $\max(S) = 20$ . For the same set with the signs flipped,  $-S = \{12, 7, 0, -5, -20\}$ , we have  $\max(-S) = 12$  and  $\min(-S) = -20$ .

So, another way to find the minimum is via the maximum on the sign-flipped values (and then flip the sign again). That is, if  $\alpha = \min\{f(X)\}$ , then  $-\alpha = \max\{-f(X)\}$ . Taking this into account by redefining  $A = -\tilde{A}$ , we find the standard form of the SDP in (2.6) to be

$$\begin{aligned} -\alpha = \text{maximize: } & \left\langle \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix}, X \right\rangle \\ \text{subject to: } & \begin{pmatrix} x_{1,2} + x_{2,1} & 0 \\ 0 & -ix_{1,2} + ix_{2,1} \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \\ & X = \begin{pmatrix} x_{1,1} & x_{1,2} \\ x_{2,1} & x_{2,2} \end{pmatrix} \in \operatorname{Pos}(\mathbb{C}^2). \end{aligned} \quad (2.13)$$

The negative sign in front of the  $\alpha$  does not affect the SDP, and is only there to remind us to negate the optimal objective function value after solving the problem.

*Example.* Let  $H \in \operatorname{Herm}(\mathcal{X})$ ,  $D(\mathcal{X}) := \{X \in \operatorname{Pos}(\mathcal{X}) : \operatorname{Tr}(X) = 1\}$  and consider the SDP

$$\begin{aligned} \alpha = \text{maximize: } & \langle H, X \rangle \\ \text{subject to: } & X \in D(\mathcal{X}). \end{aligned} \quad (2.14)$$

Its constraint contains both affine and PSD parts; we can see this by writing it in standard form as

$$\begin{aligned} \alpha = \text{maximize: } & \langle H, X \rangle \\ \text{subject to: } & \operatorname{Tr}(X) = 1 \\ & X \in \operatorname{Pos}(\mathcal{X}). \end{aligned} \quad (2.15)$$

*Example.* An SDP can have many variables and constraints; for instance,

$$\begin{aligned}
\alpha = \text{maximize: } & \langle A, X_1 \rangle \\
\text{subject to: } & \Phi_1(X_1) \succeq B_1 \\
& \Phi_2(X_2) = B_2 \\
& \Phi_3(X_3) = \Phi_4(X_4) \\
& \Phi_5(X_5) = B_3 \\
& X_1, \dots, X_4 \in \text{Pos}(\mathcal{X}) \\
& X_5 \in \text{Herm}(\mathcal{Y}) .
\end{aligned} \tag{2.16}$$

The only requirement is the maps being Hermiticity-preserving and the matrices being Hermitian.

### 2.2.1 Real versus complex data

All of the example SDPs we have seen so far from Lecture 2.1 could be solved only using real-valued matrices. This begs the question: does having complex matrices help? Are there any SDPs for which we *have to* use complex numbers?

The immediate answer is: yes, there are SDPs that require complex numbers. For instance,

$$\begin{aligned}
\alpha = \text{maximize: } & \langle \mathbb{1}, X \rangle \\
\text{subject to: } & X = \begin{pmatrix} 5 & i \\ -i & 5 \end{pmatrix} \\
& X \in \text{Pos}(\mathbb{C}^2) .
\end{aligned} \tag{2.17}$$

This is very similar to our first example SDP in (2.3), but its only feasible solution is complex. If  $X$  had to have real entries, then this SDP would be infeasible.

On the other hand, there are situations where we can safely ignore complex numbers.

**Lemma 2.2.1.** *Suppose an SDP has real  $A$ ,  $B$ , and  $\Phi$ . Then, without loss of generality, optimization can be performed over real Hermitian  $X$  (i.e.,  $X = X^T$ , or  $X$  is symmetric).*

*Proof.* Let  $W = \text{Re}(X)$  be the real part of a feasible solution  $X$ . We show that  $W$  is also feasible with the same objective function value (and so the complex part of  $X$  may be safely ignored).

*Objective function value:* We show that  $\langle A, W \rangle = \langle A, X \rangle$ . Write  $X = W + iV$ , where  $W$  and  $V$  are real. Recall that  $A$  is real. By linearity,  $\langle A, X \rangle = \langle A, W \rangle + i\langle A, V \rangle$ . Since  $\langle A, X \rangle$ ,  $\langle A, W \rangle$ , and  $\langle A, V \rangle$  are all real, it must be that  $\langle A, V \rangle = 0$ .

*Affine constraint:* We show that  $\Phi(X) = B$  implies  $\Phi(W) = B$ .  $\Phi$  being a real map means that  $\Phi(\text{Re}(X)) = \text{Re}(\Phi(X))$  (for example,  $\Phi(X) = \langle C, X \rangle$  with a real  $C$  is real). Thus, we have  $\Phi(W) = \Phi(\text{Re}(X)) = \text{Re}(\Phi(X)) = \text{Re}(B) = B$ .

*PSD constraint:* We show that  $X \in \text{Pos}(\mathcal{X})$  implies  $W \in \text{Pos}(\mathcal{X})$ . Writing again  $X = W + iV$ , we have  $X^* = W^T - iV^T$ . Since  $X = X^*$ , this means  $W = W^T$  and  $V = -V^T$ . Let  $G$  be an arbitrary Hermitian matrix. We can compute  $\langle G, X \rangle = \langle G, W \rangle + i\langle G, V \rangle$ . Since this is always real, we must have  $\langle G, V \rangle = 0$ . For any PSD matrix  $M$ , we have  $0 \leq \langle M, X \rangle = \langle M, W \rangle$ . In other words,  $W \in \text{Pos}(\mathcal{X})^* = \text{Pos}(\mathcal{X})$ .  $\square$

*Remark.* When  $X \in \text{Pos}(\mathbb{C}^n)$  is assumed to be real, we may equivalently write  $X \in \text{Pos}(\mathbb{R}^n)$ .

## 2.3 Duality theory

So far, we have seen what an SDP is and how to manipulate it, but not yet how to compute the SDP value  $\alpha$ . Finding a feasible solution provides a one-sided bound on  $\alpha$ , so how can we constrain it from the other direction?

Indeed, we have seen only half of the story, the other half being the **dual**—another SDP closely related to and shedding light on the original problem. The dual problem can bound the optimal value, prove optimality of feasible solutions, characterize optimal solutions, etc.

We refer to the original SDP as the **primal**; in standard form, the primal-dual pair is

| Primal problem                                    | Dual problem                                     |        |
|---------------------------------------------------|--------------------------------------------------|--------|
| $\alpha = \text{maximize: } \langle A, X \rangle$ | $\beta = \text{minimize: } \langle B, Y \rangle$ | (2.18) |
| subject to: $\Phi(X) = B$                         | subject to: $\Phi^*(Y) \succcurlyeq A$           |        |
| $X \in \text{Pos}(\mathcal{X})$ ,                 | $Y \in \text{Herm}(\mathcal{Y})$ .               |        |

Recall that  $\Phi^*$  is the adjoint of  $\Phi$ , the unique linear operator such that  $\langle \Phi(X), Y \rangle = \langle X, \Phi^*(Y) \rangle$  for all  $X \in \mathcal{X}$  and  $Y \in \mathcal{Y}$ . Note that the primal and the dual share the same data:  $A, B$ , and  $\Phi$ .

*Remark.* When the primal SDP has “minimize”, its dual has “maximize” and  $\Phi^*(Y) \preccurlyeq A$ .

It may at first seem unclear that the dual is an SDP. To see this, we may replace the inequality constraint  $\Phi^*(Y) \succcurlyeq A$  with an equality constraint  $\Phi^*(Y) - Z = A$  having  $Z \in \text{Pos}(\mathcal{X})$ . Then,

| Dual problem                                     |        |
|--------------------------------------------------|--------|
| $\beta = \text{minimize: } \langle B, Y \rangle$ | (2.19) |
| subject to: $\Phi^*(Y) - Z = A$                  |        |
| $Y \in \text{Herm}(\mathcal{Y})$                 |        |
| $Z \in \text{Pos}(\mathcal{X})$ .                |        |

Variables that are only constrained to be Hermitian (like  $Y$  above) are said to be *free*, and the constraint is sometimes omitted for conciseness. We likewise sometimes omit the “ $\alpha =$ ” and “ $\beta =$ ” parts of primal-dual pairs, which are understood implicitly to be the values of the respective SDPs.

*Example.* Consider the primal-dual pair with the data  $A = H \in \text{Herm}(\mathcal{X})$ ,  $B = 1$ , and  $\Phi = \text{Tr}$ .

| Primal problem                                    | Dual problem                             |        |
|---------------------------------------------------|------------------------------------------|--------|
| $\alpha = \text{maximize: } \langle H, X \rangle$ | $\beta = \text{minimize: } y$            | (2.20) |
| subject to: $\text{Tr}(X) = 1$                    | subject to: $y\mathbb{1} \succcurlyeq H$ |        |
| $X \in \text{Pos}(\mathcal{X})$ ,                 | $y \in \mathbb{R}$ .                     |        |

Recall that we have seen the above primal in Lecture 2.2, and that its constraints may be written as  $X \in \text{D}(\mathcal{X})$ . The only nontrivial part of deriving this dual is finding the adjoint of  $\Phi = \text{Tr}$ .

$$\langle \Phi(X), y \rangle = \text{Tr}(X) \cdot y = \langle X, \mathbb{1} \rangle \cdot y = \langle X, y\mathbb{1} \rangle = \langle X, \Phi^*(y) \rangle, \quad (2.21)$$

and so we conclude  $\Phi^*(y) = \text{Tr}^*(y) = y\mathbb{1}$ . Note that  $y \in \text{Herm}(\mathbb{C}^1) = \mathbb{R}$ .

We refer to the optimal value, the objective function, and the constraints of dual problems analogously to how we do for primal ones. The respective feasible regions of the two problems are

$$\mathcal{A} := \{X : \Phi(X) = B, X \in \text{Pos}(\mathcal{X})\}, \quad (2.22)$$

$$\mathcal{B} := \{Y : \Phi^*(Y) \succcurlyeq A, Y \in \text{Herm}(\mathcal{Y})\}. \quad (2.23)$$

When the dual is written with  $Z \in \text{Pos}(\mathcal{X})$  as in (2.19) we may additionally include it in the dual feasible region  $\mathcal{B}$  by defining it over  $(Y, Z)$ . We may wonder if there is a relationship between these, and in particular between the SDP values  $\alpha$  and  $\beta$ .

### 2.3.1 Weak duality

The reason for considering the dual problem is that it bounds the primal, and vice versa.

**Theorem 2.3.1** (Weak duality). *For any primal-dual SDP pair, it follows that  $\alpha \leq \beta$ .*

*Remark.* If the primal SDP has “minimize”, the sign is flipped:  $\alpha \geq \beta$ .

*Proof.* We begin by covering some simple cases:  $\alpha = -\infty$  when the primal is infeasible and  $\beta = +\infty$  when the dual is infeasible. In both cases,  $\alpha \leq \beta$  trivially, and so we can assume that both problems are feasible. Let  $X$  be a feasible solution to the primal SDP and  $(Y, Z)$  be a feasible solution to the dual SDP. As this implies  $\Phi(X) = B$  and  $\Phi^*(Y) - Z = A$ , we have

$$\langle Y, B \rangle - \langle A, X \rangle = \langle Y, \Phi(X) \rangle - \langle A, X \rangle = \langle \Phi^*(Y) - A, X \rangle = \langle Z, X \rangle \geq 0, \quad (2.24)$$

where the inequality follows from  $X, Z \in \text{Pos}(\mathcal{X})$ . Consequently, the gap between  $\alpha$  and  $\beta$  is

$$\beta - \alpha = \left( \inf_{(Y, Z) \in \mathcal{B}} \langle Y, B \rangle \right) - \left( \sup_{X \in \mathcal{A}} \langle A, X \rangle \right) \quad (2.25a)$$

$$= \inf_{\substack{X \in \mathcal{A}, \\ (Y, Z) \in \mathcal{B}}} \langle Y, B \rangle - \langle A, X \rangle = \inf_{\substack{X \in \mathcal{A}, \\ (Y, Z) \in \mathcal{B}}} \langle Z, X \rangle \geq 0, \quad (2.25b)$$

Where we use the identity  $-\sup\{f(x)\} = \inf\{-f(x)\}$ , as mentioned in Lecture 2.2.  $\square$

**Corollary 2.3.2.** *If  $Y$  is dual feasible, then  $\alpha \leq \langle B, Y \rangle$ . If  $X$  is primal feasible, then  $\langle A, X \rangle \leq \beta$ .*

*Example.* Let us apply this principle to bound the values of the example SDP pair (2.20). Consider the spectral decomposition  $H = \sum_i \lambda_i v_i v_i^*$ , let  $y = \lambda_{\max}[H] = \max_i \{\lambda_i[H]\}$ , and by resolution of the identity we have  $\mathbb{1} = \sum_i v_i v_i^*$ . Then,  $\lambda_{\max}[H] - \lambda_i \geq 0$  implies  $y\mathbb{1} - H = \sum_i (\lambda_{\max}[H] - \lambda_i) v_i v_i^* \succcurlyeq 0$ , meaning  $y = \lambda_{\max}[H]$  is dual feasible. By weak duality, we have that  $\alpha \leq \beta \leq \lambda_{\max}[H]$ .

Similarly, we show a feasible primal solution. Let  $v$  be an eigenvalue associated with  $\lambda_{\max}$ , that is,  $Hv = \lambda_{\max}[H] \cdot v$ . Let  $X = vv^*$ , then  $\langle H, X \rangle = \langle H, vv^* \rangle = v^* H v = \lambda_{\max}[H]$ . By weak duality,  $\lambda_{\max}[H] \leq \alpha \leq \beta$ . With the previous result,  $\alpha = \beta = \lambda_{\max}[H]$  and both solutions are optimal.

**Corollary 2.3.3.** *If  $X$  is primal feasible,  $Y$  is dual feasible, and  $\langle A, X \rangle = \langle B, Y \rangle$ , then  $X$  is primal optimal,  $Y$  is dual optimal and  $\alpha = \langle A, X \rangle = \langle B, Y \rangle = \beta$ .*

However, it is not always possible to find such feasible solutions that satisfy  $\langle A, X \rangle = \langle B, Y \rangle$ , so it may very well be possible that  $\alpha \neq \beta$ .

Weak duality additionally rules out certain combinations of infeasible and unbounded primal-dual pairs. In particular, we have the following.

*Example.* Let  $B \in \text{Herm}(\mathcal{X})$  be some Hermitian matrix, and consider the following SDP.

| Primal problem                                                                                                           | Dual problem                                                                      |        |
|--------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|--------|
| $\alpha = \text{maximize: } \text{Tr}(X)$<br>subject to: $\langle B, X \rangle = 0$<br>$X \in \text{Pos}(\mathcal{X})$ , | $\beta = \text{minimize: } 0 \cdot y$<br>subject to: $y \cdot B \succcurlyeq 1$ . | (2.26) |

The value of the dual depends on feasibility;  $\beta = 0$  if feasible and  $\beta = +\infty$  otherwise. For instance, if  $B = \mathbb{1}$  then  $y = 1$  is feasible, but if  $B = 0$  the dual is infeasible. Supposing the primal is unbounded ( $\alpha = +\infty$ ), by weak duality we then find that the dual cannot be feasible. Specifically, if  $Y$  is dual feasible, then  $\alpha \leq \beta \leq \langle B, Y \rangle < +\infty$  contradicts  $\alpha = +\infty$ .

The same principle extends to when the dual is unbounded.

**Corollary 2.3.4.** *If the primal is unbounded ( $\alpha = +\infty$ ), then the dual is infeasible. Similarly, if the dual is unbounded ( $\beta = -\infty$ ), then the primal is infeasible.*

*Example.* Recall that  $E_{1,1}$  is a matrix that is zero everywhere except its  $(1,1)$  entry, which is 1.

| Primal problem                                                                                                                  | Dual problem                                                                     |        |
|---------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------|--------|
| $\alpha = \text{maximize: } \text{Tr}(X)$<br>subject to: $\langle E_{1,1}, X \rangle = -1$<br>$X \in \text{Pos}(\mathcal{X})$ , | $\beta = \text{minimize: } -y$<br>subject to: $y \cdot E_{1,1} \succcurlyeq 1$ . | (2.27) |

Both parts of the above SDP are infeasible, so  $\alpha = -\infty$ ,  $\beta = +\infty$ , meaning  $\beta - \alpha$  is infinite.

**Definition 2.3.5.** The **duality gap** of an SDP with primal and dual optimal objective function values  $\alpha$  and  $\beta$  is  $\beta - \alpha$ . By weak duality, the duality gap is always nonnegative.

*Example.* It is possible to have  $\alpha \neq \beta$ , even when both are finite.

| Primal problem                                                                                                                                              | Dual problem                                                                                                                      |        |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------|--------|
| $\alpha = \text{maximize: } -t$<br>subject to: $\begin{pmatrix} t & a & b \\ a & 0 & \frac{1-t}{2} \\ b & \frac{1-t}{2} & c \end{pmatrix} \succcurlyeq 0$ , | $\beta = \text{minimize: } y$<br>subject to: $\begin{pmatrix} y & 0 & 0 \\ 0 & z & y \\ 0 & y & 0 \end{pmatrix} \succcurlyeq 0$ . | (2.28) |

Recall that in Corollary 2.1.6 we have established the following: a zero on the diagonal of a positive semidefinite matrix implies that the whole row and column must be zero as well. The elements in the second row and column of primal SDP and the elements in the third row and column of dual SDP must therefore be zero, and we have  $a = 0$ ,  $t = 1$ , and  $y = 0$ . Since  $(a, b, c, t) = (0, 0, 0, 1)$  is primal feasible and  $(y, z) = (0, 0)$  is dual feasible, we have  $\alpha = -1$  and  $\beta = 0$  (no better solutions exist since  $t$  and  $y$  are fixed). Thus,  $\alpha \neq \beta$  with a finite duality gap.

*Example.* A duality gap is also possible when only one of  $\alpha$  and  $\beta$  is finite. Consider the SDP

| Primal problem                                                                                                                                  | Dual problem                                                                                                                                                                                              |
|-------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| $\alpha = \text{maximize: } x_{1,2} + x_{2,1}$<br>$\text{subject to: } x_{1,1} = 0$<br>$x_{2,2} = x_{3,3}$<br>$X \in \text{Pos}(\mathbb{C}^3),$ | $\beta = \text{maximize: } 0$<br>$\text{subject to: } \begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & -b \end{pmatrix} \succcurlyeq \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$ |

(2.29)

Invoking the earlier principle,  $x_{1,1} = 0$  implies  $x_{1,2} = x_{2,1} = 0$ . Since  $X = 0$  is primal-feasible, we have  $\alpha = 0$ . The same principle applied to the dual reveals that  $b$  must be positive due to appearing as the (2,2) entry. However, as  $b$  is also the (3,3) entry it cannot be positive. This contradiction makes the dual infeasible, meaning  $\beta = +\infty$ . Since  $\alpha = 0$  is finite, the duality gap is *semi-infinite*.

*Example.* A semi-infinite duality gap is also possible when the dual is feasible but not the primal.

| Primal problem                                                                                                                                                                 | Dual problem                                                                                                          |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------|
| $\alpha = \text{maximize: } 0$<br>$\text{subject to: } X \succcurlyeq \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}$<br>$x_{1,1} \leq 0$<br>$X \in \text{Pos}(\mathbb{C}^2),$ | $\beta = \text{minimize: } y_{1,2} + y_{2,1}$<br>$\text{subject to: } t \cdot E_{1,1} \succcurlyeq Y \succcurlyeq 0.$ |

(2.30)

By similar reasoning,  $x_{1,1}$  must be both positive and nonpositive, meaning the primal is infeasible and  $\alpha = -\infty$ . The dual has  $y_{2,2} = 0$  from the constraints, which implies  $y_{1,2} = 0$  and  $y_{2,1} = 0$ . Since  $(t, Y) = (0, 0)$  is feasible, we have  $\beta = 0$  and again a semi-infinite duality gap.

|                                      | Dual | Unbounded<br>( $\beta = -\infty$ ) | Finite $\beta$                                | Infeasible<br>( $\beta = +\infty$ ) |
|--------------------------------------|------|------------------------------------|-----------------------------------------------|-------------------------------------|
| Primal                               |      |                                    |                                               |                                     |
| Unbounded<br>( $\alpha = +\infty$ )  |      | Impossible                         | Impossible                                    | Possible<br>(but trivial)           |
| Finite $\alpha$                      |      | Impossible                         | ★ Possible<br>( $\alpha \neq \beta$ possible) | Possible<br>(but unlikely)          |
| Infeasible<br>( $\alpha = -\infty$ ) |      | Possible<br>(but trivial)          | Possible<br>(but unlikely)                    | Possible<br>(no solutions)          |

Table 2.2: Relationship between primal and dual SDP values.

Table 2.2 summarizes the possible types of combinations of  $\alpha$  and  $\beta$  in SDPs. Three cases are ruled out by weak duality, and out of the remaining ones in this course we mainly focus on the one where both  $\alpha$  and  $\beta$  are finite. The other ones we do not consider often, for various reasons: showing  $\alpha = \beta = -\infty$  or  $\alpha = \beta = +\infty$  leaves nothing else to be solved; SDPs with both infeasible primals and duals are likely not very practical, and semi-infinite duality gaps are uncommon.

## 2.4 Dual of the dual

Following the recipe for taking the dual has given us a new SDP that reveals many properties about the original. At this point we may wonder: what happens if we take the dual *again*? Will the dual of the dual be a third SDP with yet more insights about our problem?

In this lecture, we attempt to do just that. We start with the dual SDP in standard form, writing its constraint as an equality via a PSD variable  $Z$  and switching minimization to maximization.

$$\begin{aligned} -\beta = \text{maximize: } & \langle -B, Y \rangle \\ \text{subject to: } & \Phi^*(Y) - Z = A \\ & Y \in \text{Herm}(\mathcal{Y}) \\ & Z \in \text{Pos}(\mathcal{X}) . \end{aligned} \quad (2.31)$$

We can write any  $Y \in \text{Herm}(\mathcal{X})$  as  $Y = Y_1 - Y_2$ , where  $Y_1, Y_2 \in \text{Pos}(\mathcal{X})$ . Specifically,

$$Y = \sum_i \lambda_i v_i v_i^* = \sum_{i: \lambda_i \geq 0} \lambda_i v_i v_i^* - \sum_{i: \lambda_i < 0} (-\lambda_i) v_i v_i^* =: Y_1 - Y_2 , \quad (2.32)$$

and consequently  $\langle -B, Y \rangle = \langle B, Y_2 \rangle - \langle B, Y_1 \rangle$  and  $\Phi^*(Y) = \Phi^*(Y_1) - \Phi^*(Y_2)$ . We now have three PSD constraints  $Y_1, Y_2 \in \text{Pos}(\mathcal{Y})$  and  $Z \in \text{Pos}(\mathcal{X})$ , which when combined are equivalent to

$$X' := \begin{bmatrix} Y_1 & * & * \\ * & Y_2 & * \\ * & * & Z \end{bmatrix} \in \text{Pos}(\mathcal{Y} \oplus \mathcal{Y} \oplus \mathcal{X}) , \quad (2.33)$$

where the  $*$ s denote parts of the block matrix  $X'$  we may ignore. The proof of this as a more general statement is left to the reader in Exercise 2.6.

To write the objective function and the affine constraint of the SDP in terms of  $X'$ , define

$$B' = \begin{bmatrix} -B & 0 & 0 \\ 0 & B & 0 \\ 0 & 0 & 0 \end{bmatrix} \in \text{Pos}(\mathcal{Y} \oplus \mathcal{Y} \oplus \mathcal{X}) , \quad \begin{aligned} \Psi : \text{Herm}(\mathcal{Y} \oplus \mathcal{Y} \oplus \mathcal{X}) &\rightarrow \text{Herm}(\mathcal{X}) , \\ \Psi(X') &= \Phi^*(Y_1) - \Phi^*(Y_2) - Z . \end{aligned} \quad (2.34)$$

Notice that  $\langle B', X' \rangle = \langle B, Y_2 \rangle - \langle B, Y_1 \rangle$  is the objective function and  $\Psi(X') = A$  is the affine constraint of our SDP, meaning it is now a “primal” problem in standard form with the data  $(B', A, \Psi)$ .

We are now ready to apply the dual recipe to the dual. The resulting “primal”-dual pair is:

| Primal problem                                                                                                                                                                                        | Dual problem                                                                                                                                                                                                                                      |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| $\begin{aligned} -\beta = \text{maximize: } & \langle B', X' \rangle \\ \text{subject to: } & \Psi(X') = A \\ & X' \in \text{Pos}(\mathcal{Y} \oplus \mathcal{Y} \oplus \mathcal{X}) , \end{aligned}$ | $\begin{aligned} \gamma = \text{minimize: } & \langle A, Y' \rangle \\ \text{subject to: } & \Psi^*(Y') - Z' = B' \\ & Y' \in \text{Herm}(\mathcal{Y}) \\ & Z' \in \text{Pos}(\mathcal{Y} \oplus \mathcal{Y} \oplus \mathcal{X}) . \end{aligned}$ |

(2.35)

It remains to rewrite the resulting dual in a simpler form. The affine constraint evaluates to

$$\Psi^*(Y') - Z' = \begin{bmatrix} \Phi(Y') & 0 & 0 \\ 0 & -\Phi(Y') & 0 \\ 0 & 0 & -Y' \end{bmatrix} - \begin{bmatrix} \Lambda_1 & 0 & 0 \\ 0 & \Lambda_2 & 0 \\ 0 & 0 & \Lambda_3 \end{bmatrix} = \begin{bmatrix} -B & 0 & 0 \\ 0 & B & 0 \\ 0 & 0 & 0 \end{bmatrix} = B' . \quad (2.36)$$

We leave the verification of the adjoint  $\Psi^*$  to the reader, and note that the above form of  $Z'$  with  $\Lambda_1, \Lambda_2 \in \text{Pos}(\mathcal{Y})$  and  $\Lambda_3 \in \text{Pos}(\mathcal{X})$  is equivalent due to the constraint. Consequently, this may be split into three separate constraints:

$$\Phi(Y') - \Lambda_1 = -B, \quad \text{meaning} \quad \Phi(Y') \succeq -B, \quad (2.37)$$

$$-\Phi(Y') - \Lambda_2 = B, \quad \text{meaning} \quad \Phi(Y') \preceq -B, \quad (2.38)$$

$$-Y' - \Lambda_3 = 0, \quad \text{meaning} \quad -Y' \succeq 0. \quad (2.39)$$

As expected, the PSD inequalities (2.37) and (2.38) in different directions combine into an equality.

**Lemma 2.4.1.** *If  $X \succeq Y$  and  $X \preceq Y$ , then  $X = Y$ .*

*Proof.*  $X \succeq Y$  means that  $X - Y$  has nonnegative eigenvalues. Similarly,  $X \preceq Y$  has nonpositive eigenvalues. Thus, all the eigenvalues of  $X - Y$  are 0, meaning  $X - Y = 0$  and  $X = Y$ .  $\square$

The above property lets us conclude that  $\Phi(Y') = -B$ , and so the dual can now be written as

Dual problem

$$\begin{aligned} \gamma = \text{minimize:} \quad & \langle A, Y' \rangle \\ \text{subject to:} \quad & \Phi(Y') = -B \\ & -Y' \in \text{Pos}(\mathcal{X}). \end{aligned} \quad (2.40)$$

Let  $X = -Y'$ , switch the “minimize” to a “maximize”, then jointly negate  $-\beta$  and  $-\gamma$ . The result:

Primal problem

Dual problem

$$\begin{aligned} \beta = \text{maximize:} \quad & \langle B', X' \rangle & \gamma = \text{maximize:} \quad & \langle A, X \rangle \\ \text{subject to:} \quad & \Psi(X') = A & \text{subject to:} \quad & \Phi(X) = B \\ & X' \in \text{Pos}(\mathcal{Y} \oplus \mathcal{Y} \oplus \mathcal{X}), & & X \in \text{Pos}(\mathcal{X}). \end{aligned} \quad (2.41)$$

Note that the dual problem now exactly matches the primal standard form.

In summary, by taking the dual of the dual, we have arrived back at the primal. This means the primal and dual problems of any SDP pair are dual to each other. In retrospect this may not seem surprising—it is precisely this relationship that lets them reveal information about one another.

## 2.5 A simpler way to take the dual

Lecture 2.3 taught us one way to take the dual of an SDP: convert it into standard form (2.18), use the resulting data in the dual recipe, and simplify. This approach is formal, but may be inconvenient; it can be difficult to convert to standard form without knowing it beforehand, and the steps used to standardize and then simplify may counteract each other. To avoid doing work just to undo it shortly after, we seek a way to take the dual of any SDP directly.

### 2.5.1 Lagrangian dual

We briefly take a step back from SDPs and consider the following convex optimization problem.

$$\begin{array}{ll} \text{minimize:} & f(x) \\ \text{subject to:} & g(x) = b \\ & h(x) \leq 0, \end{array} \quad \begin{array}{l} \text{where } f \text{ is convex,} \\ g \text{ is linear,} \\ h \text{ is convex.} \end{array} \quad (2.42)$$

For convenience, we may write the above affine constraint as  $g(x) = 0$ , where  $g$  is an affine function.

Let us verify that the above is a convex optimization problem according to Definition 1.6.14.

**Lemma 2.5.1.** *The feasible region  $\{x : g(x) = b, h(x) \leq 0\}$  is a convex set.*

*Proof.* First, prove that  $\{x : g(x) = b\}$  is a convex set. Let  $\lambda \in [0, 1]$  and suppose that  $x_1$  and  $x_2$  satisfy  $g(x_1) = g(x_2) = b$ . Then,

$$g(\lambda x_1 + (1 - \lambda)x_2) = \lambda g(x_1) + (1 - \lambda)g(x_2) = \lambda b + (1 - \lambda)b = b. \quad (2.43)$$

Next, prove that  $\{x : h(x) \leq 0\}$  is convex. Let  $\lambda \in [0, 1]$  and suppose that  $x_1$  and  $x_2$  satisfy  $h(x_1) \leq 0$  and  $h(x_2) \leq 0$ . Then,

$$h(\lambda x_1 + (1 - \lambda)x_2) = \lambda h(x_1) + (1 - \lambda)h(x_2) \leq 0. \quad (2.44)$$

The intersection of these two convex sets is therefore also convex (see Lemma 1.6.3).  $\square$

We explore the *Lagrangian* dual of (2.42)—a specific notion of a dual problem (among many others) with useful insights for SDPs. For every constraint, we introduce a *Lagrange multiplier*:

$$\begin{array}{ll} \text{minimize:} & f(x) \\ \text{subject to:} & g(x) = 0 \\ & h(x) \leq 0 \end{array} \quad \left| \quad \begin{array}{l} y \\ \lambda \geq 0. \end{array} \right. \quad (2.45)$$

The new variables  $y \in \mathbb{R}$  and  $\lambda \geq 0$  are used as additional parameters to the following function.

**Definition 2.5.2.** The **Lagrangian** (function) of a convex optimization problem is

$$\mathcal{L}(x, y, \lambda) := f(x) + y \cdot g(x) + \lambda \cdot h(x). \quad (2.46)$$

We claim that the convex optimization problem (2.42) can equivalently be written as follows.

$$\begin{array}{ll} \text{minimize:} & \text{maximize: } \mathcal{L}(x, y, \lambda) \\ x & y, \lambda \geq 0 \end{array} \quad (2.47)$$

This may be seen as an interaction between two players: first, a minimizing player chooses  $x$  (same as in the original problem) to make  $\alpha(x)$  as small as possible, where

$$\alpha(x) := \text{maximize}_{y, \lambda \geq 0} \mathcal{L}(x, y, \lambda) = \begin{cases} f(x) & g(x) = 0 \text{ and } h(x) \leq 0, \\ +\infty & \text{otherwise.} \end{cases} \quad (2.48)$$

Then, a maximizing player picks  $y$  and  $\lambda \geq 0$  to make  $\mathcal{L}(x, y, \lambda)$  large, while keeping  $x$  to be the value chosen by the first player. Equivalence of (2.42) and (2.47) follows from the latter equality in (2.48), which may be seen by considering what happens to  $\alpha(x)$  when  $x$  is infeasible.

Suppose the first player picks  $x$  such that  $g(x) \neq 0$ . Then, the maximizing player may choose an arbitrarily large  $y$  with its sign matching that of  $g(x)$ , meaning  $y \cdot g(x) \rightarrow +\infty$ . Similarly, any  $x$  having  $h(x) > 0$  means a large  $\lambda \geq 0$  causes  $\lambda \cdot h(x) \rightarrow +\infty$ . That is,  $\alpha(x) \rightarrow +\infty$  for infeasible  $x$ .

*Remark.* The minimizing player is hence subject to the *implicit constraints*  $g(x) = 0$  and  $h(x) \leq 0$ .

On the other hand, if the first player picks a feasible  $x$ , the maximizing player gets no advantage from  $y, \lambda$ , and  $\alpha(x) = f(x)$ . Since  $g(x) = 0$  we have  $y \cdot g(x) = 0$  for all  $y$ . Since  $h(x) \leq 0$ , choosing a positive  $\lambda > 0$  can only decrease  $\mathcal{L}(x, y, \lambda)$ , and so the second player might as well pick  $\lambda = 0$ .

We are now ready to present the Lagrangian dual of the convex optimization problem.

$$\begin{array}{ccc}
 \text{Convex opt. problem} & & \text{Lagrangian dual} \\
 \text{minimize:} & \text{maximize:} & \\
 x & y, \lambda \geq 0 & x \\
 \mathcal{L}(x, y, \lambda) & \geq & \mathcal{L}(x, y, \lambda) .
 \end{array} \quad (2.49)$$

The only thing we did was switching the order of “maximize” and “minimize”. The inequality above is weak duality (for Lagrangian duals), which follows for *any* structure of  $\mathcal{L}(x, y, \lambda)$ . More generally, switching inf and sup yields this inequality for any function of any arguments.

*Remark.* We may argue the inequality in (2.49) makes intuitive sense. Recall from the earlier two-player interpretation that the order of optimization is evaluated left-to-right; in the original problem, the minimizing player goes first, and the maximizing one second. In this interaction, going second gives an advantage from being able to see the first player’s choice and base your own answer on it. By switching the order of the players in the dual, we instead pass this advantage to the minimizing player, decreasing the resulting value.

## 2.5.2 Lagrangian dual of SDPs in standard form

We now apply the principle of Lagrangian duality to semidefinite programs, generalizing accordingly to use operators and PSD inequalities. As a first check, let us take the Lagrangian dual of an SDP in standard form and see if the resulting dual matches what we expect.

Specifically, consider the standard form of a minimization SDP (i.e., (2.1) with “minimize” in place of “maximize”). After some manipulation to more closely match (2.45), we introduce the following Lagrange multipliers:

$$\begin{array}{ccc}
 \text{minimize:} & \langle A, X \rangle & \\
 \text{subject to:} & \Phi(X) - B = 0 & Y \\
 & -X \preceq 0 & \Lambda \succeq 0 .
 \end{array} \quad (2.50)$$

Specifically,  $Y \in \text{Herm}(\mathcal{Y})$  and  $\Lambda \in \text{Pos}(\mathcal{X})$ . We may then write this problem and its dual as

$$\begin{array}{ccc}
 \text{Minimization SDP} & & \text{Lagrangian dual} \\
 \text{minimize:} & \text{maximize:} & \\
 X & Y, \Lambda \succeq 0 & X \\
 \mathcal{L}(X, Y, \Lambda) & \geq & \mathcal{L}(X, Y, \Lambda) ,
 \end{array} \quad (2.51)$$

where we define the Lagrangian function analogously as follows.

$$\mathcal{L}(X, Y, \Lambda) := \langle A, X \rangle + \langle Y, \Phi(X) - B \rangle + \langle \Lambda, -X \rangle \quad (2.52a)$$

$$= \langle X, A \rangle + \langle \Phi^*(Y), X \rangle - \langle Y, B \rangle - \langle X, \Lambda \rangle \quad (2.52b)$$

$$= \langle X, A + \Phi^*(Y) - \Lambda \rangle - \langle B, Y \rangle. \quad (2.52c)$$

In its original form (2.52a), this function is the sum of the primal objective and the inner products of the Lagrange multipliers and the constraints. Once again, this makes feasibility of  $X$  an implicit constraint in the primal problem. Specifically, if the minimizing player chooses an  $X$  such that either  $\Phi(X) - B \neq 0$  or  $-X \not\preceq 0$ , then the maximizing player can pick an arbitrarily large  $Y$  or  $\Lambda$  to make  $\mathcal{L}(X, Y, \Lambda) \rightarrow +\infty$  (e.g., via  $Y = t(\Phi(X) - B)$  or  $\Lambda = -tvv^*$  with  $t \rightarrow +\infty$ , where  $v^*Xv < 0$ ).

By rewriting the Lagrangian function as (2.52c), we have instead isolated the variable  $X$ . Notice that this pertains to the dual problem in (2.51), where  $X$  is chosen by the second player. Analogously,  $X$  is a Lagrange multiplier of the implicit constraint  $A + \Phi^*(Y) - \Lambda = 0$  (as otherwise, the minimizing player may choose  $X = t(A + \Phi^*(Y) - \Lambda)$  with  $t \rightarrow -\infty$ , causing  $\mathcal{L}(X, Y, \Lambda) \rightarrow -\infty$ ).

When  $Y$  and  $\Lambda$  are chosen to satisfy this dual implicit constraint, the remaining part of the Lagrangian function is  $\mathcal{L}(X, Y, \Lambda) = -\langle B, Y \rangle$  (which is independent of  $X$ ). The dual is then

#### Dual problem

$$\begin{aligned} \text{maximize:} \quad & -\langle B, Y \rangle \\ \text{subject to:} \quad & -\Phi^*(Y) + \Lambda = A \\ & \Lambda \succeq 0. \end{aligned} \quad (2.53)$$

After replacing  $-Y$  by  $Y$ , this exactly matches the SDP dual in standard form (2.18) (except that “minimize” is now replaced by “maximize”). We have now effectively shown that taking the Lagrangian dual yields the same result as the dual recipe in standard form.

We may repeat this verification for the standard form SDP with “maximize”:

#### Maximization SDP

$$\begin{aligned} -\alpha = \text{minimize:} \quad & \langle -A, X \rangle \\ \text{subject to:} \quad & \Phi(X) = B \\ & X \succeq 0, \end{aligned}$$

#### Lagrangian dual

$$\begin{aligned} -\beta = \text{maximize:} \quad & \langle B, Y \rangle \\ \text{subject to:} \quad & \Phi^*(Y) \preceq -A. \end{aligned} \quad (2.54)$$

Accounting for the negative signs by switching “minimize” to “maximize” and vice versa takes us back to the standard form (2.18). In summary, the Lagrangian dual of an SDP is the SDP dual.

### 2.5.3 Lagrangian dual of example SDPs

As mentioned earlier, one reason we sometimes prefer to use Lagrangian duals is that we can take them directly without first converting into standard form, often making them easier to compute.

*Example.* Consider the following SDP and its associated Lagrangian multipliers.

$$\begin{array}{ll|l} \text{maximize:} & \langle A, X \rangle & \\ \text{subject to:} & \Phi_1(X) = B_1 & Y_1 \\ & \Phi_2(X) = B_2 & Y_2 \\ & \Phi_3(X) \succeq 0 & \Lambda \succeq 0. \end{array} \quad (2.55)$$

We aim to find its dual and to demonstrate weak duality. The Lagrangian function is

$$\mathcal{L}(X, Y_1, Y_2, \Lambda) = \langle A, X \rangle + \langle Y_1, B_1 - \Phi_1(X) \rangle + \langle Y_2, B_2 - \Phi_2(X) \rangle + \langle \Lambda, \Phi_3(X) \rangle \quad (2.56a)$$

$$= \langle X, A - \Phi_1^*(Y_1) - \Phi_2^*(Y_2) + \Phi_3^*(\Lambda) \rangle + \langle Y_1, B_1 \rangle + \langle Y_2, B_2 \rangle. \quad (2.56b)$$

*Remark.* Since  $Y_1$  is Hermitian, we can use either  $\langle Y_1, B_1 - \Phi_1(X) \rangle$  or  $\langle Y_1, \Phi_1(X) - B_1 \rangle$ .

*Remark.* Since we are maximizing the primal SDP with  $\Phi_3(X) \succcurlyeq 0$  and  $\Lambda \succcurlyeq 0$ , the term  $\langle \Lambda, \Phi_3(X) \rangle$  must be preceded by a plus sign.

We have isolated  $X$  in (2.56b), the other term in its inner product being the implicit constraint. The remaining part of the Lagrangian function becomes the dual objective.

#### Dual problem

$$\begin{aligned} \text{minimize: } & \langle B_1, Y_1 \rangle + \langle B_2, Y_2 \rangle \\ \text{subject to: } & A - \Phi_1^*(Y_1) - \Phi_2^*(Y_2) + \Phi_3^*(\Lambda) = 0 \\ & \Lambda \succcurlyeq 0. \end{aligned} \quad (2.57)$$

We may verify weak duality by computing the objective gap:

$$\langle B_1, Y_1 \rangle + \langle B_2, Y_2 \rangle - \langle A, X \rangle = \langle X, \Phi_1^*(Y_1) + \Phi_2^*(Y_2) - A \rangle \quad (2.58a)$$

$$= \langle X, \Phi_3^*(\Lambda) \rangle = \langle \Phi_3(X), \Lambda \rangle \succcurlyeq 0. \quad (2.58b)$$

*Example.* For an example of an SDP with data that is more concrete, consider:

$$\begin{array}{ll|l} \text{maximize: } & \langle E_{1,1}, X \rangle & \\ \text{subject to: } & \text{Tr}(X) = 1 & y \\ & X \succcurlyeq 0 & \Lambda \succcurlyeq 0. \end{array} \quad (2.59)$$

This SDP is already in standard form, but we may as well take the Lagrangian approach anyway. Note also that  $y$  is now a scalar since  $\mathcal{Y} = \mathbb{C}^1$ . The Lagrangian function is

$$\mathcal{L}(X, y, \Lambda) = \langle E_{1,1}, X \rangle + y(1 - \text{Tr}(X)) + \langle \Lambda, X \rangle \quad (2.60a)$$

$$= \langle X, E_{1,1} - y\mathbb{1} + \Lambda \rangle + y. \quad (2.60b)$$

Once again,  $\langle \Lambda, X \rangle$  is preceded by a plus sign since we are maximizing. The dual is then

#### Dual problem

$$\begin{aligned} \text{minimize: } & y \\ \text{subject to: } & E_{1,1} - y\mathbb{1} + \Lambda = 0 \\ & \Lambda \succcurlyeq 0, \end{aligned} \quad (2.61)$$

where the two constraints may succinctly be written as  $y\mathbb{1} \succcurlyeq E_{1,1}$ . Let us verify weak duality:

$$y - \langle E_{1,1}, X \rangle = \text{Tr}(X) \cdot y - \langle X, E_{1,1} \rangle = \langle X, y\mathbb{1} - E_{1,1} \rangle \geq 0. \quad (2.62)$$

*Example.* With the Lagrangian dual, we do not need to combine the constraints or variables.

$$\begin{array}{ll|l}
 \text{maximize:} & \langle A_1, X_1 \rangle + \langle A_2, X_2 \rangle & \\
 \text{subject to:} & \Phi_1(X_1) = B_1 & Y_1 \\
 & \Phi_2(X_2) = B_2 & Y_2 \\
 & X_1 + X_2 \preceq \mathbb{1} & \Lambda_1 \succeq 0 \\
 & X_1 \succeq 0 & \Lambda_2 \succeq 0.
 \end{array} \tag{2.63}$$

The Lagrangian is a function of *all* the primal variables and Lagrange multipliers.

$$\mathcal{L}(X_1, X_2, Y_1, Y_2, \Lambda_1, \Lambda_2) = \langle A_1, X_1 \rangle + \langle A_2, X_2 \rangle + \langle Y_1, B_1 - \Phi_1(X_1) \rangle \tag{2.64a}$$

$$+ \langle Y_2, B_2 - \Phi_2(X_2) \rangle + \langle \Lambda_1, \mathbb{1} - X_1 - X_2 \rangle + \langle \Lambda_2, X_1 \rangle \tag{2.64b}$$

$$= \langle X_1, A_1 - \Phi_1^*(Y_1) - \Lambda_1 + \Lambda_2 \rangle + \langle X_2, A_2 - \Phi_2^*(Y_2) - \Lambda_1 \rangle \tag{2.64c}$$

$$+ \langle B_1, Y_1 \rangle + \langle B_2, Y_2 \rangle + \langle \mathbb{1}, \Lambda_1 \rangle. \tag{2.64d}$$

The two original variables in the primal now correspond to two implicit constraints in the dual.

#### Dual problem

$$\begin{array}{ll}
 \text{minimize:} & \langle B_1, Y_1 \rangle + \langle B_2, Y_2 \rangle + \langle \mathbb{1}, \Lambda_1 \rangle \\
 \text{subject to:} & A_1 - \Phi_1^*(Y_1) - \Lambda_1 + \Lambda_2 = 0 \\
 & A_2 - \Phi_2^*(Y_2) - \Lambda_1 = 0 \\
 & \Lambda_1, \Lambda_2 \succeq 0,
 \end{array} \tag{2.65}$$

which, with some minor rearrangements and a relabeling of  $\Lambda_1 = \Lambda$ , may be written as

#### Dual problem

$$\begin{array}{ll}
 \text{minimize:} & \langle B_1, Y_1 \rangle + \langle B_2, Y_2 \rangle + \text{Tr}(\Lambda) \\
 \text{subject to:} & \Phi_1^*(Y_1) + \Lambda \succeq A_1 \\
 & \Phi_2^*(Y_2) + \Lambda = A_2 \\
 & \Lambda \succeq 0.
 \end{array} \tag{2.66}$$

Utilizing all the constraints of the primal and dual allows us to verify weak duality.

$$\langle B_1, Y_1 \rangle + \langle B_2, Y_2 \rangle + \text{Tr}(\Lambda) - \langle A_1, X_1 \rangle - \langle A_2, X_2 \rangle \tag{2.67a}$$

$$= \langle \Phi_1(X_1), Y_1 \rangle + \langle \Phi_2(X_2), Y_2 \rangle + \langle \mathbb{1}, \Lambda \rangle - \langle X_1, A_1 \rangle - \langle X_2, A_2 \rangle \tag{2.67b}$$

$$= \langle X_1, \Phi_1^*(Y_1) - A_1 \rangle + \langle X_2, \Phi_2^*(Y_2) - A_2 \rangle + \langle \mathbb{1}, \Lambda \rangle \tag{2.67c}$$

$$\geq \langle X_1, -\Lambda \rangle + \langle X_2, -\Lambda \rangle + \langle \mathbb{1}, \Lambda \rangle = \langle \mathbb{1} - X_1 - X_2, \Lambda \rangle \geq 0. \tag{2.67d}$$

*Remark.* Primal variables become dual constraints; primal constraints become dual variables.

*Example.* Lagrangian duals also help when we have fewer constraints than the standard form.

$$\begin{array}{ll|l}
 \text{minimize:} & \text{Tr}(X) & \\
 \text{subject to:} & X \succeq \mathbb{1} & \Lambda \succeq 0.
 \end{array} \tag{2.68}$$

Since we are minimizing with  $X - \mathbb{1} \succcurlyeq 0$  and  $\Lambda \succcurlyeq 0$ , the corresponding term needs a negative sign.

$$\mathcal{L}(X, \Lambda) = \text{Tr}(X) - \langle \Lambda, X - \mathbb{1} \rangle = \langle X, \mathbb{1} - \Lambda \rangle + \langle \Lambda, \mathbb{1} \rangle. \quad (2.69)$$

The implicit constraint is hence  $\mathbb{1} - \Lambda = 0$ , and the dual is

#### Dual problem

$$\begin{aligned} \text{maximize: } & \text{Tr}(\Lambda) \\ \text{subject to: } & \Lambda = \mathbb{1} \\ & \Lambda \succcurlyeq 0. \end{aligned} \quad (2.70)$$

Note that the PSD constraint is redundant. Let us verify weak duality:

$$\text{Tr}(X) - \text{Tr}(\Lambda) = \langle \mathbb{1}, X \rangle - \langle \Lambda, \mathbb{1} \rangle = \langle \Lambda, X \rangle - \langle \Lambda, \mathbb{1} \rangle = \langle \Lambda, X - \mathbb{1} \rangle \succcurlyeq 0. \quad (2.71)$$

Both the primal and the dual in this case are very simple, and have obvious solutions.

For an additional example of taking duals, see Exercise 2.7.

## 2.6 Optimality conditions: Slater's theorem

When we introduced SDP duals in Lecture 2.3, we have seen that they can bound optimal objective function values thanks to weak duality  $\alpha \leq \beta$ . In particular, finding *any* pair of feasible primal and dual solutions places both  $\alpha$  and  $\beta$  in the closed interval between the resulting values. A special instance of this is when the values coincide, in which case we exactly know  $\alpha$  and  $\beta$ , and further know that the feasible solutions we found are optimal. However, this is only possible when  $\alpha = \beta$ .

In this lecture, we look at some conditions that can tell us that  $\alpha$  and  $\beta$  are equal.

### 2.6.1 Slater's theorem

Specifically, the considered conditions have to do with strict feasibility. Recall that for an SDP in standard form (2.18), the primal is strictly feasible when there exists a feasible  $X \in \text{Pd}(\mathcal{X})$ .

**Theorem 2.6.1** (Slater's theorem). *If the primal is strictly feasible, then we have **strong duality** (meaning  $\alpha = \beta$ ). Moreover, if  $\alpha$  is finite, then  $\beta$  is attained (i.e., an optimal dual solution exists).*

*Proof.* We know that the primal is feasible, so  $\alpha > -\infty$ . If  $\alpha = +\infty$ , then the dual is infeasible, and  $\beta = +\infty = \alpha$ . It remains to prove the case when  $\alpha$  is finite. Define the following sets.

$$Q := \{Y \in \text{Herm}(\mathcal{Y}) : \exists X \in \text{Herm}(\mathcal{X}), \Phi(X) = Y\}, \quad (2.72)$$

$$M := \{(S, V, t) \in \text{Herm}(\mathcal{X}) \times Q \times \mathbb{R} : \exists X \succcurlyeq S, V = B - \Phi(X), \langle A, X \rangle \geq t\}, \quad (2.73)$$

$$N := \{(0, 0, s) \in \text{Herm}(\mathcal{X}) \times Q \times \mathbb{R} : s > \alpha\}. \quad (2.74)$$

Note that  $Q$  is the *image* of  $\Phi$ . Primal feasibility implies  $B \in Q$ , and hence  $Q \neq \emptyset$ . The last parameter of the sets  $M$  and  $N$  represents values (respectively) below and above the value  $\alpha$  of the primal SDP, with  $S$  and  $V$  being the amounts of infeasibility/gap in each constraint. By optimality of  $\alpha$ , we then have  $M \cap N = \emptyset$ . By the separating hyperplane theorem (see

Theorem 1.6.7), there exist  $(Z, Y, \lambda) \in \text{Herm}(\mathcal{X}) \times Q \times \mathbb{R}$  (not all zero) and  $c \in \mathbb{R}$  such that

$$(S, V, t) \in M \quad \Rightarrow \quad \langle Z, S \rangle + \langle Y, V \rangle + \lambda t \leq c, \quad (2.75)$$

$$(S, V, t) \in N \quad \Rightarrow \quad \langle Z, S \rangle + \langle Y, V \rangle + \lambda t \geq c. \quad (2.76)$$

As  $S = 0$  and  $V = 0$  in  $N$ , (2.76) simplifies to  $\lambda s \geq c$  for all  $s > \alpha$ , implying  $\lambda \geq 0$  and  $c \leq \lambda \alpha$ . In summary, we now know that there exist  $Z \in \text{Herm}(\mathcal{X})$ ,  $Y \in Q$ , and  $\lambda \geq 0$  such that

$$(Z, Y, \lambda) \neq (0, 0, 0), \quad (S, V, t) \in M \quad \Rightarrow \quad \langle Z, S \rangle + \langle Y, V \rangle + \lambda t \leq \lambda \alpha. \quad (2.77)$$

Prove that  $Z \succeq 0$ . For any feasible  $X$  (which exists) and  $S \preceq 0$ , we have  $(S, 0, \langle A, X \rangle) \in M$ , by (2.77) implying  $\langle Z, S \rangle + \lambda \langle A, X \rangle \leq \lambda \alpha$ . If  $Z$  has a negative eigenvalue, we can choose  $S \preceq 0$  large enough to violate this bound by making the  $\langle Z, S \rangle$  term arbitrarily large.

Suppose  $\lambda = 0$  for purpose of a contradiction. For any strictly feasible  $X$  (which exists), we have  $(X, 0, \langle A, X \rangle) \in M$ , by (2.77) implying  $\langle Z, X \rangle \leq 0$ . Since  $Z \succeq 0$  and  $X \in \text{Pd}(\mathcal{X})$ , we must have  $Z = 0$ . Notice that  $(S, B - \Phi(S), \langle A, S \rangle) \in M$  for any  $S \in \text{Herm}(\mathcal{X})$ , by (2.77) implying  $\langle Y, B - \Phi(S) \rangle \leq 0$ . Since  $Y \in Q$ , we must have  $Y = 0$ , as otherwise we can find  $S \in \text{Herm}(\mathcal{X})$  to violate the inequality. Thus,  $(Z, Y, \lambda) = (0, 0, 0)$ , a contradiction.

Knowing  $\lambda > 0$  allows us to rescale  $Z$  and  $Y$  in (2.77) to eliminate  $\lambda$ . We now have that

$$\exists Z \in \text{Pos}(\mathcal{X}), \quad Y \in Q, \quad (S, V, t) \in M \quad \Rightarrow \quad \langle Z, S \rangle + \langle Y, V \rangle + t \leq \alpha. \quad (2.78)$$

For any  $S \in \text{Herm}(\mathcal{X})$  we have  $(S, B - \Phi(S), \langle A, S \rangle) \in M$ , and  $\langle S, Z - \Phi^*(Y) + A \rangle + \langle B, Y \rangle \leq \alpha$  by (2.78). We have  $Z - \Phi^*(Y) + A = 0$ , as otherwise some Hermitian  $S$  violates this bound. Thus,  $(Y, Z)$  is dual feasible with  $\langle B, Y \rangle \leq \alpha$ . Combining with weak duality, we now have  $\alpha \leq \beta \leq \langle B, Y \rangle \leq \alpha$ , meaning  $\alpha = \beta$  and  $(Y, Z)$  is an optimal dual solution.  $\square$

Since the dual of the dual is the primal (see Lecture 2.4), we have the following corollary.

**Corollary 2.6.2.** *If the dual is strictly feasible, then we have strong duality (meaning  $\alpha = \beta$ ). Moreover, if  $\beta$  is finite, then  $\alpha$  is attained (i.e., an optimal primal solution exists).*

**Remark.** Strict feasibility for the dual means that there exists a  $Z \in \text{Pd}(\mathcal{X})$  that is feasible in (2.19).

**Corollary 2.6.3.** *If the primal and dual are both strictly feasible, then we have strong duality (meaning  $\alpha = \beta$ ) and both  $\alpha$  and  $\beta$  are attained (optimal solutions of both the primal and the dual exist).*

Note that  $\alpha$  and  $\beta$  in this case must be finite; otherwise, one of the problems would be infeasible.

In practice, we may quite often find that  $\alpha = \beta$  after solving an SDP. Now, we have the ability to show the same without the need to find optimal solutions or compute the exact values.

**Example.** Recall the following primal-dual SDP pair, where  $D(\mathcal{X}) := \{X \in \text{Pos}(\mathcal{X}) : \text{Tr}(X) = 1\}$ .

| Primal problem                   | Dual problem                          |        |
|----------------------------------|---------------------------------------|--------|
| maximize: $\langle H, X \rangle$ | minimize: $y$                         | (2.79) |
| subject to: $X \in D(X)$ ,       | subject to: $y\mathbb{1} \succeq H$ . |        |

The following primal and dual solutions are both strictly feasible:

$$X = \frac{1}{\dim(\mathcal{X})} \mathbb{1}_{\mathcal{X}}, \quad y = \lambda_{\max}[H] + 1. \quad (2.80)$$

By Corollary 2.6.3, we have  $\alpha = \beta$  and both problems attain an optimal solution.

*Remark.* This approach is simple, but it does not tell us what  $\alpha$ ,  $\beta$ , or the optimal solutions are.

## 2.6.2 Complementary Slackness

First, let us prove a technical result about positive semidefinite matrices.

**Lemma 2.6.4.** *Let  $X, Y \succcurlyeq 0$ . We have  $\langle X, Y \rangle = 0$  if and only if  $XY = 0$ .*

*Proof.* Since  $X$  and  $Y$  are Hermitian,  $X = X^*$  and  $Y = Y^*$ . Show the two directions separately:

$\Leftarrow$ : Suppose  $XY = 0$ . Then,  $\langle X, Y \rangle = \text{Tr}(X^*Y) = \text{Tr}(XY) = 0$ .

$\Rightarrow$ : Suppose  $\langle X, Y \rangle = 0$ . Take the spectral decompositions  $X = \sum_i \lambda_i v_i v_i^*$  and  $Y = \sum_j \mu_j w_j w_j^*$ .

$$0 = \langle X, Y \rangle = \sum_{i,j} \lambda_i \mu_j \text{Tr}(v_i v_i^* w_j w_j^*) = \sum_{i,j} \lambda_i \mu_j \langle v_i^* | w_j \rangle \langle w_j^* | v_i \rangle = \sum_{i,j} \lambda_i \mu_j |\langle v_i, w_j \rangle|^2. \quad (2.81)$$

Since  $\lambda_i, \mu_j, |\langle v_i, w_j \rangle|^2 \geq 0$ , equality can only hold if  $\lambda_i \mu_j |\langle v_i, w_j \rangle|^2 = 0$  for each pair  $i, j$ . Then,

$$XY = \sum_{i,j} \lambda_i \mu_j v_i (v_i^* w_j) w_j^* = 0, \quad (2.82)$$

since for every  $i, j$  we have either  $\lambda_i = 0$ ,  $\mu_j = 0$ , or  $\langle v_i, w_j \rangle = 0$ .  $\square$

We now introduce a condition that could help us find or characterize optimal solutions.

**Theorem 2.6.5** (Complementary slackness). *Suppose  $X$  is primal feasible and  $Y$  is dual feasible.*

$$\langle A, X \rangle = \langle B, Y \rangle \quad \Leftrightarrow \quad \Phi^*(Y)X = AX. \quad (2.83)$$

*Remark.* Equivalently, we may use  $X\Phi^*(Y) = XA$  as the right-hand-side condition.

Specifically, *complementary slackness* refers to the equality on the right-hand side of (2.83).

*Proof.* Since  $X$  and  $Y$  are feasible, we have  $\Phi(X) = B$ ,  $X \succcurlyeq 0$ , and  $\Phi^*(Y) - A \succcurlyeq 0$ .

$$\langle A, X \rangle = \langle B, Y \rangle \Leftrightarrow \langle \Phi^*(Y) - A, X \rangle = 0 \quad (\text{since } B = \Phi(X)), \quad (2.84a)$$

$$\Leftrightarrow (\Phi^*(Y) - A)X = 0 \quad (\text{by Lemma 2.6.4 with } \Phi^*(Y) - A, X \succcurlyeq 0). \quad (2.84b)$$

After rearranging, we get  $\Phi^*(Y)X = AX$  and the proof is complete.  $\square$

Suppose we find feasible solutions  $X$  and  $Y$  such that  $\Phi^*(Y)X = AX$ . Then, Theorem 2.6.5 tells us those solutions are optimal with  $\alpha = \beta$ , which we may easily calculate. Moreover, if the primal and dual are both strictly feasible, Theorem 2.6.1 tells us that such solutions must exist.

## 2.7 Clark's theorem and ascent directions

Recall the following properties of bounded SDPs (having  $\alpha < +\infty$ ) from Definition 2.1.3:

- Infeasible SDPs are considered bounded (since  $\alpha = -\infty < +\infty$ ).
- If the feasible region  $\mathcal{A}$  is a bounded set, then the SDP cannot be unbounded (and  $\alpha < +\infty$ ).

*Proof.* Suppose  $\mathcal{A} \subseteq B_r(0)$ ; that is,  $X \in \mathcal{A}$  when  $\|X\|_\infty \leq r$ . Then,

$$\alpha = \underset{X \in \mathcal{A}}{\text{maximize:}} \langle A, X \rangle \leq \underset{X : \|X\|_\infty \leq r}{\text{maximize:}} |\langle A, X \rangle| = r \cdot \|A\|_1 < +\infty. \quad (2.85)$$

We note that we have the flexibility of choosing the norm of  $B_r(0)$  in the above.  $\square$

- $\alpha$  could be finite even when the feasible region  $\mathcal{A}$  is unbounded (see the SDP below).

$$\begin{aligned} \alpha = \underset{t}{\text{maximize:}} \quad & t \\ \text{subject to:} \quad & t \leq 1. \end{aligned} \quad (2.86)$$

In other words, unbounded SDPs have unbounded feasible regions, but the opposite is not always true. In this lecture we look at unbounded feasible regions: their causes, and how to mitigate them.

### 2.7.1 Clark's Theorem

As it turns out, we have to deal with unbounded feasible regions quite frequently; the same condition that gives us strong duality by Theorem 2.6.1 also implies an unbounded feasible region.

**Theorem 2.7.1** (Clark's theorem). *If the primal is strictly feasible, then either the primal or the dual has an unbounded feasible region.*

*Proof.* We first prove the statement in the special case of  $A = \epsilon \mathbb{1}$  for some  $\epsilon > 0$ .

| Primal problem                                     | Dual problem                                               |        |
|----------------------------------------------------|------------------------------------------------------------|--------|
| maximize: $\langle \epsilon \mathbb{1}, X \rangle$ | minimize: $\langle B, Y \rangle$                           | (2.87) |
| subject to: $\Phi(X) = B$                          | subject to: $\Phi^*(Y) \succcurlyeq \epsilon \mathbb{1}$ . |        |
| $X \in \text{Pos}(\mathcal{X})$ ,                  |                                                            |        |

If the primal feasible region  $\mathcal{A}$  is unbounded, we are done—so, suppose otherwise (and aim to show that  $\mathcal{B}$  is unbounded). From  $\mathcal{A}$  being bounded we know that  $\alpha$  is finite, which together with the initial assumption of the primal being strictly feasible lets us apply Theorem 2.6.1 to show the existence of an optimal dual solution  $Y$ . Consider the parametrized solution  $Y(t) = tY$  with  $t \geq 1$ , which is feasible since  $\Phi^*(tY) = t\Phi^*(Y) \succcurlyeq t\epsilon \mathbb{1} \succcurlyeq \epsilon \mathbb{1}$ . Additionally, we see that  $Y \neq 0$  since  $\Phi^*(0) = 0 \not\succcurlyeq \epsilon \mathbb{1}$ , and hence  $\|Y\| > 0$ . We then have  $\|tY\| = t\|Y\| \rightarrow +\infty$  as  $t \rightarrow +\infty$ , and so the dual feasible region  $\mathcal{B}$  is unbounded.

We now return to the general case of a standard form SDP (2.18) with  $A \in \text{Herm}(\mathcal{X})$ . Let  $\epsilon = \lambda_{\max}[A]$ ; from the result in the special case we have an unbounded set of solutions  $\{Y(t) : t \geq 1\}$  that are feasible in the dual of (2.87), meaning  $\Phi^*(Y(t)) \succcurlyeq \epsilon \mathbb{1}$ . We have

$\lambda_{\max}[A] \cdot \mathbb{1} \succcurlyeq A$  from taking spectral decompositions, so  $\Phi^*(Y(t)) \succcurlyeq \lambda_{\max}[A] \cdot \mathbb{1} \succcurlyeq A$ , and hence the same unbounded set of solutions is feasible for the standard form dual.  $\square$

Since the dual of the dual is the primal (see Lecture 2.4), we have the following corollary.

**Corollary 2.7.2.** *If the dual is strictly feasible, then either the primal or the dual has an unbounded feasible region.*

## 2.7.2 Bounding feasible regions

We sometimes wish that the feasible regions of SDPs are bounded. Clark's theorem tells us that dealing with at least one unbounded feasible region is necessary for strictly feasible SDPs. Fortunately, strictly feasible solutions can also be used to bound these feasible regions—in such a way that we never cut off any optimal solutions.

Suppose  $\hat{X} \succ 0$  is strictly primal feasible and  $\hat{Y}$  is strictly dual feasible (i.e.,  $\hat{Z} = \Phi^*(\hat{Y}) - A \succ 0$ ). Consider adding to the standard form primal SDP (2.1) the constraint  $\langle X, \hat{Z} \rangle \leq 2\langle \hat{X}, \hat{Z} \rangle$ , to get:

Primal problem

$$\begin{aligned} \hat{\alpha} = \text{maximize: } & \langle A, X \rangle \\ \text{subject to: } & \Phi(X) = B \\ & \langle X, \hat{Z} \rangle \leq 2\langle \hat{X}, \hat{Z} \rangle \\ & X \in \text{Pos}(\mathcal{X}) . \end{aligned} \quad (2.88)$$

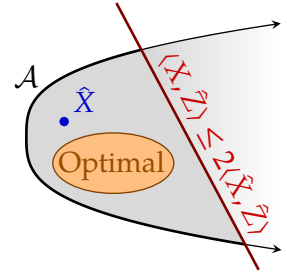


Figure 2.2: Bounding the feasible region.

Does the new constraint affect the SDP or its value? Note that strict feasibility implies by Slater's theorem (specifically, Corollary 2.6.3) that there exists at least one optimal primal solution  $X'$ . But, does the new constraint cut off  $X'$  (or other optimal primal solutions)?

**Lemma 2.7.3.** *Suppose  $\hat{X}$  and  $\hat{Y}$  are strictly feasible. Both SDPs (2.1) and (2.88) have the same set of optimal solutions. Moreover, (2.88) has a bounded feasible region and is still strictly feasible.*

We aim to show a situation as in Figure 2.2; the set of optimal solutions and  $\hat{X}$  are included within the constraint, but it cuts off the unbounded part of the original feasible region.

*Proof.* Let  $X'$  be optimal in (2.1). Feasibility implies  $\Phi(X') = B$  and  $\Phi(\hat{X}) = B$ , and so

$$\langle X', \hat{Z} \rangle = \langle B, \hat{Y} \rangle - \langle A, X' \rangle, \quad \langle \hat{X}, \hat{Z} \rangle = \langle B, \hat{Y} \rangle - \langle A, \hat{X} \rangle. \quad (2.89)$$

Since  $X'$  is optimal,  $\langle A, X' \rangle \geq \langle A, \hat{X} \rangle$ . Combining with (2.89) gives  $\langle X', \hat{Z} \rangle \leq \langle \hat{X}, \hat{Z} \rangle \leq 2\langle \hat{X}, \hat{Z} \rangle$ , meaning  $X'$  is feasible in (2.88) and so  $\hat{\alpha} \geq \langle A, X' \rangle = \alpha$ . Since (2.88) has the same objective function as (2.1) but a smaller feasible region, we must also have  $\hat{\alpha} \leq \alpha$ . Combined, we have  $\hat{\alpha} = \langle A, X' \rangle$ , meaning  $X'$  is also optimal in the constrained SDP (2.88).

Conversely, suppose  $X''$  is optimal in the constrained SDP (2.88).  $X''$  is then feasible in the original SDP (2.1), and by the earlier computation  $\alpha = \hat{\alpha} = \langle A, X'' \rangle$  it is optimal there as well. This lets us conclude that the sets of optimal solutions of the two SDPs are equal.

Since  $\hat{X} \succ 0$  and  $\hat{Z} \succ 0$ , one can show that  $\langle \hat{X}, \hat{Z} \rangle > 0$  (see Exercise 2.8). In addition to  $\hat{X}$  being strictly feasible in the original SDP, we then have  $\langle \hat{X}, \hat{Z} \rangle < 2\langle \hat{X}, \hat{Z} \rangle$ , meaning  $\hat{X}$  is also strictly feasible with respect to the new constraint and hence for the whole of (2.88).

Finally, we show that the feasible region of the constrained SDP is bounded. We first note a general property: for any  $X \in \text{Pos}(\mathcal{X})$ , we have  $\lambda_i[X] \geq 0$  and hence

$$\text{Tr}(X) = \sum_i \lambda_i[X] = \sum_i |\lambda_i[X]| = \sum_i \sigma_i[X] = \|X\|_1. \quad (2.90)$$

Suppose  $X$  is feasible in (2.88). Since  $X$  satisfies the added constraint, we have

$$2\langle \hat{X}, \hat{Z} \rangle \geq \langle X, \hat{Z} \rangle \geq \langle X, \lambda_{\min}[\hat{Z}] \cdot \mathbb{1} \rangle = \lambda_{\min}[\hat{Z}] \cdot \text{Tr}(X) = \lambda_{\min}[\hat{Z}] \cdot \|X\|_1. \quad (2.91)$$

$\lambda_{\min}[\hat{Z}]$  and  $\langle \hat{X}, \hat{Z} \rangle$  are constants independent of  $X$ , so the trace norm of  $X$  is upper bounded by a constant. As this applies to any feasible  $X$ , the feasible region of (2.88) is bounded.  $\square$

*Remark.* The proof follows with any number greater than 1 in place of 2 in the constraint of (2.88).

### 2.7.3 Ascent directions

Next, we formalize a way of checking for unbounded primal and hence infeasible dual problems.

**Definition 2.7.4.** An **ascent direction**  $D$  of a primal SDP in standard form (2.1) is such that:

$$\langle A, D \rangle > 0 \quad (\text{has a positive objective function value}), \quad (2.92a)$$

$$\Phi(D) = 0 \quad (\text{is in the kernel of } \Phi), \quad (2.92b)$$

$$D \in \text{Pos}(\mathcal{X}) \quad (\text{is positive semidefinite}). \quad (2.92c)$$

Moving in this direction increases the objective function value while retaining feasibility.

**Lemma 2.7.5.** *If the primal is feasible and has an ascent direction, then it is unbounded.*

*Proof.* Suppose  $X$  is feasible. Then, for all  $t \geq 0$  we have  $X + tD \succcurlyeq 0$  and

$$\langle A, X + tD \rangle = \langle A, X \rangle + t \cdot \langle A, D \rangle \rightarrow +\infty \quad \text{as } t \rightarrow \infty, \quad (2.93a)$$

$$\Phi(X + tD) = \Phi(X) + t \cdot \Phi(D) = \Phi(X) = B. \quad (2.93b)$$

This means  $X + tD$  is feasible with an objective function value that grows unbounded.  $\square$

On the other hand, feasibility of the primal is not required to show that the dual is infeasible.

**Lemma 2.7.6.** *If the primal has an ascent direction, then the dual is infeasible.*

*Proof.* Since  $B$  only appears in the objective of the standard form dual (2.18), we may change  $B$  without affecting the dual feasible region. Let  $B = 0$ , making  $X = 0$  feasible. The ascent direction then implies the primal is unbounded, meaning the dual must be infeasible.  $\square$

We may wonder if the converse of the above is true, meaning whether an ascent direction is necessary for the dual to be infeasible. This turns out to be true when the dual is *very* infeasible.

| Modified primal                                                                                                                               | Modified dual                                                                                          |        |
|-----------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------|--------|
| $\alpha' = \text{maximize: } \langle A, X \rangle$<br>subject to: $\Phi(X) = 0$<br>$\text{Tr}(X) \leq 1$<br>$X \in \text{Pos}(\mathcal{X})$ , | $\beta' = \text{minimize: } t$<br>subject to: $\Phi^*(Y) + t\mathbb{1} \succcurlyeq A$<br>$t \geq 0$ . | (2.94) |

In particular, very infeasible means that  $\beta' > 0$  above. In other words, it means that the dual needs a lot of help to be feasible. If  $t \rightarrow 0$  in the limit, we have  $\beta' = 0$  and the dual is *almost* feasible.

**Lemma 2.7.7.** *If the dual is infeasible with  $\beta' > 0$  in (2.94), then there exists an ascent direction.*

*Proof.* The modified dual SDP in (2.94) is strictly feasible. In particular, for any  $Y$  we may pick a sufficiently large  $t > 0$  such that  $\Phi^*(Y) + t\mathbb{1} \succ A$ . Since  $\beta'$  is bounded below by 0 and the dual is feasible,  $\beta'$  is finite. By Corollary 2.6.2, we then have  $\alpha' = \beta' > 0$  and an optimal solution  $D$  exists for the modified primal. We then have  $D \in \text{Pos}(\mathcal{X})$  with  $\langle A, D \rangle > 0$  and  $\Phi(D) = 0$ , meaning  $D$  is an ascent direction.  $\square$

## 2.7.4 General approach for analyzing SDPs

We have now introduced all the necessary basic tools for analyzing SDPs. It may seem a bit daunting at this point to know which tool fits best to a particular problem—while this mainly comes from experience, there is a general sequence of steps we may follow:

1. Check if SDP is obviously infeasible, unbounded, etc. It may not even be worth investigating.
2. Calculate the dual (see Lecture 2.5).
3. Check for strictly feasible solutions. Infer whatever you can about optimal solutions, values, and the duality gap. Even if you just want to solve them numerically, looking for strictly feasible solutions can be a good predictor of how well your algorithm performs.

We will take this approach when looking at quantum applications in Chapter 3.

## Chapter 2 exercises

**Exercise 2.1.** Prove that an SDP in standard form (2.1) with “minimize” is a convex optimization problem. That is, show that its feasible region  $\mathcal{A}$  is a convex set and  $\langle A, X \rangle$  is a convex function.

**Exercise 2.2.** Prove that the set of optimal solutions of any SDP is a convex set.

**Exercise 2.3.** Find an SDP with a finite  $\alpha$ , but an unbounded set of optimal solutions.

**Exercise 2.4.** Prove that if the feasible region is nonempty and bounded, then  $\alpha$  is finite and attained (i.e., there exists an optimal solution).

*Hint:* First prove that the feasible region is closed.

**Exercise 2.5.** Put the following SDP into standard form:

$$\begin{aligned} &\text{maximize: } a + b + c \\ &\text{subject to: } \begin{pmatrix} a & 1 & 1 \\ 1 & b & 1 \\ 1 & 1 & c \end{pmatrix} \in \text{Pos}(\mathbb{C}^3). \end{aligned} \quad (2.95)$$

*Hint:* Consider whether complex solutions are needed.

**Exercise 2.6.** Prove that a matrix being positive semidefinite is equivalent to its diagonal blocks being positive semidefinite. Specifically, let  $\mathcal{X} = \mathcal{Y}_1 \oplus \mathcal{Y}_2 \oplus \dots \oplus \mathcal{Y}_n$  and

$$X = \begin{bmatrix} Y_1 & * & \cdots & * \\ * & Y_2 & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \cdots & Y_n \end{bmatrix} \in \text{Pos}(\mathcal{X}), \quad (2.96)$$

and show that

1. If the  $*$ s are such that  $X \in \text{Pos}(\mathcal{X})$ , then  $Y_i \in \text{Pos}(\mathcal{Y}_i)$  for all  $i \in \{1, \dots, n\}$ .
2. If  $Y_i \in \text{Pos}(\mathcal{Y}_i)$  for all  $i \in \{1, \dots, n\}$ , then there exists a choice of  $*$ s such that  $X \in \text{Pos}(\mathcal{X})$ .

**Exercise 2.7.** Take the dual of the SDP

$$\begin{aligned} &\text{maximize: } \langle A, X \rangle \\ &\text{subject to: } C \preceq X \preceq D, \end{aligned} \quad (2.97)$$

where  $A, C$ , and  $D$  are Hermitian. Verify that weak duality holds.

**Exercise 2.8.** Let  $X, Y \in \text{Pd}(\mathcal{X})$  and prove that this implies  $\langle X, Y \rangle > 0$ .

**Exercise 2.9.** Give a first-principles proof for the dual version of Clark’s theorem (Corollary 2.7.2).

*Hint:* Look for a descent direction via  $\beta'$  when  $\beta' < 0$  and finite.

## Chapter 3

# Quantum Information

We now reach the second half of the course, in which we apply various concepts from semidefinite programming to quantum information. This is far from the only application of SDPs, although the positive semidefinite nature of quantum objects often makes this application straightforward.

Some sets used as examples in previous lectures (e.g.,  $D(\mathcal{X})$ ,  $\text{Eff}(\mathcal{X})$ ) make a more prominent reappearance here. This is no coincidence—we had gently increased mathematical familiarity with these concepts, and are now ready to give them operational meanings. This chapter aims beyond merely introducing quantum information; readers with more interest in the semidefinite programming aspect may appreciate it for grounding SDP concepts in practical problems.

For a more thorough resource on using SDPs in quantum information, see:

- Paul Skrzypczyk, Daniel Cavalcanti, *Semidefinite Programming in Quantum Information Science*, IOP Publishing, 2023.

### 3.1 Quantum states and measurements

The purpose of this lecture is to introduce several basic classes of objects used in quantum information theory. This should also reveal their intrinsic connection to positive semidefinite matrices.

#### 3.1.1 Quantum states

In actuality, quantum states describe physical things—be they electrons, ions, photons, etc. Here, we abstract such implementation details until we are left with purely mathematical objects. We aim to capture the full power of quantum mechanics, while avoiding unneeded redundancy.

The *state* of a system is one of several positions or configurations that it can be in. As shown later, the objects we use to describe states relate to the probability the system is found to be in such states when observed. But, they have a quantum flavor in addition to regular probabilities.

#### Density operators

We now examine a set that has been featured in various examples in the first part of the course.

**Definition 3.1.1.** Let  $\mathcal{X}$  be a complex Euclidean space. A **density operator** is a positive semidefinite operator with trace equal to 1. The set of all such operators is denoted by

$$D(\mathcal{X}) := \{\rho \in \text{Pos}(\mathcal{X}) : \text{Tr}(\rho) = 1\} . \quad (3.1)$$

This set is particularly important since density operators represent quantum states. We usually denote density operators by letters such as  $\rho$  or  $\sigma$  (Greek letters “rho” and “sigma”).

**Example.** Let  $(p_1, \dots, p_n)$  be a probability vector. Each following  $\rho$  is a density operator.

$$\rho = \frac{1}{n} \mathbb{1}_n \quad \text{Maximally mixed state,} \quad (3.2)$$

$$\rho = \begin{pmatrix} p_1 & & \\ & \ddots & \\ & & p_n \end{pmatrix} \quad \text{Diagonal state,} \quad (3.3)$$

$$\rho = \begin{pmatrix} 1/2 & 1/4 \\ 1/4 & 1/2 \end{pmatrix} \quad \text{(no special name).} \quad (3.4)$$

When observed in the standard way, the diagonal state as seen above behaves identically to its underlying probability vector.

The set of density operators has several nice properties such as the following.

**Lemma 3.1.2.** *The set  $D(\mathcal{X})$  is convex and compact.*

*Proof.* We now show that  $D(\mathcal{X}) = \{X \in \text{Pos}(\mathcal{X}) : \text{Tr}(X) = 1\}$  is convex, bounded, and closed.

*Convex:* Let  $\rho, \sigma \in D(\mathcal{X})$ , and let  $\lambda \in [0, 1]$ . We have that  $\lambda\rho + (1 - \lambda)\sigma \in \text{Pos}(\mathcal{X})$  by convexity of  $\text{Pos}(\mathcal{X})$ . To show that  $\lambda\rho + (1 - \lambda)\sigma \in D(\mathcal{X})$ , we additionally require

$$\text{Tr}(\lambda\rho + (1 - \lambda)\sigma) = \lambda \text{Tr}(\rho) + (1 - \lambda) \text{Tr}(\sigma) = \lambda + (1 - \lambda) = 1. \quad (3.5)$$

*Bounded:* Let  $\rho \in D(\mathcal{X})$ . We have that  $D(\mathcal{X})$  is bounded with respect to  $\|\cdot\|_1$ , as

$$\text{Tr}(\rho) = \sum_{i=1}^n \lambda_i = \|\rho\|_1 = 1 \quad (\text{since } \lambda_i \geq 0). \quad (3.6)$$

*Closed:* Motivated by the above, we may write  $D(\mathcal{X}) = \text{Pos}(\mathcal{X}) \cap \{X : \|X\|_1 = 1\}$ . This means  $D(\mathcal{X})$  is an intersection of two closed sets, implying it is closed.

Since a set is compact when it is closed and bounded, the proof is complete.  $\square$

## Pure states

While density operators may by now be familiar from a mathematical perspective, we are yet to see how they relate to physical systems. To better understand them, we define a simplified notion of quantum states—one that often suffices in many applications such as quantum computing despite the lost generality.

**Definition 3.1.3.** A **pure state** is a rank-1 density operator. That is, a pure state  $\rho \in D(\mathcal{X})$  may be written as  $\rho = vv^*$  for some unit vector  $v \in \mathcal{X}$ , meaning  $\|v\|_2 = 1$ . This vector representation is not unique.

To understand the equivalency, note that a rank-1 linear operator  $A \in L(\mathcal{X}, \mathcal{Y})$  may be written as  $A = uv^*$  for some  $u \in \mathcal{Y}, v \in \mathcal{X}$ . Additionally restricting  $A \in \text{Pos}(\mathcal{X})$  implies  $A = vv^*$  for some  $v \in \mathcal{X}$ . Note that by the cyclic property of the trace,  $\text{Tr}(\rho) = \text{Tr}(vv^*) = \text{Tr}(v^*v) = \|v\|_2^2 = 1$ .

*Remark.* A density operator that is *not* pure is said to be **mixed**.

*Remark.* A common alternative, *Dirac notation*, represents column unit vectors as  $v = |\psi\rangle$  (called “ket psi”) and row unit vectors as  $v^* = \langle\psi|$  (called “bra psi”). Pure states can then be written as  $\rho = |\psi\rangle\langle\psi|$ .

*Example.* The following choices of  $v \in \mathbb{C}^2$  result in  $\rho = vv^* \in D(\mathbb{C}^2)$  being a pure state.

$$v = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{Zero state } |0\rangle, \quad (3.7)$$

$$v = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad \text{One state } |1\rangle, \quad (3.8)$$

$$v = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} \quad \text{Plus state } |+\rangle, \quad (3.9)$$

$$v = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix} \quad \text{Minus state } |-\rangle. \quad (3.10)$$

Due to having two dimensions, these are also called “qubits”—a quantum analogue of bits.

The above examples may be somewhat reminiscent of vector representations of probability vectors, except that now we have  $\|v\|_2 = 1$  instead of  $\|p\|_1 = 1$ , and the entries may be negative or even complex. These entries are thus not probabilities, and we give them a different name—*amplitudes*. They are still related to probabilities, though to fully see this we first need to define measurements. As a brief motivation, recall the diagonal state (3.3), which had probabilities on its diagonal. Note that for a pure state with the  $i$ th value in its vector being  $v_i$ , the corresponding entry on the diagonal of its density operator is  $|v_i|^2 \geq 0$ .

In addition to encoding probabilities via their magnitudes, amplitudes additionally have a *phase*. The uniquely quantum aspect of system evolution is that it is done via combinations of amplitudes, not probabilities. This leads to interference (constructive when phases align and destructive otherwise), a behavior crucial to many quantum algorithms.

## Ensembles and mixtures

We now have two ways of describing quantum states; a general one via density operators, and a perhaps more motivated one via pure state vectors. The next step is to connect these two notions.

**Definition 3.1.4.** Let  $\mathcal{X}$  be a complex Euclidean space, and let  $n \in \mathbb{N}$ . Suppose we are given a density operator from  $\rho_1, \dots, \rho_n \in D(\mathcal{X})$  according to some probability vector  $(p_1, \dots, p_n)$ . The **ensemble** representing such a sampling is the tuple  $(p_1\rho_1, \dots, p_n\rho_n) \in \text{Pos}(\mathcal{X})^{\times n}$ , and our knowledge about the state is the **mixture**

$$\rho = \sum_{i=1}^n p_i \rho_i \in D(\mathcal{X}) \quad (\text{by convexity of } D(\mathcal{X})). \quad (3.11)$$

This gives us the needed connection between pure and mixed states. We may pick distinct and pure  $\rho_1, \dots, \rho_n$  and observe that their mixture is now a mixed state. Density operators are thus simply probabilistic combinations of pure states, their generality allowing us to encode the amplitudes of pure states in a probabilistic manner.

A reasonable follow-up question to the above definitions is whether they uniquely correspond to each other. This is indeed true for ensembles; given  $(\sigma_1, \dots, \sigma_n) \in \text{Pos}(\mathcal{X})^{\times n}$ , we may recover  $p_i = \text{Tr}(\sigma_i)$  and  $\rho_i = \sigma_i / \text{Tr}(\sigma_i)$  for each  $i \in \{1, \dots, n\}$ . On the other hand, mixtures do not have this property—a density operator may be a mixture of multiple distinct ensembles.

*Example.* Note that each of the below matrices is positive semidefinite.

$$\frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 1/2 & -1/2 \\ -1/2 & 1/2 \end{pmatrix} = \frac{1}{2} \mathbb{1}_2 \in D(\mathbb{C}^2). \quad (3.12)$$

This shows that the density operator  $\frac{1}{2} \mathbb{1}_2$  is a mixture of two distinct ensembles.

While this means ensembles capture more information than density operators, the information obscured by the density operator representation happens to be entirely redundant. Any two ensembles having the same mixture are physically indistinguishable (we discuss distinguishability in Lecture 3.2), which happens to be a crucial property in quantum mechanics.

**Lemma 3.1.5.** *Every density operator is some mixture of pure states.*

*Proof.* Let  $\mathcal{X}$  be an  $n$ -dimensional complex Euclidean space and  $\rho \in D(\mathcal{X})$  a density operator. Since  $\rho$  is normal, it has a spectral decomposition of the form

$$\rho = \sum_{i=1}^n \lambda_i v_i v_i^*, \quad (3.13)$$

$\{v_1, \dots, v_n\}$  being an orthonormal basis for  $\mathcal{X}$ . As  $\rho \in D(\mathcal{X})$ , we have  $\rho \in \text{Pos}(\mathcal{X})$  and  $\text{Tr}(\rho) = 1$ , which in turn imply  $\lambda_i \geq 0$  for all  $i \in \{1, \dots, n\}$  and  $\sum_{i=1}^n \lambda_i = 1$ . Thus,  $(\lambda_1, \dots, \lambda_n)$  is a probability vector. Since  $\|v_i\|_2 = 1$  for all  $i \in \{1, \dots, n\}$ , each  $v_i v_i^*$  is a pure state. Thus,  $\rho$  is the mixture of an ensemble  $(\lambda_1 v_1 v_1^*, \dots, \lambda_n v_n v_n^*)$  of pure states.  $\square$

The role of eigenvalues as probabilities in the above ensemble necessitates the constraints we have used when defining density operators. Eigenvalues must be nonnegative and sum to one, and thus their corresponding density operators must be positive semidefinite and trace-one.

## Multiple quantum states

What if we want to combine multiple quantum states, perhaps ones shared by separate parties? We can do this via their *tensor product*, a notion introduced in Lectures 1.1 and 1.2. Given  $\rho \in D(\mathcal{X})$  and  $\sigma \in D(\mathcal{Y})$  for some complex Euclidean spaces  $\mathcal{X}$  and  $\mathcal{Y}$ , their **joint state** is  $\rho \otimes \sigma \in D(\mathcal{X} \otimes \mathcal{Y})$ .

**Lemma 3.1.6.** *If  $\rho \in D(\mathcal{X})$  and  $\sigma \in D(\mathcal{Y})$ , then  $\rho \otimes \sigma \in D(\mathcal{X} \otimes \mathcal{Y})$ .*

*Proof.* Suppose  $|\mathcal{X}| = m$  and  $|\mathcal{Y}| = n$ . Let  $(\lambda_1, u_1), \dots, (\lambda_m, u_m)$  be  $m$  linearly independent eigenpairs of  $\rho$ , and  $(\mu_1, v_1), \dots, (\mu_n, v_n)$  be  $n$  linearly independent eigenpairs of  $\sigma$ ; that is,  $\rho u_i = \lambda_i u_i$  for all  $i \in \{1, \dots, m\}$  and  $\sigma v_j = \mu_j v_j$  for all  $j \in \{1, \dots, n\}$ . We now show that  $\rho \in \text{Pos}(\mathcal{X})$  and  $\sigma \in \text{Pos}(\mathcal{Y})$  imply  $\rho \otimes \sigma \in \text{Pos}(\mathcal{X} \otimes \mathcal{Y})$  and  $\text{Tr}(\rho \otimes \sigma) = \text{Tr}(\rho) \text{Tr}(\sigma)$ .

*Positive semidefiniteness:* We have that

$$(\rho \otimes \sigma)(u_i \otimes v_j) = \rho u_i \otimes \sigma v_j = \lambda_i u_i \otimes \mu_j v_j = \lambda_i \mu_j (u_i \otimes v_j), \quad (3.14)$$

meaning  $(\lambda_i \mu_j, u_i \otimes v_j)$  form  $mn = |\mathcal{X} \otimes \mathcal{Y}|$  linearly independent eigenpairs of  $\rho \otimes \sigma$ . Each has a nonnegative eigenvalue, as  $\lambda_i, \mu_j \geq 0$  implies  $\lambda_i \mu_j \geq 0$ . Hence,  $\rho \otimes \sigma \in \text{Pos}(\mathcal{X} \otimes \mathcal{Y})$ .

*Trace:* We have that  $\text{Tr}(\rho) = \sum_{i=1}^m \lambda_i$  and  $\text{Tr}(\sigma) = \sum_{j=1}^n \mu_j$ , and hence

$$\text{Tr}(\rho \otimes \sigma) = \sum_{i=1}^m \sum_{j=1}^n \lambda_i \mu_j = \left( \sum_{i=1}^m \lambda_i \right) \left( \sum_{j=1}^n \mu_j \right) = \text{Tr}(\rho) \text{Tr}(\sigma). \quad (3.15)$$

We thus have  $\rho \otimes \sigma \in \text{Pos}(\mathcal{X} \otimes \mathcal{Y})$  and  $\text{Tr}(\rho \otimes \sigma) = 1$ , which completes the proof.  $\square$

*Remark.* Not every state in the space  $\mathcal{X} \otimes \mathcal{Y}$  can be represented as a product state  $\rho \otimes \sigma$ .

### 3.1.2 Measurements

We have on multiple occasions alluded to observing quantum states, but have not yet defined how to do so. One means of obtaining tangible classical information out of a quantum state is via measurements. Generally, we may think of a measurement as some device that consumes a quantum state, and in doing so informing us of one of  $n$  possible outcomes.

**Definition 3.1.7.** Let  $\mathcal{X}$  be a complex Euclidean space, and let  $n \in \mathbb{N}$ . An  $n$ -outcome **positive operator-valued measure (POVM)** is an  $n$ -tuple

$$(M_1, \dots, M_n) \in \text{Pos}(\mathcal{X})^{\times n} \quad \text{having} \quad \sum_{i=1}^n M_i = \mathbb{1}_{\mathcal{X}}. \quad (3.16)$$

Upon measuring  $\rho \in \text{D}(\mathcal{X})$  using this POVM, outcome  $i \in \{1, \dots, n\}$  is produced with probability  $p_i = \langle \rho, M_i \rangle$ . This behavior is known as *Born's rule*.

Note that  $(p_1, \dots, p_n)$  is a proper probability vector;  $\rho, M_i \in \text{Pos}(\mathcal{X})$  implies  $p_i \geq 0$ , and

$$\sum_{i=1}^n p_i = \sum_{i=1}^n \langle \rho, M_i \rangle = \left\langle \rho, \sum_{i=1}^n M_i \right\rangle = \langle \rho, \mathbb{1}_{\mathcal{X}} \rangle = \text{Tr}(\rho) = 1. \quad (3.17)$$

*Remark.* The summation constraint in (3.16) induces one of the POVM elements. For instance, an **effect**  $E \in \text{Eff}(\mathcal{X}) := \{E : 0 \preceq E \preceq \mathbb{1}\}$  corresponds to a unique 2-outcome POVM  $(E, \mathbb{1} - E)$ .

*Remark.* We assume the state to be destroyed upon measurement. While it is possible to define a more general notion of measurements that preserves some quantum state post-measurement, the POVM definition above suffices for our purposes.

**Lemma 3.1.8.** *The set of  $n$ -outcome POVMs acting on a space  $\mathcal{X}$ ,*

$$\left\{ (M_1, \dots, M_n) \in \text{Pos}(\mathcal{X})^{\times n} : \sum_{i=1}^n M_i = \mathbb{1}_{\mathcal{X}} \right\}, \quad (3.18)$$

*is convex.*

*Proof.* Let  $(M_1, \dots, M_n)$  and  $(N_1, \dots, N_n)$  be  $n$ -outcome POVMs, and let  $\lambda \in [0, 1]$ . Then,

$$\lambda(M_1, \dots, M_n) + (1 - \lambda)(N_1, \dots, N_n) = (\lambda M_1 + (1 - \lambda)N_1, \dots, \lambda M_n + (1 - \lambda)N_n). \quad (3.19)$$

We have that  $\lambda M_i + (1 - \lambda)N_i \in \text{Pos}(\mathcal{X})$  by convexity of  $\text{Pos}(\mathcal{X})$ . Additionally,

$$\sum_{i=1}^n [\lambda M_i + (1 - \lambda)N_i] = \lambda \sum_{i=1}^n M_i + (1 - \lambda) \sum_{i=1}^n N_i = \lambda \mathbb{1}_{\mathcal{X}} + (1 - \lambda) \mathbb{1}_{\mathcal{X}} = \mathbb{1}_{\mathcal{X}}. \quad (3.20)$$

This completes the proof.  $\square$

**Example.** Let  $n \in \mathbb{N}$ , and consider the POVM formed by the following matrices.

$$M_1 = E_{1,1} = \begin{pmatrix} 1 & & \\ & 0 & \\ & & \ddots \\ & & & 0 \end{pmatrix}, \quad M_2 = E_{2,2} = \begin{pmatrix} 0 & & \\ & 1 & \\ & & \ddots \\ & & & 0 \end{pmatrix}, \quad \dots, \quad M_n = E_{n,n} = \begin{pmatrix} 0 & & \\ & 0 & \\ & & \ddots \\ & & & 1 \end{pmatrix}. \quad (3.21)$$

Recall that  $E_{i,i} = e_i e_i^*$  for  $i \in \{1, \dots, n\}$ . As such, the above POVM may be seen as measuring in the standard basis  $\{e_1, e_2, \dots, e_n\}$ .

More generally, given an orthonormal basis  $\{v_1, \dots, v_n\}$ , the *basis measurement* with respect to it is defined by  $M_i = v_i v_i^*$  for  $i \in \{1, \dots, n\}$ . When each  $M_i = M_i^2$  is a projection, we get a *projective measurement*. Note that each basis measurement is necessarily a projective measurement.

**Example.** Here is another example of a POVM, this time on  $\mathbb{C}^2$ .

$$M_1 = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}, \quad M_2 = \begin{pmatrix} 1/2 & -1/2 \\ -1/2 & 1/2 \end{pmatrix}. \quad (3.22)$$

This is a basis measurement with respect to  $\{|+\rangle, |-\rangle\}$ , with said plus and minus states as defined in (3.9) and (3.10).

Suppose we are given either the plus state or the minus state. Measuring in the standard basis yields the same probabilities in both cases, yet using the POVM in (3.22) distinguishes them perfectly. This reveals another interesting property of quantum states—viewing them via different measurements may yield entirely different outcomes. Since quantum states are destroyed when measured (and cannot be cloned), extracting information from them may require knowing the right measurement ahead of time. This forms a basis for many quantum cryptographic protocols.

## 3.2 Distinguishing between two states

In this and the following lecture, we analyze more closely the task of distinguishing quantum states. We formalize the problem, formulate it as an SDP, and extract an explicit solution for two states.

Informally, the problem of state discrimination is as follows:

*Suppose Alice prepares a quantum state and sends it to Bob, can Bob identify the state Alice prepared?*

Specifically, Alice prepares a quantum state from a fixed set of finitely many quantum states, and Bob aims to guess the index of the prepared state while knowing the descriptions of what the states could be. If successful, Bob in effect extracts classical information from a physical quantum object.

**Definition 3.2.1** (Quantum state discrimination). Let  $\mathcal{X}$  be a complex Euclidean space and  $n \in \mathbb{N}$ . Suppose  $\rho_1, \dots, \rho_n \in \mathcal{D}(\mathcal{X})$  and a probability vector  $(p_1, \dots, p_n)$  are publicly known.

*Alice:* Randomly sample an index  $i \in \{1, \dots, n\}$  according to  $(p_1, \dots, p_n)$ . Create a physical copy of the state  $\rho_i$  and send it to Bob, without revealing  $i$ .

*Bob:* Apply an  $n$ -outcome POVM (see Definition 3.1.7) to the received state and obtain an outcome  $j \in \{1, \dots, n\}$ . Output  $j$  as the prediction for the state's index.

Bob is successful at distinguishing the states precisely when this guess is correct, i.e.,  $j = i$ .

In this lecture, we focus on a special case where  $n = 2$  and  $p_1 = p_2 = 1/2$ . Recall that a 2-outcome POVM is defined by  $(M_1, M_2) \in \text{Pos}(\mathcal{X})^{\times 2}$  with  $M_1 + M_2 = \mathbb{1}$ , and outputs index  $j$  with probability  $\langle \rho_i, M_j \rangle$ . For a given POVM, we may compute Bob's success probability as follows.

$$\Pr[\text{success}] = \Pr[\text{success}, \rho_1 \text{ sent}] + \Pr[\text{success}, \rho_2 \text{ sent}] \quad (3.23a)$$

$$= \Pr[\rho_1 \text{ sent}] \cdot \Pr[\text{success} \mid \rho_1 \text{ sent}] + \Pr[\rho_2 \text{ sent}] \cdot \Pr[\text{success} \mid \rho_2 \text{ sent}] \quad (3.23b)$$

$$= \frac{1}{2} \langle \rho_1, M_1 \rangle + \frac{1}{2} \langle \rho_2, M_2 \rangle. \quad (3.23c)$$

Bob aims to choose  $M_1$  and  $M_2$  that maximize the above success probability. Doing so while ensuring that  $(M_1, M_2)$  is a valid POVM can be represented by the following primal SDP:

| Primal problem                                                                                                 | Dual problem                                            |        |
|----------------------------------------------------------------------------------------------------------------|---------------------------------------------------------|--------|
| $\alpha = \text{maximize: } \frac{1}{2} \langle \rho_1, M_1 \rangle + \frac{1}{2} \langle \rho_2, M_2 \rangle$ | $\beta = \text{minimize: } \text{Tr}(Y)$                | (3.24) |
| $\text{subject to: } M_1 + M_2 = \mathbb{1}$                                                                   | $\text{subject to: } Y \succcurlyeq \frac{1}{2} \rho_1$ |        |
| $M_1, M_2 \succcurlyeq 0,$                                                                                     | $Y \succcurlyeq \frac{1}{2} \rho_2.$                    |        |

To find the above dual, we introduce the Lagrange multipliers  $Y$  for the affine constraint and  $\Lambda_1, \Lambda_2 \succcurlyeq 0$  for the PSD ones, and consider the Lagrangian function:

$$\mathcal{L}(M_1, M_2, Y, \Lambda_1, \Lambda_2) = \frac{1}{2} \langle \rho_1, M_1 \rangle + \frac{1}{2} \langle \rho_2, M_2 \rangle + \langle Y, \mathbb{1} - M_1 - M_2 \rangle + \langle \Lambda_1, M_1 \rangle + \langle \Lambda_2, M_2 \rangle \quad (3.25a)$$

$$= \langle M_1, \frac{1}{2} \rho_1 - Y + \Lambda_1 \rangle + \langle M_2, \frac{1}{2} \rho_2 - Y + \Lambda_2 \rangle + \langle \mathbb{1}, Y \rangle. \quad (3.25b)$$

As  $(M_1, M_2) = (\frac{1}{2} \mathbb{1}, \frac{1}{2} \mathbb{1})$  is strictly feasible for the primal problem and  $Y = \mathbb{1}$  is strictly feasible for the dual, Slater's theorem (see Theorem 2.6.1) implies  $\alpha = \beta$  and both SDPs attain optimal solutions. In particular, Bob's optimal success probability in the 2-state discrimination problem is realized by some POVM.

### 3.2.1 Two-state discrimination probability

We next leverage the SDP (3.24) to compute Bob's optimal probability of identifying one in two states. In particular, we are hoping to express  $\alpha$  as some simple function of  $\rho_1$  and  $\rho_2$  that does not require evaluating an entire SDP. First, we take the spectral decomposition of  $\rho_1 - \rho_2$ .

$$\rho_1 - \rho_2 = \sum_{i=1}^n \lambda_i v_i v_i^* . \quad (3.26)$$

Note that while  $\rho_1$  and  $\rho_2$  are PSD, their difference does not have to be—meaning the eigenvalues  $\lambda_i$  may be negative. Let us then define the measurement operators as follows:

$$M_1 = \sum_{i:\lambda_i>0} v_i v_i^* \succcurlyeq 0 , \quad M_2 = \sum_{i:\lambda_i\leq 0} v_i v_i^* \succcurlyeq 0 . \quad (3.27)$$

$(M_1, M_2)$  is a valid POVM by resolution of the identity;  $\mathbb{1} = \sum_{i=1}^n v_i v_i^* = M_1 + M_2$ , and is thus a feasible solution to the primal of (3.24). Its objective function value is

$$\frac{1}{2}\langle M_1, \rho_1 \rangle + \frac{1}{2}\langle M_2, \rho_2 \rangle = \frac{1}{2}\langle M_1, \rho_1 \rangle + \frac{1}{2}\langle \mathbb{1} - M_1, \rho_2 \rangle \quad (3.28a)$$

$$= \frac{1}{2}\langle \mathbb{1}, \rho_2 \rangle + \frac{1}{2}\langle M_1, \rho_1 - \rho_2 \rangle \quad (3.28b)$$

$$= \frac{1}{2} + \frac{1}{2} \sum_{i:\lambda_i>0} \lambda_i . \quad (3.28c)$$

Since the operator  $\rho_1 - \rho_2$  is traceless, we have  $\sum_i \lambda_i = \text{Tr}(\rho_1 - \rho_2) = 0$ , meaning we can write

$$\sum_{i:\lambda_i>0} \lambda_i = \sum_{i:\lambda_i<0} |\lambda_i| = \frac{1}{2} \sum_i |\lambda_i| = \frac{1}{2} \|\rho_1 - \rho_2\|_1 , \quad (3.29)$$

where the last equality follows by definition of the trace norm. From the constructed feasible solution, we now have that the value of the SDP (3.24) is lower-bounded by  $\alpha \geq \frac{1}{2} + \frac{1}{4} \|\rho_1 - \rho_2\|_1$ .

It remains to match this value via a feasible dual solution. We do this twice, firstly by defining

$$Y = \frac{1}{4}\rho_1 + \frac{1}{4}\rho_2 + \frac{1}{4} \sum_i |\lambda_i| v_i v_i^* , \quad \text{noting that} \quad \text{Tr}(Y) = \frac{1}{2} + \frac{1}{4} \|\rho_1 - \rho_2\|_1 . \quad (3.30)$$

We may verify that this solution is feasible by checking the PSD constraints of (3.24):

$$Y - \frac{1}{2}\rho_1 = \sum_i (|\lambda_i| - \lambda_i) v_i v_i^* \succcurlyeq 0 , \quad Y - \frac{1}{2}\rho_2 = \sum_i (|\lambda_i| + \lambda_i) v_i v_i^* \succcurlyeq 0 . \quad (3.31)$$

Alternatively, we may also verify the same via  $Y = \frac{1}{2}M_1\rho_1 + \frac{1}{2}M_2\rho_2$ . To check the constraints,

$$Y = \frac{1}{2}(\mathbb{1} - M_2)\rho_1 + \frac{1}{2}M_2\rho_2 = \frac{1}{2}\rho_1 - \frac{1}{2}M_2(\rho_1 - \rho_2) = \frac{1}{2}\rho_1 + \frac{1}{2} \sum_{i:\lambda_i<0} |\lambda_i^2| v_i v_i^* , \quad (3.32)$$

$$Y = \frac{1}{2}M_1\rho_1 + \frac{1}{2}(\mathbb{1} - M_1)\rho_2 = \frac{1}{2}\rho_2 + \frac{1}{2}M_1(\rho_1 - \rho_2) = \frac{1}{2}\rho_2 + \frac{1}{2} \sum_{i:\lambda_i\geq 0} |\lambda_i^2| v_i v_i^* , \quad (3.33)$$

which certify that  $Y \succcurlyeq \frac{1}{2}\rho_1$ ,  $Y \succcurlyeq \frac{1}{2}\rho_2$ , and additionally that  $Y \in \text{Herm}(\mathcal{X})$ . The demonstrated feasible solutions  $Y$  bound the value of the dual,  $\beta \leq \frac{1}{2} + \frac{1}{4} \|\rho_1 - \rho_2\|_1$ . As we had  $\alpha = \beta$  by strong duality, we conclude that Bob's optimal discrimination probability between two quantum states is

$$\alpha = \beta = \frac{1}{2} + \frac{1}{4} \|\rho_1 - \rho_2\|_1 . \quad (3.34)$$

We note that  $\|\rho_1 - \rho_2\|_1$  in the above is between 0 and 2; the former happens when  $\rho_1 = \rho_2$  and the latter when  $\langle \rho_1, \rho_2 \rangle = 0$ . Correspondingly, in these cases the discrimination probability evaluates to  $\alpha = 1/2$  (random guessing) and  $\alpha = 1$  (complete certainty).

### 3.2.2 Data processing inequality

In Definition 3.2.1, we model Bob's strategy as a POVM. This may seem like an unnecessary restriction; was there something else Bob could have done to increase the discrimination probability?

To this end, we consider quantum *channels*—general operations encompassing unitaries, measurements, and other actions such as creating or discarding part of a state. While we do not formally introduce channels until Lecture 3.5, for now it suffices to note that a channel is a linear map  $\Phi$ , that, among other things, ensures  $\Phi(X) \geq 0$  when  $X \geq 0$  and  $\text{Tr}(\Phi(X)) = \text{Tr}(X)$ . This means  $\Phi(\rho_1)$  and  $\Phi(\rho_2)$  are also quantum states, and we may attempt to discriminate these instead.

The central question is then: are  $\rho_1$  and  $\rho_2$  more or less distinguishable than  $\Phi(\rho_1)$  and  $\Phi(\rho_2)$ ? Decreasing distinguishability is trivial; for instance, via a channel  $\Phi(X) = \text{Tr}(X) \cdot \rho_3$  that discards its input and creates some fixed state  $\rho_3$ . While  $\rho_1$  and  $\rho_2$  may be distinguishable, two identical copies of  $\rho_3$  are certainly not. But, can some  $\Phi$  make  $\Phi(\rho_1)$  and  $\Phi(\rho_2)$  more distinguishable than the original states? Let us analyze the 2-state discrimination SDP with the processed states.

| Primal problem                                                                                                                                                                                  | Dual problem                                                                                                                              |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------|
| $\alpha' = \text{maximize: } \frac{1}{2}\langle \Phi(\rho_1), M_1 \rangle + \frac{1}{2}\langle \Phi(\rho_2), M_2 \rangle$<br>$\text{subject to: } M_1 + M_2 = \mathbb{1}$<br>$M_1, M_2 \geq 0,$ | $\beta' = \text{minimize: } \text{Tr}(Y')$<br>$\text{subject to: } Y' \geq \frac{1}{2}\Phi(\rho_1)$<br>$Y' \geq \frac{1}{2}\Phi(\rho_2).$ |

(3.35)

We again have strong duality with  $\alpha' = \beta'$ , and attempt to show whether or not  $\alpha' > \alpha$  is possible.

Recall that an optimal dual solution  $Y$  of the original SDP (3.24) is guaranteed to exist, having  $\text{Tr}(Y) = \beta = \alpha$  and  $Y \geq \frac{1}{2}\rho_i$  for  $i \in \{1, 2\}$ . By positivity and linearity of  $\Phi$ , the PSD constraints give

$$Y - \frac{1}{2}\rho_i \geq 0, \quad \Phi(Y - \frac{1}{2}\rho_i) \geq 0, \quad \Phi(Y) \geq \frac{1}{2}\Phi(\rho_i). \quad (3.36)$$

This implies  $Y' = \Phi(Y)$  is feasible in the dual of (3.35). Since  $\Phi$  is trace-preserving,

$$\alpha' = \beta' \leq \text{Tr}(Y') = \text{Tr}(\Phi(Y)) = \text{Tr}(Y) = \beta = \alpha, \quad (3.37)$$

letting us conclude that  $\alpha' \leq \alpha$ , meaning that  $\Phi$  cannot increase distinguishability.

Analogously solving the SDP (3.35) gives  $\alpha' = \frac{1}{2} + \frac{1}{4}\|\Phi(\rho_1) - \Phi(\rho_2)\|_1$ . Together with (3.34), our earlier insight of  $\alpha' \leq \alpha$  may be written as the following general information-theoretic statement.

**Theorem 3.2.2** (Data-processing inequality for  $\|\cdot\|_1$ ). *Suppose  $\Phi \in \mathcal{T}(\mathcal{X}, \mathcal{Y})$  is positive and trace-preserving. For any  $\rho_1, \rho_2 \in \mathcal{D}(\mathcal{X})$ , we have*

$$\|\rho_1 - \rho_2\|_1 \geq \|\Phi(\rho_1) - \Phi(\rho_2)\|_1. \quad (3.38)$$

### 3.3 Distinguishing between many states

We now return to quantum state discrimination (Definition 3.2.1) in full generality: with  $n$  possible states and arbitrary probabilities  $(p_1, \dots, p_n)$ . Analogously, Bob's success probability is

$$\Pr[\text{success}] = \sum_{i=1}^n \Pr[\text{success}, \rho_i \text{ sent}] \quad (3.39a)$$

$$= \sum_{i=1}^n \Pr[\rho_i \text{ sent}] \cdot \Pr[\text{success} \mid \rho_i \text{ sent}] = \sum_{i=1}^n p_i \langle \rho_i, M_i \rangle. \quad (3.39b)$$

We aim to find  $M_1, \dots, M_n \in \text{Pos}(\mathcal{X})$  with  $\sum_{i=1}^n M_i = \mathbb{1}$  to maximize the above. As an SDP, this is

| Primal problem                                                            | Dual problem                                    |
|---------------------------------------------------------------------------|-------------------------------------------------|
| $\alpha = \text{maximize: } \sum_{i=1}^n p_i \langle \rho_i, M_i \rangle$ | $\beta = \text{minimize: } \text{Tr}(Y)$        |
| $\text{subject to: } \sum_{i=1}^n M_i = \mathbb{1}$                       | $\text{subject to: } Y \succcurlyeq p_i \rho_i$ |
| $M_1, \dots, M_n \in \text{Pos}(\mathcal{X}),$                            | $\forall i \in \{1, \dots, n\}.$                |

(3.40)

Note that the dual has  $n$  PSD constraints due to the primal having  $n$  PSD variables. We leave derivation of the dual to the reader as Exercise 3.1.

While the general state discrimination SDP (3.40) is no longer simple enough for us to write Bob's success probability as a function of  $\rho_i$ , we may nevertheless use a feasible dual solution analogous to that in Lecture 3.2 to derive an equivalent condition for optimality.

**Theorem 3.3.1** (Holevo-Kennedy-Yuen-Lax). *A POVM  $(M_1, \dots, M_n)$  is optimal for  $n$ -state discrimination (see Definition 3.2.1) if and only if all of the following conditions hold:*

$$Y := \sum_{k=1}^n p_k M_k \rho_k \in \text{Herm}(\mathcal{X}), \quad Y \succcurlyeq p_i \rho_i \quad \forall i \in \{1, \dots, n\}. \quad (3.41)$$

*Proof.* Let  $(M_1, \dots, M_n)$  be a POVM, meaning  $\sum_{i=1}^n M_i = \mathbb{1}$  and  $M_i \succcurlyeq 0$  for all  $i \in \{1, \dots, n\}$ .

$\Leftarrow$ : Supposing (3.41) holds, we have that  $Y$  is dual feasible. Also,  $(M_1, \dots, M_n)$  is a POVM and is hence primal feasible. The objective function values of these two solutions are

$$\alpha = \sum_{i=1}^n p_i \langle \rho_i, M_i \rangle = \text{Tr} \left( \sum_{i=1}^n p_i M_i \rho_i \right) = \text{Tr}(Y) = \beta. \quad (3.42)$$

Weak duality then implies that both solutions are optimal.

$\Rightarrow$ : Suppose that  $(M_1, \dots, M_n)$  is an optimal primal solution for  $n$ -state discrimination. Since both problems are strictly feasible (see Exercise 3.2), by Slater's theorem there also exists a dual optimal solution  $Y$  with a matching objective function value. This  $Y$  must be feasible, meaning  $Y \in \text{Herm}(\mathcal{X})$  and  $Y \succcurlyeq p_i \rho_i$ . By complementary slackness (see Theorem 2.6.5),

$$M_i Y = M_i (p_i \rho_i) \quad \forall i \in \{1, \dots, n\} \quad \Rightarrow \quad \sum_{i=1}^n M_i Y = \sum_{i=1}^n p_i M_i \rho_i, \quad (3.43)$$

where we add together the constraints to obtain the right-hand side. Noting  $\sum_{i=1}^n M_i = \mathbb{1}$  yields (3.41), meaning the optimal solution  $Y$  is unique and must take said form.  $\square$

We may also derive an upper bound on the optimal quantum state discrimination probability.

**Theorem 3.3.2.** *Consider  $n$ -state discrimination with uniform probabilities (meaning  $p_i = 1/n$  for all  $i \in \{1, \dots, n\}$ ). Then,  $\alpha \leq \dim(\mathcal{X})/n$  in (3.40).*

For a reasonable chance of discriminating between many states, said states must hence be embedded in a state of large enough dimension.

*Proof.* Since  $\rho_i \geq 0$  and  $\text{Tr}(\rho_i) = 1$ , all eigenvalues of  $\rho_i$  are in  $[0, 1]$ , and hence  $\rho_i \leq \mathbb{1}$ . Then,

$$\alpha = \sum_{i=1}^n \frac{1}{n} \langle \rho_i, M_i \rangle \leq \sum_{i=1}^n \frac{1}{n} \langle \mathbb{1}, M_i \rangle = \frac{1}{n} \langle \mathbb{1}, \mathbb{1} \rangle = \frac{1}{n} \dim(\mathcal{X}), \quad (3.44)$$

where in the above we use  $\sum_{i=1}^n M_i = \mathbb{1}$  by validity of the POVM and  $\text{Tr}(\mathbb{1}) = \dim(\mathcal{X})$ .  $\square$

### 3.3.1 Error-free discrimination

When is it possible to have *perfect* quantum state discrimination? As this is equivalent to  $\alpha = 1$ , by Theorem 3.3.2 we know that  $\dim(\mathcal{X}) \geq n$  is necessary. Next, we derive a sufficient condition.

**Lemma 3.3.3.** *The state discrimination SDP (3.40) has  $\alpha = 1$  if and only if  $\langle \rho_i, \rho_j \rangle = 0$  for all  $i \neq j$ .*

*Proof.* Since  $\sum_{i=1}^n p_i = 1$ , the condition  $\alpha = 1$  is equivalent to  $\langle \rho_i, M_i \rangle = 1$  for all  $i \in \{1, \dots, n\}$ .

$\Rightarrow$ : Since  $\sum_{i=1}^n M_i = \mathbb{1}$ , perfect distinguishability is equivalent to  $\langle \rho_i, M_j \rangle = 0$  for all  $i \neq j$ , i.e., no mistakes are made. As  $\rho_i, M_j \geq 0$ , Lemma 2.6.4 lets us write this as  $\rho_i M_j = M_j \rho_i = 0$ .

$$\rho_i \rho_j = \rho_i \mathbb{1} \rho_j = \rho_i \left( \sum_{k=1}^n M_k \right) \rho_j = \left( \sum_{k=1}^n \rho_i M_k \right) \rho_j = (\rho_i M_i) \rho_j = \rho_i (M_i \rho_j) = \begin{cases} \rho_i^2 & i = j, \\ 0 & i \neq j. \end{cases} \quad (3.45)$$

Having  $\rho_i \rho_j = 0$  when  $i \neq j$  is again equivalent to  $\langle \rho_i, \rho_j \rangle = 0$  by Lemma 2.6.4.

$\Leftarrow$ : Supposing  $\langle \rho_i, \rho_j \rangle = 0$  for  $i \neq j$ , it is again equivalent to  $\rho_i \rho_j = \rho_j \rho_i = 0$ . This implies that the states  $\rho_i$  commute, meaning there exists a basis  $\{v_1, \dots, v_k\}$  that mutually diagonalizes all of them. Define a POVM  $(M_1, \dots, M_n)$  as follows:

$$M_i = \sum_{k: v_k^* \rho_i v_k > 0} v_k v_k^* \quad \text{for } i \in \{2, \dots, n\}, \quad M_1 = \mathbb{1} - \sum_{i=2}^n M_i. \quad (3.46)$$

We leave it as Exercise 3.3 to the reader to check that this yields  $\alpha = 1$ .  $\square$

### 3.3.2 Existence of nontrivial strategies

On the other hand, which situations are the *worst* for state discrimination? First note that  $\alpha \geq 1/n$ , as we may always attempt to randomly guess the index (that is,  $M_1 = \dots = M_n = \mathbb{1}/n$  is always feasible). Alternatively, we may simply output the index of the state with the highest  $p_i$  (that is,  $M_i = \mathbb{1}$  and  $M_j = 0$  for  $j \neq i$ ) to succeed with probability  $p_i \geq 1/n$ . Are there situations where nothing better than these trivial strategies exists, and if so, what are they?

**Lemma 3.3.4.** *The  $n$ -state discrimination SDP (3.40) with uniform probabilities (meaning  $p_i = 1/n$  for all  $i \in \{1, \dots, n\}$ ) has  $\alpha = 1/n$  if and only if  $\rho_1 = \dots = \rho_n$ .*

*Proof.* Throughout this proof the states are equiprobable, i.e.,  $p_i = 1/n$  for all  $i \in \{1, \dots, n\}$ .

$\Leftarrow$ : Suppose all states are the same, i.e.,  $\rho_1 = \dots = \rho_n$ . The SDP objective function is then

$$\sum_{i=1}^n \frac{1}{n} \langle \rho_i, M_i \rangle = \frac{1}{n} \sum_{i=1}^n \langle \rho_1, M_i \rangle = \frac{1}{n} \langle \rho_1, \mathbb{1} \rangle = \frac{1}{n}. \quad (3.47)$$

Substituting  $\rho_i$  by  $\rho_1$  above allows combining  $\sum_{i=1}^n M_i = \mathbb{1}$ . So,  $\alpha = 1/n$  in this case.

$\Rightarrow$ : It remains to show that  $\alpha = 1/n$  is not possible in any other case. For a contradiction, suppose  $\alpha = 1/n$  but  $\rho_i \neq \rho_j$  for some  $i$  and  $j$ . Set  $M_k = 0$  for all other  $k \notin \{i, j\}$ , effectively reducing the problem to 2-state discrimination. The success probability is then given by the same primal SDP as (3.24), but with  $\frac{1}{n}$  in place of  $\frac{1}{2}$  in the objective, for an overall  $\frac{2}{n}$  of its value. Leveraging (3.34), optimizing over  $M_i$  and  $M_j$  gives the success probability

$$\frac{2}{n} \left( \frac{1}{2} + \frac{1}{4} \|\rho_i - \rho_j\|_1 \right) = \frac{1}{n} + \frac{1}{2n} \|\rho_i - \rho_j\|_1 > \frac{1}{n}, \quad (3.48)$$

where the inequality follows since  $\rho_i \neq \rho_j$ . We thus have  $\alpha > 1/n$ , a contradiction.  $\square$

## 3.4 Duality of Uhlmann's and Alberti's theorems

In this lecture we introduce an important function in quantum information theory—the *fidelity*. We analyze it using semidefinite programming, in the process deriving and showing the duality of two theorems for expressing fidelity.

### 3.4.1 Fidelity

**Definition 3.4.1.** Let  $\mathcal{X}$  be a complex Euclidean space. The **fidelity** of  $P, Q \in \text{Pos}(\mathcal{X})$  is

$$F(P, Q) := \left\| \sqrt{P} \sqrt{Q} \right\|_1 = \text{Tr} \left( \sqrt{\sqrt{Q} P \sqrt{Q}} \right). \quad (3.49)$$

*Remark.* Some sources instead define fidelity as the above quantity, squared.

Recall that for a PSD operator  $P = \sum_{i=1}^n \lambda_i v_i v_i^*$  with  $\lambda_i \geq 0$  its square root is  $\sqrt{P} = \sum_{i=1}^n \sqrt{\lambda_i} v_i v_i^*$ . Recall also that  $\|A\|_1 = \text{Tr}(\sqrt{A^* A})$  is the trace norm, the sum of the singular values  $\sigma_i[A]$ .

Fidelity is often used as a measure of *similarity* between quantum states. Note that the fidelity of a state  $\rho \in \mathcal{D}(\mathcal{X})$  with itself is  $F(\rho, \rho) = \|\rho\|_1 = \text{Tr}(\rho) = 1$ . On the other hand, when  $\rho, \sigma \in \mathcal{D}(\mathcal{X})$  are orthogonal (meaning  $\langle \rho, \sigma \rangle = 0$ ), we have  $F(\rho, \sigma) = 0$ . In particular, for quantum states we have  $0 \leq F(\rho, \sigma) \leq 1$ , where the fidelity is higher when the states are closer to each other.

*Remark.* Despite being defined based on  $\|\cdot\|_1$ , which is a metric, the fidelity itself is *not* a metric. Nevertheless, it may be used to define a metric (see *Bures distance*).

We next express fidelity as the value of an SDP. Let  $P, Q \in \text{Pos}(\mathcal{X})$  and consider the following.

| Primal problem                                                                                                                                                                                      | Dual problem                                                                                                                                                                                                                             |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| $\alpha = \text{maximize: } \frac{1}{2} \text{Tr}(X) + \frac{1}{2} \text{Tr}(X^*)$<br>subject to: $\begin{bmatrix} P & X \\ X^* & Q \end{bmatrix} \in \text{Pos}(\mathcal{X} \oplus \mathcal{X})$ , | $\beta = \text{minimize: } \frac{1}{2} \langle P, Y_1 \rangle + \frac{1}{2} \langle Q, Y_2 \rangle$<br>subject to: $\begin{bmatrix} Y_1 & \mathbb{1} \\ \mathbb{1} & Y_2 \end{bmatrix} \in \text{Pos}(\mathcal{X} \oplus \mathcal{X})$ . |

(3.50)

*Remark.* Without loss of generality, we may restrict  $Y_1 \in \text{Pd}(\mathcal{X})$  and  $Y_2 = Y_1^{-1}$  in the above dual. While this may lose attainment of optimal solutions (which might not exist), it does not affect  $\beta$ .

Derivation of the dual SDP is left to the reader as Exercise 3.4.

**Lemma 3.4.2.** *The SDP pair in (3.50) has  $\alpha = \beta = F(P, Q)$ .*

*Proof.* Since  $X = 0$  is primal feasible and  $(Y_1, Y_2) = (2\mathbb{1}, 2\mathbb{1})$  is strictly dual feasible,  $\alpha = \beta$  and the primal attains an optimal solution. It remains to show that  $\alpha$  and  $\beta$  equal the fidelity.

$\alpha \geq F(P, Q)$ : For any unitary  $U \in \mathcal{U}(\mathcal{X})$ , we may define  $X = \sqrt{P} U^* \sqrt{Q}$ . We then have

$$\begin{bmatrix} P & X \\ X^* & Q \end{bmatrix} = \begin{bmatrix} \sqrt{P}\sqrt{P} & \sqrt{P}U^*\sqrt{Q} \\ \sqrt{Q}U\sqrt{P} & \sqrt{Q}\sqrt{Q} \end{bmatrix} = \begin{bmatrix} \sqrt{P} \\ \sqrt{Q}U \end{bmatrix} \begin{bmatrix} \sqrt{P} \\ \sqrt{Q}U \end{bmatrix}^* \in \text{Pos}(\mathcal{X} \oplus \mathcal{X}), \quad (3.51)$$

meaning  $X$  is primal feasible with an objective function value  $\text{Re}(\text{Tr}(\sqrt{Q}\sqrt{P}U^*))$ .

We thus aim to find a  $U \in \mathcal{U}(\mathcal{X})$  to make this value match the fidelity, which would then imply  $\alpha \geq F(P, Q)$ . Define  $U$  based on the singular value decomposition of  $\sqrt{Q}\sqrt{P}$ ;

$$\sqrt{Q}\sqrt{P} = \sum_{i=1}^r \sigma_i u_i v_i^*, \quad U := \sum_{i=1}^r u_i v_i^*. \quad (3.52)$$

Note that  $U$  is unitary by resolution of the identity. The objective function value is then

$$\text{Re}(\text{Tr}(\sqrt{Q}\sqrt{P}U^*)) = \text{Re}(\text{Tr}(\sum_{i=1}^r \sigma_i u_i v_i^* \sum_{j=1}^r v_j u_j^*)) = \sum_{i=1}^r \sigma_i = \|\sqrt{P}\sqrt{Q}\|_1. \quad (3.53)$$

$\beta \leq F(P, Q)$ : Analogously, we construct a feasible dual solution. For any  $W \in \text{Pd}(\mathcal{X})$ , let

$$\begin{bmatrix} W & \mathbb{1} \\ \mathbb{1} & W^{-1} \end{bmatrix} = \begin{bmatrix} \sqrt{W} \\ \sqrt{W^{-1}} \end{bmatrix} \begin{bmatrix} \sqrt{W} \\ \sqrt{W^{-1}} \end{bmatrix}^* \in \text{Pos}(\mathcal{X} \oplus \mathcal{X}), \quad (3.54)$$

so  $(Y_1, Y_2) = (W, W^{-1})$  is dual feasible with an objective value  $\frac{1}{2}\langle P, W \rangle + \frac{1}{2}\langle Q, W^{-1} \rangle$ .

If  $P$  was invertible, we could set  $W = P^{-1/2}(\sqrt{P}Q\sqrt{P})^{1/2}P^{-1/2}$  to attain  $F(P, Q)$ . As invertibility is not guaranteed for  $P \in \text{Pos}(\mathcal{X})$ , we show this using limits. For  $\epsilon > 0$ , define

$$\begin{aligned} P_\epsilon &:= P + \epsilon \mathbb{1} \in \text{Pos}(\mathcal{X}), & W_\epsilon &:= P_\epsilon^{-1/2}(\sqrt{P_\epsilon}Q_\epsilon\sqrt{P_\epsilon})^{1/2}P_\epsilon^{-1/2}. \end{aligned} \quad (3.55)$$

Then,  $(Y_1, Y_2) = (W_\epsilon, W_\epsilon^{-1})$  is dual feasible. We can check that its value is bounded by

$$\frac{1}{2}\langle P, W_\epsilon \rangle + \frac{1}{2}\langle Q, W_\epsilon^{-1} \rangle \leq \frac{1}{2}\langle P_\epsilon, W_\epsilon \rangle + \frac{1}{2}\langle Q_\epsilon, W_\epsilon^{-1} \rangle = F(P_\epsilon, Q_\epsilon) \rightarrow F(P, Q), \quad (3.56)$$

as  $\epsilon \downarrow 0$ . This family of feasible dual solutions then lets us conclude that  $\beta \leq F(P, Q)$ .

Combined, the statements yield  $\alpha = \beta = F(P, Q)$ , and we are done.  $\square$

In the above proof, we could as well have used  $X = \sqrt{Q}U^*\sqrt{P}$ , which is also primal optimal.

*Remark.* We have  $W_\epsilon := P_\epsilon^{-1} \# Q_\epsilon$  in (3.55), ‘#’ being the operator *geometric mean* (see Lecture 3.6).

Now that we have expressed  $F(P, Q)$  as the SDP (3.50), what can we do with it? One important consequence is that, since the primal constraint is linear in  $P$  and  $Q$ , we can now easily extend it to maximizing fidelity while varying  $P$  and/or  $Q$ . This could be interesting if, for instance,  $P$  and  $Q$  are quantum states with affine constraints.

### 3.4.2 Alberti’s and Uhlmann’s theorems

Recall the remark following the fidelity SDP (3.50): without affecting  $\beta$ , we may take  $Y_1 \in \text{Pd}(\mathcal{X})$  (let us also relabel it to  $R$ ) and  $Y_2 = Y_1^{-1}$  in the dual. This gives us the left-hand-side SDP below.

| Dual problem                                                                                                                                      | Modified dual                                                                                                                     |
|---------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------|
| $\beta = \text{minimize: } \frac{1}{2}\langle P, R \rangle + \frac{1}{2}\langle Q, R^{-1} \rangle$<br>subject to: $R \in \text{Pd}(\mathcal{X}),$ | $\beta' = \text{minimize: } \sqrt{\langle P, R \rangle \langle Q, R^{-1} \rangle}$<br>subject to: $R \in \text{Pd}(\mathcal{X}).$ |

(3.57)

We have  $\beta \geq \beta'$  by the arithmetic-geometric mean inequality (i.e.,  $(x + y)/2 \geq \sqrt{xy}$  for all  $x, y \geq 0$ ). Notice also that the modified dual is scale-invariant; that is,

$$\sqrt{\langle P, R \rangle} \sqrt{\langle Q, R^{-1} \rangle} = \sqrt{\langle P, \lambda R \rangle} \sqrt{\langle Q, (\lambda R)^{-1} \rangle} \quad \forall \lambda > 0. \quad (3.58)$$

So, we can scale  $R$  such that  $\langle P, \lambda R \rangle = \langle Q, (\lambda R)^{-1} \rangle$ , saturating the inequality and implying  $\beta = \beta'$ . Since  $\beta = F(P, Q)$  by Lemma 3.4.2, fidelity can now also be expressed via the modified dual.

**Theorem 3.4.3** (Alberti’s theorem). *For any  $P, Q \in \text{Pos}(\mathcal{X})$ , we have*

$$\begin{aligned} F(P, Q)^2 &= \text{minimize: } \langle P, R \rangle \langle Q, R^{-1} \rangle \\ &\text{subject to: } R \in \text{Pd}(\mathcal{X}). \end{aligned} \quad (3.59)$$

We have proven this to be true. We now use this to prove Uhlmann’s theorem, stated below.

**Theorem 3.4.4** (Uhlmann's Theorem). *Let  $\mathcal{Y}$  be a complex Euclidean space of sufficiently large dimension  $\dim(\mathcal{Y}) \geq \dim(\mathcal{X})$ . Let  $v \in \mathcal{X} \otimes \mathcal{Y}$  be a purification of  $Q$ , i.e.,  $\text{Tr}_{\mathcal{Y}}(vv^*) = Q$ . Then,*

$$\begin{aligned} F(P, Q)^2 &= \text{maximize: } \langle W, vv^* \rangle \\ &\text{subject to: } \text{Tr}_{\mathcal{Y}}(W) = P \\ &W \succcurlyeq 0. \end{aligned} \quad (3.60)$$

*Proof.* Denote the value of the SDP (3.60) by  $\alpha$ . This SDP is in standard form, and its dual is

Dual problem

$$\begin{aligned} \beta &= \text{minimize: } \langle P, Y \rangle \\ &\text{subject to: } Y \otimes \mathbb{1}_{\mathcal{Y}} \succcurlyeq vv^*, \end{aligned} \quad (3.61)$$

since  $\text{Tr}_{\mathcal{Y}}^*(Y) = Y \otimes \mathbb{1}_{\mathcal{Y}}$ . We aim to show that  $\alpha = \beta = F(P, Q)^2$ . As  $Y = 2 \text{Tr}(Q) \mathbb{1}$  is strictly dual feasible and  $W = P \otimes \rho$  for any  $\rho \in D(\mathcal{Y})$  is primal feasible,  $\alpha = \beta$  and the primal attains an optimal solution. It remains to show these equal  $F(P, Q)^2$ , which we do via the dual.

Note that the constraint in the dual problem (3.61) necessitates  $Y \in \text{Pos}(\mathcal{X})$ . We may further restrict  $Y \in \text{Pd}(\mathcal{X})$  without affecting  $\beta$ , though possibly losing attainment of an optimal solution (which might not exist in the first place). Supposing  $Y \succ 0$ ,

$$Y \otimes \mathbb{1}_{\mathcal{Y}} \succcurlyeq vv^* \quad \Leftrightarrow \quad \mathbb{1} \succcurlyeq (Y \otimes \mathbb{1}_{\mathcal{Y}})^{-1/2} vv^* (Y \otimes \mathbb{1}_{\mathcal{Y}})^{-1/2} \quad (3.62a)$$

$$\Leftrightarrow \quad 1 \geq \lambda_{\max}[(Y \otimes \mathbb{1}_{\mathcal{Y}})^{-1/2} vv^* (Y \otimes \mathbb{1}_{\mathcal{Y}})^{-1/2}]. \quad (3.62b)$$

The matrix in (3.62b) has rank 1, meaning  $\lambda_{\max}$  is equal to its trace. Thus,

$$Y \otimes \mathbb{1}_{\mathcal{Y}} \succcurlyeq vv^* \quad \Leftrightarrow \quad 1 \geq \text{Tr}((Y \otimes \mathbb{1}_{\mathcal{Y}})^{-1/2} vv^* (Y \otimes \mathbb{1}_{\mathcal{Y}})^{-1/2}) \quad (3.62c)$$

$$= \langle Y^{-1} \otimes \mathbb{1}_{\mathcal{Y}}, vv^* \rangle = \langle Y^{-1}, \text{Tr}_{\mathcal{Y}}(vv^*) \rangle = \langle Y^{-1}, Q \rangle. \quad (3.62d)$$

Replacing the PSD constraint of the dual (3.61) by this equivalent statement,

$$\begin{aligned} \beta &= \text{minimize: } \langle P, Y \rangle &= \text{minimize: } \langle P, Y \rangle \langle Y^{-1}, Q \rangle \\ &\text{subject to: } \langle Y^{-1}, Q \rangle \leq 1 &\text{subject to: } \langle Y^{-1}, Q \rangle = 1. \end{aligned} \quad (3.63)$$

The right-hand-side SDP follows by replacing the inequality in the constraint with an equality, then multiplying the objective function by 1. Specifically, if  $\langle Y^{-1}, Q \rangle < 1$ , we may rescale  $Y$  to  $\lambda Y$  with  $\lambda = \langle Y^{-1}, Q \rangle \in (0, 1)$  to make  $\langle (\lambda Y)^{-1}, Q \rangle = 1$  while only improving the objective function value. Note that the new objective function is scale-invariant, letting us remove the equality constraint entirely. Considering the implicit constraint  $Y \in \text{Pd}(\mathcal{X})$  (which is needed to take  $Y^{-1}$ ), the dual now exactly matches the SDP (3.59) in Alberti's theorem, letting us conclude that  $\alpha = \beta = F(P, Q)^2$ .  $\square$

The above proof shows that the dual of Uhlmann's SDP (3.60) is Alberti's SDP (3.59), meaning that Uhlmann's and Alberti's theorems are dual to each other.

### 3.5 Channel norms and fidelity properties

In Lecture 3.4, we have derived four SDP representations of the fidelity: (3.50), (3.57), (3.59) and (3.60). This lecture features a few useful results that follow as a consequence of these SDPs.

We have previously alluded to the concept of a *channel*—a linear map with physical significance in quantum information theory. The necessary restrictions for a quantum channel are as follows.

**Definition 3.5.1.** A **channel** is a linear map  $\Phi \in T(\mathcal{X}, \mathcal{Y})$  having the following properties.

$$(\Phi \otimes \mathbb{1}_{L(\mathcal{Z})})(X) \in \text{Pos}(\mathcal{Y} \otimes \mathcal{Z}) \quad \forall \mathcal{Z}, \forall X \in \text{Pos}(\mathcal{X} \otimes \mathcal{Z}) \quad (\text{completely positive}), \quad (3.64)$$

$$\text{Tr}(\Phi(X)) = \text{Tr}(X) \quad \forall X \in L(\mathcal{X}) \quad (\text{trace-preserving}). \quad (3.65)$$

The set of all channels from  $L(\mathcal{X})$  to  $L(\mathcal{Y})$  is denoted by  $C(\mathcal{X}, \mathcal{Y})$ , or  $C(\mathcal{X})$  when the spaces match.

The above conditions ensure that when acting on quantum states  $\rho \in D(\mathcal{X})$  and  $\sigma \in D(\mathcal{X} \otimes \mathcal{Z})$ , the outputs  $\Phi(\rho) \in D(\mathcal{Y})$  and  $(\Phi \otimes \mathbb{1}_{L(\mathcal{Z})})(\sigma) \in D(\mathcal{Y} \otimes \mathcal{Z})$  are likewise proper quantum states.

*Example.* The partial trace  $\text{Tr}_{\mathcal{X}} : L(\mathcal{X} \otimes \mathcal{Y}) \rightarrow L(\mathcal{Y})$  is a channel. For all  $A \in L(\mathcal{X} \otimes \mathcal{Y} \otimes \mathcal{Z})$ ,

$$x^*(\text{Tr}_{\mathcal{X}} \otimes \mathbb{1}_{L(\mathcal{Z})})(A)x = \sum_{i=1}^{\dim(\mathcal{Z})} (x \otimes e_i)^* A (x \otimes e_i) \geq 0 \quad \forall x \in \mathcal{Y} \otimes \mathcal{Z}, \quad (3.66)$$

and  $\text{Tr}(\text{Tr}_{\mathcal{X}}(B)) = \text{Tr}(B)$  for all  $B \in L(\mathcal{X} \otimes \mathcal{Y})$ . Operationally, the partial trace represents discarding a part of a quantum state.

Note that complete positivity (3.64) is a stronger condition than positivity (i.e.,  $\Phi(X) \in \text{Pos}(\mathcal{Y})$  for all  $X \in \text{Pos}(\mathcal{X})$ ). It is possible for a map to be positive, but not completely positive.

*Example.* The transpose  $\Phi(X) = X^T$  is not a channel. While  $T$  is positive (since  $vX^T v = \bar{v}X^* \bar{v} \geq 0$  for any  $X \in \text{Pos}(\mathcal{X})$ ), the *partial transpose*  $T \otimes \mathbb{1}_{L(\mathcal{Y})} =: T_{\mathcal{X}}$  is not (so  $T$  is not completely positive).

$$T_{\mathcal{X}} \left( \begin{pmatrix} \begin{matrix} 1 & 0 \\ 0 & 0 \end{matrix} & \begin{matrix} 0 & 1 \\ 0 & 0 \end{matrix} \\ \begin{matrix} 0 & 0 \\ 1 & 0 \end{matrix} & \begin{matrix} 0 & 0 \\ 0 & 1 \end{matrix} \end{pmatrix} \right) = \begin{pmatrix} \begin{matrix} 1 & 0 \\ 0 & 0 \end{matrix} & \begin{matrix} 0 & 0 \\ 1 & 0 \end{matrix} \\ \begin{matrix} 0 & 1 \\ 0 & 0 \end{matrix} & \begin{matrix} 0 & 0 \\ 0 & 1 \end{matrix} \end{pmatrix} \not\geq 0, \quad (3.67)$$

which may be checked with the eigenpair  $\lambda = -1, v = (0 \ 1 \ -1 \ 0)^T$ .

#### 3.5.1 The completely bounded trace norm

Analogously to how we used vector norms to induce operator norms in Lecture 1.2, we may now use operator norms to induce superoperator norms. Define the trace norm of  $\Phi \in T(\mathcal{X}, \mathcal{Y})$  as

$$\|\Phi\|_1 := \max \{ \|\Phi(X)\|_1 : X \in L(\mathcal{X}), \|X\|_1 \leq 1 \}. \quad (3.68)$$

That is, the (superoperator) trace norm is the largest (operator) trace norm of any output. While this could be a useful quantity in quantum information theory, it has an oversight; akin to the definition of channels, we need to consider how a linear map acts on one part of a larger system.

**Definition 3.5.2.** The **completely bounded trace norm** of a linear map  $\Phi \in T(\mathcal{X}, \mathcal{Y})$  is

$$\|\Phi\|_1 := \sup_{k \geq 1} \|\Phi \otimes \mathbb{1}_{L(\mathbb{C}^k)}\|_1 = \|\Phi \otimes \mathbb{1}_{L(\mathcal{X})}\|_1, \quad (3.69)$$

where  $\|\cdot\|_1$  is the superoperator trace norm. Note that this norm uses triple vertical bars.

*Remark.* The completely bounded trace norm is also called the “diamond norm”, denoted by  $\|\cdot\|_\diamond$ .

Completely bounded norms are significantly more useful in quantum information theory.

For quantum states  $\rho, \sigma \in D(\mathcal{X})$ , the trace norm can measure their similarity;  $\|\rho - \sigma\|_1$  appears in the 2-state discrimination probability (see Lecture 3.2), and  $\|\sqrt{\rho}\sqrt{\sigma}\|_1$  defines the fidelity (see Definition 3.4.1). Analogously, the completely bounded trace norm is closely related to the maximum fidelity between the outputs of two positive maps.

**Definition 3.5.3.** The **maximum output fidelity** of two positive maps  $\Phi, \Psi \in T(\mathcal{X}, \mathcal{Y})$  is

$$F_{\max}(\Phi, \Psi) := \max_{\rho, \sigma \in D(\mathcal{X})} F(\Phi(\rho), \Psi(\sigma)). \quad (3.70)$$

Since  $\Phi$  and  $\Psi$  are linear, we may express maximum output fidelity as an SDP by leveraging the primal of (3.50).

$$\begin{aligned} F_{\max}(\Phi, \Psi) = \text{maximize: } & \frac{1}{2} \text{Tr}(X) + \frac{1}{2} \text{Tr}(X^*) \\ \text{subject to: } & \begin{bmatrix} \Phi(\rho) & X \\ X^* & \Psi(\sigma) \end{bmatrix} \in \text{Pos}(\mathcal{Y} \oplus \mathcal{Y}) \\ & \rho, \sigma \in D(\mathcal{X}). \end{aligned} \quad (3.71)$$

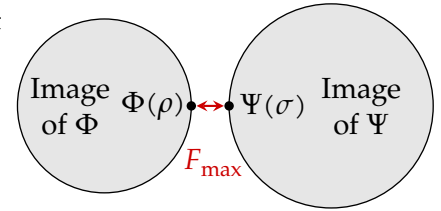


Figure 3.1:  $F_{\max}$  is the similarity between the images of  $\Phi$  and  $\Psi$ .

We now relate the completely bounded trace norm and the maximum output fidelity. Consider the Stinespring representation (see Definition 1.3.2) of any linear map

$$\Phi \in T(\mathcal{X}, \mathcal{Y}), \quad \Phi(X) = \text{Tr}_{\mathcal{Z}}(A_0 X A_1^*), \quad \text{where } A_0, A_1 \in L(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Z}). \quad (3.72)$$

Use the operators  $A_0$  and  $A_1$  above to define two new maps:

$$\Phi_0 \in T(\mathcal{X}, \mathcal{Z}), \quad \Phi_0(X) = \text{Tr}_{\mathcal{Y}}(A_0 X A_0^*), \quad (3.73a)$$

$$\Phi_1 \in T(\mathcal{X}, \mathcal{Z}), \quad \Phi_1(X) = \text{Tr}_{\mathcal{Y}}(A_1 X A_1^*). \quad (3.73b)$$

Since  $\Phi_0$  and  $\Phi_1$  have the same isometries on both sides, one can show they are completely positive. We then have  $\|\Phi\|_1 = F_{\max}(\Phi_0, \Phi_1)$  (though we do not prove this here). The completely bounded trace norm  $\|\Phi\|_1$  is then equivalently expressed as the SDP (3.71) with  $\Phi_0, \Phi_1$  in place of  $\Phi, \Psi$ .

### 3.5.2 Properties of the fidelity function

We next look at various properties of the fidelity, proving them via the previously derived SDPs.

**Lemma 3.5.4** (Properties of fidelity). *The following statements hold for any  $P, Q \in \text{Pos}(\mathcal{X})$ .*

$$F(P, Q) \geq 0 \quad \text{with equality if and only if } PQ = 0, \quad (3.74)$$

$$F(P, Q)^2 \leq \text{Tr}(P) \cdot \text{Tr}(Q) \quad \text{with equality if and only if } P \propto Q \text{ or } Q \propto P, \quad (3.75)$$

$$\sqrt{\lambda} F(P, Q) = F(P, \lambda Q) = F(\lambda P, Q), \quad (3.76)$$

$$F(P, Q) = F(Q, P), \quad (3.77)$$

$$F(P, Q) \leq F(\Phi(P), \Phi(Q)) \quad \text{for all channels } \Phi \in \mathcal{C}(\mathcal{X}, \mathcal{Y}), \quad (3.78)$$

$$F(P, Q) = F(UPU^*, UQU^*) \quad \text{for all unitaries } U \in \mathcal{U}(\mathcal{X}). \quad (3.79)$$

*Proof.* Recall that  $X = \sqrt{P} U^* \sqrt{Q}$  is feasible in the primal fidelity SDP (3.50) for all  $U \in \mathcal{U}(\mathcal{X})$ .

(3.74): Since  $X = 0$  is feasible in the primal fidelity SDP (3.50), we have  $F(P, Q) \geq 0$ . It remains to prove that  $F(P, Q) = 0 \Leftrightarrow PQ = 0$ , where we show the two directions separately.

$\Rightarrow$ : Suppose  $F(P, Q) = 0$ . Since  $X = \sqrt{P}\sqrt{Q}$  is feasible in the primal fidelity SDP (3.50),

$$F(P, Q) \geq \text{Tr}(\sqrt{P}\sqrt{Q}) = \|P^{1/4}Q^{1/4}\|_2 = 0 \quad \Rightarrow \quad P^{1/4}Q^{1/4} = 0, \quad (3.80)$$

letting us conclude that  $PQ = P^{3/4}(P^{1/4}Q^{1/4})Q^{3/4} = 0$ .

$\Leftarrow$ : Suppose  $PQ = 0$ . For  $\epsilon > 0$ , define  $Q_\epsilon = Q + \epsilon \mathbb{1}$ . Then, by Alberti's Theorem (3.59), setting  $R = Q_\epsilon$  gives  $F(P, Q_\epsilon)^2 \leq \epsilon \text{Tr}(P) \dim(\mathcal{X})$ . Taking limits,  $F(P, Q) = 0$ .

(3.75): The inequality follows from noting that  $R = \mathbb{1}$  is feasible in Alberti's theorem (3.59). It remains to show the equality condition  $F(P, Q)^2 = \text{Tr}(P) \text{Tr}(Q) \Leftrightarrow P \propto Q$  or  $Q \propto P$ . Both sides hold trivially when  $P = 0$  or  $Q = 0$ , so we may assume otherwise.

$\Leftarrow$ : Since  $X = \sqrt{P}\sqrt{Q}$  is feasible in the primal fidelity SDP (3.50),  $F(P, Q) \geq \text{Tr}(\sqrt{P}\sqrt{Q})$ .

Suppose  $P \propto Q$ , meaning  $P = \lambda Q$  for  $\lambda > 0$ . Then  $F(P, Q) \geq \sqrt{\lambda} \text{Tr}(Q)$ , implying  $F(P, Q)^2 \geq \lambda \text{Tr}(Q)^2 = \text{Tr}(P) \text{Tr}(Q)$ . From the already proven inequality part of (3.75), this must be an equality. The  $Q \propto P$  case follows analogously.

$\Rightarrow$ : Suppose  $F(P, Q)^2 = \text{Tr}(P) \text{Tr}(Q)$ , and aim to show  $P$  and  $Q$  are proportional. Let  $\lambda > 0$  be such that  $\langle P, \lambda \mathbb{1} \rangle = \langle Q, \lambda^{-1} \mathbb{1} \rangle$ , and note that  $R = \lambda \mathbb{1}$  is feasible in the dual fidelity SDP (3.57). We then have  $F(P, Q) \leq \frac{1}{2} \langle P, R \rangle + \frac{1}{2} \langle Q, R^{-1} \rangle = \sqrt{\langle P, R \rangle \langle Q, R^{-1} \rangle} = \sqrt{\text{Tr}(P) \text{Tr}(Q)}$ . By the original condition the above is an equality, and  $R$  is optimal. Since there exists an optimal primal solution  $X$ , complementary slackness yields

$$0 = \begin{bmatrix} P & X \\ X^* & Q \end{bmatrix} \begin{bmatrix} \lambda \mathbb{1} & \mathbb{1} \\ \mathbb{1} & \lambda^{-1} \mathbb{1} \end{bmatrix} = \begin{bmatrix} \lambda P + X & P + \lambda^{-1} X \\ \lambda X^* + Q & X^* + \lambda^{-1} Q \end{bmatrix}, \quad (3.81)$$

which implies  $Q = \lambda^2 P$ .

(3.76): This follows immediately from Alberti's theorem (3.59).

(3.77): This follows immediately from Alberti's theorem (3.59) by switching  $R$  and  $R^{-1}$ .

(3.78): Use the primal fidelity SDP (3.50) to compute both  $F(P, Q)$  and  $F(\Phi(P), \Phi(Q))$ . Let  $X$  be an optimal solution of the  $F(P, Q)$  SDP. Since  $\mathbb{1}_{L(\mathbb{C}^2)} \otimes \Phi$  is positive,

$$\begin{bmatrix} P & X \\ X^* & Q \end{bmatrix} \in \text{Pos}(\mathcal{X} \oplus \mathcal{X}) \quad \Rightarrow \quad \begin{bmatrix} \Phi(P) & \Phi(X) \\ \Phi(X^*) & \Phi(Q) \end{bmatrix} \in \text{Pos}(\mathcal{Y} \oplus \mathcal{Y}). \quad (3.82)$$

Thus,  $\Phi(X)$  is feasible in the  $F(\Phi(P), \Phi(Q))$  SDP (but maybe not optimal), implying that

$$F(\Phi(P), \Phi(Q)) \geq \text{Re}(\text{Tr}(\Phi(X))) = \text{Re}(\text{Tr}(X)) = F(P, Q). \quad (3.83)$$

(3.79): Use the dual fidelity SDP (3.57) to compute  $F(P, Q)$  and  $F(UPU^*, UQU^*)$ . Note that

$$\frac{1}{2}\langle UPU^*, W \rangle + \frac{1}{2}\langle UQU^*, W^{-1} \rangle = \frac{1}{2}\langle P, U^*WU \rangle + \frac{1}{2}\langle Q, U^*W^{-1}U \rangle. \quad (3.84)$$

If  $Y$  is feasible in the former SDP, then  $W = UYU^*$  is feasible in the latter with the same objective function value. If  $W$  is feasible in the latter SDP, then  $Y = U^*WU$  is feasible in the former with the same objective function value.  $\square$

## 3.6 Operator geometric means

In this lecture, we define geometric means for operators, and prove several related properties.

We begin with a technical result about PSD operators: the method of checking for a non-negative determinant in a  $2 \times 2$  matrix also works for block matrices.

**Lemma 3.6.1** (Schur complement). *Let  $A, B, D \in L(\mathcal{X})$  and  $C \in \text{Pd}(\mathcal{X})$ . We have*

$$X := \begin{bmatrix} A & B \\ B^* & C \end{bmatrix} \succcurlyeq 0 \quad \Leftrightarrow \quad A - BC^{-1}B^* \succcurlyeq 0. \quad (3.85)$$

*Proof.* Define the block operators

$$M = \begin{bmatrix} \mathbb{1} & BC^{-1} \\ 0 & \mathbb{1} \end{bmatrix}, \quad X' = \begin{bmatrix} A - BC^{-1}B^* & 0 \\ 0 & C \end{bmatrix}. \quad (3.86)$$

We then have  $X = MX'M^*$  and  $X' = M^{-1}X(M^{-1})^*$ , and the following chain of implications:

$$X \succcurlyeq 0 \quad \Leftrightarrow \quad X' \succcurlyeq 0 \quad \Leftrightarrow \quad A - BC^{-1}B^* \succcurlyeq 0. \quad (3.87)$$

The right-hand-side condition does not require  $C \succcurlyeq 0$  as it follows from  $C \in \text{Pd}(\mathcal{X})$ .  $\square$

### 3.6.1 Definition of the geometric mean

The geometric mean of two real numbers  $x, y \geq 0$  is  $\sqrt{xy}$ . We consider an operator analogue:

**Definition 3.6.2.** The operator **geometric mean** of  $P, Q \in \text{Pd}(\mathcal{X})$  is defined as

$$P \# Q := P^{1/2} (P^{-1/2} Q P^{-1/2})^{1/2} P^{1/2}. \quad (3.88)$$

Positive definiteness lets us take inverses. We soon give an extension of this for  $P, Q \in \text{Pos}(\mathcal{X})$ .

**Lemma 3.6.3.** Suppose  $\{A_t \in \text{Pos}(\mathcal{X}) : t > 0\}$  is an infinite family of positive semidefinite matrices satisfying  $A_{t_1} \succcurlyeq A_{t_2}$  for  $t_1 \geq t_2$ . Then, there exists  $A_0 \in \text{Pos}(\mathcal{X})$  such that  $A_0 = \lim_{t \downarrow 0} A_t$ .

*Proof.* Let  $\{B_1, \dots, B_n\}$  be a basis for  $\text{Herm}(\mathcal{X})$ , where  $B_i \in \text{Pos}(\mathcal{X})$ . Define  $b_{k,t} := \langle B_k, A_t \rangle$  and  $b_k := \lim_{t \downarrow 0} b_{k,t}$  (which exists since  $b_{k,t}$  is a non-increasing set of nonnegative numbers). Since  $\{B_1, \dots, B_n\}$  is a basis, there exists a unique matrix  $A$  such that  $\langle B_k, A \rangle = b_k$ . Since  $\lim_{t \downarrow 0} \langle B_k, A_t \rangle = \langle B_k, A \rangle$ , we know that  $\lim_{t \downarrow 0} A_t = A$ . Finally, we know that  $A \in \text{Pos}(\mathcal{X})$  since  $\text{Pos}(\mathcal{X})$  is closed.  $\square$

**Lemma 3.6.4.** For  $P, Q \in \text{Pd}(\mathcal{X})$  and  $W \in \text{Herm}(\mathcal{X})$ , we have

$$\begin{bmatrix} P & W \\ W & Q \end{bmatrix} \succcurlyeq 0 \quad \Rightarrow \quad W \preccurlyeq P \# Q, \quad \begin{bmatrix} P & P \# Q \\ P \# Q & Q \end{bmatrix} \succcurlyeq 0. \quad (3.89)$$

*Proof.* By taking Schur complements, we have  $Q \succcurlyeq W P^{-1} W^*$ . Appending  $P^{-1/2}$  to both sides,

$$P^{-1/2} Q P^{-1/2} \succcurlyeq P^{-1/2} W P^{-1/2} P^{-1/2} W^* P^{-1/2} = (P^{-1/2} W P^{-1/2})^2. \quad (3.90)$$

Taking square roots preserves positive semidefiniteness, so

$$(P^{-1/2} Q P^{-1/2})^{1/2} \succcurlyeq |P^{-1/2} W P^{-1/2}| \succcurlyeq P^{-1/2} W P^{-1/2}, \quad (3.91)$$

where for  $A = \sum_{i=1}^n \lambda_i v_i v_i^*$ , its absolute value is  $|A| = \sum_{i=1}^n |\lambda_i| v_i v_i^*$ . The above is equivalent to  $P^{1/2} (P^{-1/2} Q P^{-1/2})^{1/2} P^{1/2} = P \# Q \succcurlyeq W$ . To finish the proof, notice that

$$\begin{bmatrix} P & P \# Q \\ P \# Q & Q \end{bmatrix} = R R^*, \quad \text{where} \quad R = \begin{bmatrix} P^{1/2} \\ P^{1/2} (P^{-1/2} Q P^{-1/2})^{1/2} \end{bmatrix}, \quad (3.92)$$

and so the inequality on the right-hand-side of (3.89) holds unconditionally.  $\square$

*Remark.* The converse of (3.89) is false; squaring does *not* preserve positive semidefiniteness.

**Corollary 3.6.5.** For  $P, Q \in \text{Pos}(\mathcal{X})$ , the limit  $\lim_{\epsilon \downarrow 0} (P + \epsilon \mathbb{1}) \# (Q + \epsilon \mathbb{1})$  exists. Moreover, if  $P, Q \in \text{Pd}(\mathcal{X})$ , we have  $\lim_{\epsilon \downarrow 0} (P + \epsilon \mathbb{1}) \# (Q + \epsilon \mathbb{1}) = P \# Q$ .

*Proof.* Let  $\epsilon \geq 0$  and define  $W_\epsilon := (P + \epsilon \mathbb{1}) \# (Q + \epsilon \mathbb{1}) \succcurlyeq 0$ . From Lemma 3.6.4, we know that

$$\begin{bmatrix} P + \epsilon \mathbb{1} & W_\epsilon \\ W_\epsilon & P + \epsilon \mathbb{1} \end{bmatrix} \succcurlyeq 0, \quad \begin{bmatrix} P + \delta \mathbb{1} & W_\epsilon \\ W_\epsilon & Q + \delta \mathbb{1} \end{bmatrix} \succcurlyeq 0 \quad \text{for } \delta > \epsilon. \quad (3.93)$$

Thus, we have  $W_\epsilon \preccurlyeq W_\delta$ . By Lemma 3.6.3,  $\lim_{\epsilon \downarrow 0} W_\epsilon$  exists. Lastly, since  $f(X) = 1/X$  and  $f(X) = X^{1/2}$  are continuous,  $\lim_{\epsilon \downarrow 0} W_\epsilon = P \# Q$  when  $P, Q \in \text{Pd}(\mathcal{X})$ .  $\square$

We have now shown a limit that converges onto  $P \# Q$  when  $P, Q \in \text{Pd}(\mathcal{X})$ , and is well-defined for all  $P, Q \in \text{Pos}(\mathcal{X})$ . We may hence use it to extend the geometric mean to PSD operators.

**Definition 3.6.6.** For  $P, Q \in \text{Pos}(\mathcal{X})$ , define  $P \# Q := \lim_{\epsilon \downarrow 0} (P + \epsilon \mathbb{1}) \# (Q + \epsilon \mathbb{1})$ .

We may likewise extend Lemma 3.6.4 to positive semidefinite operators.

**Lemma 3.6.7.** For  $P, Q \in \text{Pos}(\mathcal{X})$  and  $W \in \text{L}(\mathcal{X})$ , we have

$$\begin{bmatrix} P & W \\ W & Q \end{bmatrix} \succcurlyeq 0 \quad \Rightarrow \quad W \preccurlyeq P \# Q, \quad \begin{bmatrix} P & P \# Q \\ P \# Q & Q \end{bmatrix} \succcurlyeq 0. \quad (3.94)$$

*Proof.* For any  $\epsilon > 0$  and  $W \in \text{L}(\mathcal{X})$ , we have

$$\begin{bmatrix} P & W \\ W & Q \end{bmatrix} \succcurlyeq 0 \quad \Rightarrow \quad \begin{bmatrix} P + \epsilon \mathbb{1} & W \\ W & Q + \epsilon \mathbb{1} \end{bmatrix} \succcurlyeq 0. \quad (3.95)$$

Thus,  $W \preccurlyeq W_\epsilon := (P + \epsilon \mathbb{1}) \# (Q + \epsilon \mathbb{1})$  for all  $\epsilon > 0$ , and  $W \preccurlyeq \lim_{\epsilon \downarrow 0} W_\epsilon = P \# Q$ . Finally,

$$\lim_{\epsilon \downarrow 0} \begin{bmatrix} P + \epsilon \mathbb{1} & W_\epsilon \\ W_\epsilon & Q + \epsilon \mathbb{1} \end{bmatrix} = \begin{bmatrix} P & P \# Q \\ P \# Q & Q \end{bmatrix} \succcurlyeq 0, \quad (3.96)$$

using the extended Definition 3.6.6 and the fact that  $\text{Pos}(\mathcal{X})$  is closed.  $\square$

### 3.6.2 Properties of the geometric mean

Let us first define a couple of concepts relevant to the operator geometric mean.

**Definition 3.6.8.** Let  $P, Q \in \text{Pos}(\mathcal{X})$ . The **hypograph** of  $P \# Q$  is

$$H(P, Q) := \left\{ Z \in \text{Herm}(\mathcal{X}) : \exists W \in \text{Herm}(\mathcal{X}), \begin{bmatrix} P & W \\ W & Q \end{bmatrix} \succcurlyeq 0, W \succcurlyeq Z \right\}. \quad (3.97)$$

**Definition 3.6.9.** The **lower closure** of  $W \in \text{Herm}(\mathcal{X})$  is

$$\downarrow W := \{ Z \in \text{Herm}(\mathcal{X}) : W \succcurlyeq Z \}. \quad (3.98)$$

**Lemma 3.6.10.** For any  $P, Q \in \text{Pos}(\mathcal{X})$ , we have  $H(P, Q) = \downarrow(P \# Q)$ .

*Proof.* We show the two directions of membership implication separately.

$\Rightarrow$ : Suppose  $Z \in H(P, Q)$ . Then, there exists a  $W$  such that  $\begin{bmatrix} P & W \\ W & Q \end{bmatrix} \succcurlyeq 0$  and  $W \succcurlyeq Z$ , implying  $W \preccurlyeq P \# Q$ . Since  $Z \preccurlyeq W$ , we have  $Z \preccurlyeq W \preccurlyeq P \# Q$ , meaning  $Z \in \downarrow(P \# Q)$ .

$\Leftarrow$ : Suppose  $Z \in \downarrow(P \# Q)$ , i.e.,  $Z \preccurlyeq P \# Q =: W$ . Then,  $\begin{bmatrix} P & W \\ W & Q \end{bmatrix} \succcurlyeq 0$ , and  $Z \in H(P, Q)$ .  $\square$

The definition of the hypergraph reveals that changing the order of its inputs has no effect.

**Corollary 3.6.11.** For any  $P, Q \in \text{Pos}(\mathcal{X})$ , we have  $H(P, Q) = H(Q, P)$ .

We may then use lower closures to show that the same is true for the geometric mean.

**Lemma 3.6.12.** For any  $P, Q \in \text{Pos}(\mathcal{X})$ , we have  $P \# Q = Q \# P$ .

*Proof.* From Lemma 3.6.10, we have that

$$P \# Q \in \downarrow(P \# Q) = H(P, Q) = H(Q, P) = \downarrow(Q \# P) \ni Q \# P. \quad (3.99)$$

implying  $P \# Q \preccurlyeq Q \# P$  and  $P \# Q \succcurlyeq Q \# P$ , and hence  $P \# Q = Q \# P$ .  $\square$

**Lemma 3.6.13.** For any  $P, Q, R \in \text{Pos}(\mathcal{X})$  satisfying  $R \succcurlyeq Q$ , we have  $P \# R \succcurlyeq P \# Q$ .

*Proof.* We have

$$\begin{bmatrix} P & P \# Q \\ P \# Q & R \end{bmatrix} \succcurlyeq \begin{bmatrix} P & P \# Q \\ P \# Q & Q \end{bmatrix} \succcurlyeq 0. \quad (3.100)$$

Thus,  $P \# Q \preccurlyeq P \# R$ .  $\square$

**Lemma 3.6.14.** For any  $P, Q \in \text{Pos}(\mathcal{X})$ , if  $P$  commutes with  $Q$ , then  $P \# Q = \sqrt{P}\sqrt{Q}$ .

*Proof.* If  $P$  and  $Q$  commute, then  $PQ = QP$ . Also,

$$(P + \epsilon \mathbb{1})(Q + \epsilon \mathbb{1}) = PQ + \epsilon Q + \epsilon P + \epsilon^2 \mathbb{1}. \quad (3.101)$$

So,  $P + \epsilon \mathbb{1}$  commutes with  $Q + \epsilon \mathbb{1}$ . The rest follows by commuting the terms in  $P \# Q$ .  $\square$

Geometric means reveal the following SDP characterization, which we use in Lecture 3.7.

*Remark.* “vec” denotes *vectorization*, which stacks the rows of a matrix. For  $\mathbb{1} = \sum_{i=1}^n e_i e_i^*$  we have

$$\text{vec}(\mathbb{1}) = \sum_{i=1}^n e_i \otimes e_i. \quad (3.102)$$

**Lemma 3.6.15.** *Let  $P, Q \in \text{Pos}(\mathcal{X})$  and  $m \in \mathbb{N}$ . We have*

$$\begin{aligned} \langle P^{1-2^{-m}}, Q^{2^{-m}} \rangle = \text{maximize: } & \text{vec}(\mathbb{1}_{\mathcal{X} \otimes \mathcal{X}})^* W_m \text{vec}(\mathbb{1}_{\mathcal{X} \otimes \mathcal{X}}) \\ \text{subject to: } & \begin{bmatrix} P \otimes \mathbb{1}_{\mathcal{X}} & W_k \\ W_k & W_{k-1} \end{bmatrix} \succcurlyeq 0 \text{ for all } k \in \{1, \dots, m\} \\ & W_0 = \mathbb{1}_{\mathcal{X}} \otimes Q^T \\ & W_0, \dots, W_m \in \text{Pos}(\mathcal{X} \otimes \mathcal{X}). \end{aligned} \quad (3.103)$$

*Proof.* We may rewrite the recursive constraints in the SDP explicitly in terms of  $P$  and  $Q$ .

$$W_1 \preccurlyeq (P \otimes \mathbb{1}_{\mathcal{X}}) \# W_0 \quad (\text{from the block matrix constraint}) \quad (3.104a)$$

$$= (P \otimes \mathbb{1}_{\mathcal{X}}) \# (\mathbb{1}_{\mathcal{X}} \otimes Q^T) \quad (\text{from } W_0 = \mathbb{1}_{\mathcal{X}} \otimes Q^T) \quad (3.104b)$$

$$= (P \otimes \mathbb{1}_{\mathcal{X}})^{1/2} (\mathbb{1}_{\mathcal{X}} \otimes Q^T)^{1/2} \quad (\text{from Lemma 3.6.14}) \quad (3.104c)$$

$$= P^{1/2} \otimes (Q^T)^{1/2}. \quad (3.104d)$$

Similarly, we can show  $W_2 \preccurlyeq P^{3/4} \otimes (Q^T)^{1/4}$ , and more generally  $W_k \preccurlyeq P^{1-2^{-k}} \otimes (Q^T)^{2^{-k}}$  for  $k \in \{1, \dots, m\}$ . We can then bound the objective function value as

$$\text{vec}(\mathbb{1}_{\mathcal{X} \otimes \mathcal{X}})^* W_m \text{vec}(\mathbb{1}_{\mathcal{X} \otimes \mathcal{X}}) \leq \text{vec}(\mathbb{1}_{\mathcal{X} \otimes \mathcal{X}})^* (P^{1-2^{-m}} \otimes (Q^T)^{2^{-m}}) \text{vec}(\mathbb{1}_{\mathcal{X} \otimes \mathcal{X}}) \quad (3.105a)$$

$$= \langle P^{1-2^{-m}}, Q^{2^{-m}} \rangle. \quad (3.105b)$$

The last equality follows from the properties of vec. Consider the feasible solution

$$W_0 = \mathbb{1}_{\mathcal{X}} \otimes Q^T, \quad W_k = (P \otimes \mathbb{1}_{\mathcal{X}}) \# W_{k-1} \quad \text{for } k \in \{1, \dots, m\}, \quad (3.106)$$

noting that it achieves the bound (3.105b).  $\square$

## 3.7 Quantum entropy functions

In this lecture, we define and show the properties of a couple of quantum *entropies*.

### 3.7.1 Definitions of quantum entropy functions

Recall that for any  $X = \sum_{i=1}^n \lambda_i v_i v_i^* \in \text{Pd}(\mathcal{X})$ , we define  $\log(X) = \sum_{i=1}^n \log(\lambda_i) v_i v_i^* \in \text{Herm}(\mathcal{X})$ .

**Definition 3.7.1.** The **von Neumann** (relative) entropy of  $P, Q \in \text{Pd}(\mathcal{X})$  is

$$S(P) := -\langle P, \log_2(P) \rangle, \quad S(P \parallel Q) := \langle P, \log_2(P) - \log_2(Q) \rangle. \quad (3.107)$$

Since the limit  $\lim_{X \downarrow 0} X \log(X) = 0$  exists, we may extend these definitions to  $P, Q \in \text{Pos}(\mathcal{X})$ .

$$S(P) := \lim_{\epsilon \downarrow 0} S(P + \epsilon \mathbb{1}), \quad S(P \parallel Q) := \lim_{\epsilon \downarrow 0} S(P + \epsilon \mathbb{1} \parallel Q + \epsilon \mathbb{1}). \quad (3.108)$$

**Definition 3.7.2.** For  $t \in (0, 1)$ , the **Tsallis** (relative) entropy of  $P, Q \in \text{Pos}(\mathcal{X})$  is

$$S_t(P) := \frac{1}{t} \langle \mathbb{1}, P^{1-t} - P \rangle, \quad S_t(P \parallel Q) := \frac{1}{t} \langle \mathbb{1}, P - P^{1-t} Q^t \rangle. \quad (3.109)$$

We first prove a connection between these two entropies.

**Lemma 3.7.3.** For all  $P \in \text{Pd}(\mathcal{X})$ , we have

$$S(P) = \lim_{t \rightarrow 0} \log_2(e) S_t(P), \quad S(P \parallel Q) = \lim_{t \rightarrow 0} \log_2(e) S_t(P \parallel Q). \quad (3.110)$$

*Proof.* Let  $P, Q \in \text{Pd}(\mathcal{X})$ , and check  $\ln(P) = \lim_{t \rightarrow 0} \frac{1}{t} (P^t - \mathbb{1})$  using L'Hopital's rule. We have

$$\lim_{t \rightarrow 0} \log_2(e) S_t(P \parallel Q) = \log_2(e) \lim_{t \rightarrow 0} \frac{\langle \mathbb{1}, P - P^{1-t} Q^t \rangle}{t} \quad (3.111a)$$

$$= -\log_2(e) \lim_{t \rightarrow 0} \text{Tr} \left( P^{1-t} \left( \frac{P^t - \mathbb{1}}{t} - \frac{Q^t - \mathbb{1}}{t} \right) \right) \quad (3.111b)$$

$$= -\log_2(e) \text{Tr} \left( P (\ln(P) - \ln(Q)) \right) \quad (3.111c)$$

$$= S(P \parallel Q). \quad (3.111d)$$

The other case follows by noting that  $S(P) = -S(P \parallel \mathbb{1})$  and  $S_t(P) = -S_t(P \parallel \mathbb{1})$ .  $\square$

We may now express entropy via an SDP shown in Lemma 3.6.15.

**Corollary 3.7.4.** Let  $P \in \text{Pos}(\mathcal{X})$  and  $m \in \mathbb{N}$ . We have

$$\begin{aligned} S_{2^{-m}}(P \parallel Q) = \text{minimize:} \quad & -2^m \text{vec}(\mathbb{1}_{\mathcal{X} \otimes \mathcal{X}})^* W_m \text{vec}(\mathbb{1}_{\mathcal{X} \otimes \mathcal{X}}) + 2^m \langle \mathbb{1}_{\mathcal{X}}, P \rangle \\ \text{subject to:} \quad & \begin{bmatrix} P \otimes \mathbb{1}_{\mathcal{X}} & W_k \\ W_k & W_{k-1} \end{bmatrix} \succcurlyeq 0 \text{ for all } k \in \{1, \dots, m\} \\ & W_0 = \mathbb{1}_{\mathcal{X}} \otimes Q^T \\ & W_0, \dots, W_m \in \text{Pos}(\mathcal{X} \otimes \mathcal{X}), \end{aligned} \quad (3.112)$$

and analogously for  $S_{2^{-m}}(P)$ , with  $Q^T$  in the constraint replaced by  $\mathbb{1}_{\mathcal{X}}$ , the objective negated, and "minimize" switched to "maximize".

These two SDPs families (indexed by  $m \in \mathbb{N}$ ) give a hierarchy that converges onto the von Neumann entropy. Thus, we can optimize with SDPs (to some approximation level) problems

of the following form:

$$\begin{array}{ll}
\text{maximize:} & S(P) \\
\text{subject to:} & \Phi(P) = B \\
& P \in \text{Pos}(\mathcal{X}) , \\
\text{minimize:} & S(P \parallel Q) \\
\text{subject to:} & \Phi(P, Q) = B \\
& P, Q \in \text{Pos}(\mathcal{X}) .
\end{array} \tag{3.113}$$

### 3.7.2 Properties of quantum entropies

The SDP representation (3.112) yields simpler proofs for the joint convexity of the von Neumann entropy. Strong convexity consists of two conditions, which we show separately.

**Lemma 3.7.5.** *Let  $P_0, P_1, Q_0, Q_1 \in \text{Pos}(\mathcal{X})$ . We have*

$$S(P_0 + P_1 \parallel Q_0 + Q_1) \leq S(P_0 \parallel Q_0) + S(P_1 \parallel Q_1) . \tag{3.114}$$

*Proof.* Let  $(W_0, W_1, \dots, W_m)$  be an optimal solution for  $S_{2-m}(P_0 \parallel Q_0)$  in (3.112) (which exists by compactness). Let  $(V_0, V_1, \dots, V_m)$  be an optimal solution for  $S_{2-m}(P_0 + P_1 \parallel Q_0 + Q_1)$  with the value  $S_{2-m}(P_0 \parallel Q_0) + S_{2-m}(P_1 \parallel Q_1)$ . The proof follows by taking limits.  $\square$

**Lemma 3.7.6.** *Let  $P, Q \in \text{Pos}(\mathcal{X})$  and  $\lambda \geq 0$ . We have*

$$S(\lambda P \parallel \lambda Q) = \lambda S(P \parallel Q) . \tag{3.115}$$

*Proof.* Using the SDP (3.112), we may express  $\lambda S_{2-m}(P \parallel Q)$  by adding  $\lambda$  to the objective, e.g. by prepending to make  $\lambda W_m$  and  $\lambda P$ . By the change of variables  $V_k = \lambda W_k$ , the former of these two may be moved to the terms  $\lambda P \otimes \mathbb{1}_{\mathcal{X}}$  and  $\lambda Q^T$  in the constraints. We now have the same SDP as in (3.112) but with  $P$  and  $Q$  replaced by  $\lambda P$  and  $\lambda Q$ , meaning its value is  $S_{2-m}(\lambda P \parallel \lambda Q)$ . The result holds by taking limits.  $\square$

**Theorem 3.7.7** (Joint convexity). *Let  $P_0, P_1, Q_0, Q_1 \in \text{Pos}(\mathcal{X})$  and  $\lambda_0, \lambda_1 > 0$ . We have that*

$$S(\lambda_0 P_0 + \lambda_1 P_1 \parallel \lambda_0 Q_0 + \lambda_1 Q_1) \leq S(\lambda_0 P_0 \parallel \lambda_0 Q_0) + S(\lambda_1 P_1 \parallel \lambda_1 Q_1) . \tag{3.116}$$

*Proof.* This is verified by combining Lemmas 3.7.5 and 3.7.6.

$$S(\lambda_0 P_0 + \lambda_1 P_1 \parallel \lambda_0 Q_0 + \lambda_1 Q_1) \leq S(\lambda_0 P_0 \parallel \lambda_0 Q_0) + S(\lambda_1 P_1 \parallel \lambda_1 Q_1) \tag{3.117a}$$

$$= \lambda_0 S(P_0 \parallel Q_0) + \lambda_1 S(P_1 \parallel Q_1) . \tag{3.117b}$$

We remark that joint convexity commonly refers to the case when  $\lambda_0 + \lambda_1 = 1$ .  $\square$

We next show a version of the data-processing inequality for the von Neumann entropy. First, we demonstrate it in a restricted case.

**Lemma 3.7.8.** Let  $P, Q \in \text{Pos}(\mathcal{X})$ , and let  $\Phi \in C(\mathcal{X}, \mathcal{Y})$  be unital (meaning  $\Phi(\mathbb{1}) = \mathbb{1}$ ). We have

$$S(\Phi(P) \parallel \Phi(Q)) \leq S(P \parallel Q). \quad (3.118)$$

*Proof.* Recall that since  $\Phi \in C(\mathcal{X}, \mathcal{Y})$  is completely positive and trace-preserving, we have

$$\begin{bmatrix} A & B \\ B^* & C \end{bmatrix} \succcurlyeq 0 \Rightarrow \begin{bmatrix} \Phi(A) & \Phi(B) \\ \Phi(B^*) & \Phi(C) \end{bmatrix} \succcurlyeq 0, \quad \text{Tr}(\Phi(X)) = \text{Tr}(X) \quad \forall X \in L(\mathcal{X}). \quad (3.119)$$

If  $(W_0, \dots, W_m)$  is feasible in the SDP for  $S_{2-m}(P \parallel Q)$ , then it is straightforward to verify that  $((\Phi \otimes \Phi)(W_0), \dots, (\Phi \otimes \Phi)(W_m))$  is feasible in the SDP for  $S_{2-m}(\Phi(P) \parallel \Phi(Q))$ . Comparing the objective function values of these feasible solutions in the limit yields the result.  $\square$

**Lemma 3.7.9.** Let  $P, Q \in \text{Pos}(\mathcal{X})$ , and let  $V \in U(\mathcal{X}, \mathcal{Y})$  be an isometry. We then have

$$S(VPV^* \parallel VQV^*) = S(P \parallel Q). \quad (3.120)$$

*Proof.* Use the SDP (3.112) to express both  $S(VPV^* \parallel VQV^*)$  and  $S(P \parallel Q)$ , noting that the objective functions match. Prove (3.120) by splitting it into two inequalities.

$\leq$ : Show that feasibility in the  $S(P \parallel Q)$  SDP implies feasibility in the  $S(VPV^* \parallel VQV^*)$  SDP.

$$0 \leq \begin{bmatrix} P \otimes \mathbb{1}_{\mathcal{X}} & W_k \\ W_k & W_{k-1} \end{bmatrix} \Rightarrow 0 \leq \begin{bmatrix} V \otimes V & 0 \\ 0 & V \otimes V \end{bmatrix} \begin{bmatrix} P \otimes \mathbb{1}_{\mathcal{X}} & W_k \\ W_k & W_{k-1} \end{bmatrix} \begin{bmatrix} V^* \otimes V^* & 0 \\ 0 & V^* \otimes V^* \end{bmatrix} \quad (3.121a)$$

$$= \begin{bmatrix} VPV^* \otimes VV^* & (V \otimes V)W_k(V^* \otimes V^*) \\ (V \otimes V)W_k(V^* \otimes V^*) & (V \otimes V)W_{k-1}(V^* \otimes V^*) \end{bmatrix} \quad (3.121b)$$

$$\leq \begin{bmatrix} VPV^* \otimes \mathbb{1} & (V \otimes V)W_k(V^* \otimes V^*) \\ (V \otimes V)W_k(V^* \otimes V^*) & (V \otimes V)W_{k-1}(V^* \otimes V^*) \end{bmatrix}. \quad (3.121c)$$

Let  $W'_k = (V \otimes V)W_k(V^* \otimes V^*)$ , so  $W'_0 = (V \otimes V)W_0(V^* \otimes V^*) = VV^* \otimes VQV^* \leq \mathbb{1} \otimes VQV^*$ .

$\geq$ : Show that feasibility in the  $S(VPV^* \parallel VQV^*)$  SDP implies feasibility in the  $S(P \parallel Q)$  SDP.

$$0 \leq \begin{bmatrix} VPV^* \otimes \mathbb{1}_{\mathcal{X}} & W_k \\ W_k & W_{k-1} \end{bmatrix} \Rightarrow 0 \leq \begin{bmatrix} V^* \otimes V^* & 0 \\ 0 & V^* \otimes V^* \end{bmatrix} \begin{bmatrix} VPV^* \otimes \mathbb{1}_{\mathcal{X}} & W_k \\ W_k & W_{k-1} \end{bmatrix} \begin{bmatrix} V \otimes V & 0 \\ 0 & V \otimes V \end{bmatrix} \quad (3.122a)$$

$$= \begin{bmatrix} P \otimes I_{\mathcal{X}} & (V \otimes V)W_k(V^* \otimes V^*) \\ (V \otimes V)W_k(V^* \otimes V^*) & (V \otimes V)W_{k-1}(V^* \otimes V^*) \end{bmatrix}. \quad (3.122b)$$

Let  $W'_k = (V \otimes V)W_k(V^* \otimes V^*)$ , so  $W'_0 = (V^* \otimes V^*)W_0(V \otimes V) = \mathbb{1} \otimes Q^T$ .  $\square$

**Theorem 3.7.10** (Data-processing inequality). Let  $P, Q \in \text{Pos}(\mathcal{X})$ ,  $\Phi \in C(\mathcal{X}, \mathcal{Y})$ . We have

$$S(\Phi(P) \parallel \Phi(Q)) \leq S(P \parallel Q). \quad (3.123)$$

*Proof.* Since  $\Phi$  is a channel, one can show that it may be written in Stinespring representation as  $\Phi(X) = \text{Tr}_{\mathcal{Z}}(VXV^*)$  for some isometry  $V \in \mathcal{U}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Z})$ . Leveraging previous results,

$$S(\Phi(P) \parallel \Phi(Q)) = S(\text{Tr}_{\mathcal{Z}}(VPV^*) \parallel \text{Tr}_{\mathcal{Z}}(VQV^*)) \quad (3.124a)$$

$$\leq S(VPV^* \parallel VQV^*) \quad (\text{by Lemma 3.7.8}), \quad (3.124b)$$

$$= S(P \parallel Q) \quad (\text{by Lemma 3.7.9}). \quad (3.124c)$$

The restricted version of data-processing applies since  $\frac{1}{\dim(\mathcal{Z})} \text{Tr}_{\mathcal{Z}}$  is unital, with subsequent rescaling done via Lemma 3.7.6.  $\square$

**Theorem 3.7.11** (Strong subadditivity). *Let  $\rho \in \mathcal{D}(\mathcal{X} \otimes \mathcal{Y} \otimes \mathcal{Z})$  and denote its subsystems as*

$$\rho_{\mathcal{Z}} := \text{Tr}_{\mathcal{X} \otimes \mathcal{Y}}(\rho), \quad \rho_{\mathcal{XZ}} := \text{Tr}_{\mathcal{Y}}(\rho), \quad \rho_{\mathcal{YZ}} := \text{Tr}_{\mathcal{X}}(\rho). \quad (3.125)$$

*We then have*

$$S(\rho) + S(\rho_{\mathcal{Z}}) \leq S(\rho_{\mathcal{XZ}}) + S(\rho_{\mathcal{YZ}}). \quad (3.126)$$

*Proof.* We prove this for  $\rho \in \text{Pd}(\mathcal{X} \otimes \mathcal{Y} \otimes \mathcal{Z})$  and the rest follows by continuity. Note that

$$S(\rho \parallel \frac{1}{n} \mathbb{1}_{\mathcal{X}} \otimes \rho_{\mathcal{YZ}}) = -S(\rho) + S(\rho_{\mathcal{YZ}}) + \log(n), \quad (3.127a)$$

$$S(\rho_{\mathcal{XZ}} \parallel \frac{1}{n} \mathbb{1}_{\mathcal{X}} \otimes \rho_{\mathcal{Z}}) = -S(\rho_{\mathcal{XZ}}) + S(\rho_{\mathcal{Z}}) + \log(n). \quad (3.127b)$$

By taking partial traces, we have  $S(\rho_{\mathcal{XZ}} \parallel \frac{1}{n} \mathbb{1}_{\mathcal{X}} \otimes \rho_{\mathcal{Z}}) \leq S(\rho \parallel \frac{1}{n} \mathbb{1}_{\mathcal{X}} \otimes \rho_{\mathcal{YZ}})$ .  $\square$

## Chapter 3 exercises

**Exercise 3.1.** Derive the dual quantum state discrimination SDP (3.40).

*Hint:* Introduce the Lagrange multipliers  $Y$  for the affine constraint and  $\Lambda_i \succcurlyeq 0$  for the PSD ones, and consider the Lagrangian function

$$\mathcal{L}(M_1, \dots, M_n, Y, \Lambda_1, \dots, \Lambda_n) = \sum_{i=1}^n p_i \langle \rho_i, M_i \rangle + \left\langle Y, \mathbb{1} - \sum_{i=1}^n M_i \right\rangle + \sum_{i=1}^n \langle \Lambda_i, M_i \rangle. \quad (3.128)$$

**Exercise 3.2.** Show that the primal and dual state discrimination SDPs (3.40) are strictly feasible.

**Exercise 3.3.** Suppose a set of  $n$  quantum states  $\rho_i$  is mutually diagonalizable by the basis  $\{v_1, \dots, v_k\}$ . Define  $M_i$  as in (3.46). Prove that  $(M_1, \dots, M_n)$  is a valid POVM and that it attains  $\alpha = 1$  in (3.40).

**Exercise 3.4.** Verify that the dual fidelity SDP (3.50) follows from the corresponding primal.

*Hint:* By relabeling variables, the primal SDP for the fidelity may alternatively be expressed as

Primal problem

$$\begin{aligned} \alpha = \text{maximize:} \quad & \left\langle \tilde{X}, \frac{1}{2} \begin{bmatrix} 0 & \mathbb{1} \\ \mathbb{1} & 0 \end{bmatrix} \right\rangle \\ \text{subject to:} \quad & X_{1,1} = P \\ & X_{2,2} = Q \\ & \tilde{X} := \begin{bmatrix} X_{1,1} & X_{1,2} \\ X_{2,1} & X_{2,2} \end{bmatrix} \succcurlyeq 0. \end{aligned} \quad (3.129)$$