Semidefinite Programming as an Analytical and Computational Tool for Quantum Information

Marco Túlio Quintino

Sorbonne Université, CNRS, LIP6, F-75005, Paris, France
3rd September 2026

These notes accompany two QALYPSO 2026 lectures on semidefinite programming (SDP) in quantum information and are available, together with the slides for Lectures 1 and 2, on my personal website.

The first lecture explains how to recognise an SDP and formulates minimum-error quantum-state discrimination as one. The second lecture develops dual certificates, strong duality, analytical bounds, numerical solvers, and a simple route from numerical evidence to an exact proof.

These notes are intentionally simple, short, and direct. The goal is to provide a gentle introduction. We refer the reader to more complete and advanced texts below.

Comments and corrections are very welcome at Marco.Quintino@lip6.fr.

Contents

Part I

Lecture 1: What is an SDP?

1 Warming up

Semidefinite programming is built from two objects that already occur throughout quantum theory: linear maps and positive-semidefinite operators. Let us begin by fixing these elementary ideas.

1.1 Linearity and operator spaces

A map $f : \mathcal{H} \to \mathcal{H}'$ between complex vector spaces is linear when

$$f(\alpha x + \beta y) = \alpha f(x) + \beta f(y) \tag{1}$$

for every $x, y \in \mathcal{H}$ and every $\alpha, \beta \in \mathbb{C}$. We write $\mathcal{L}(\mathbb{C}^d)$ for the space of linear operators on $\mathbb{C}^d$. After choosing a basis, an element $X \in \mathcal{L}(\mathbb{C}^d)$ is simply a complex d-by-d matrix.

We will also use linear maps between operator spaces,

$$\Lambda : \mathcal{L}(\mathbb{C}^d) \longrightarrow \mathcal{L}(\mathbb{C}^{d'}). \tag{2}$$

Linearity now means

$$\Lambda(\alpha X + \beta Y) = \alpha\Lambda(X) + \beta\Lambda(Y). \tag{3}$$

This is familiar in quantum theory. A quantum evolution sends a state ρ to another state $\Lambda(\rho)$, and the Born rule is linear in both the state and the measurement effect:

$$p(i \mid \rho, \{M_i\}_i) = \mathrm{tr}(M_i\rho). \tag{4}$$

The adjoint of X is denoted by $X^\dagger$. We call X Hermitian, or self-adjoint, when $X = X^\dagger$.

1.2 Positive-semidefinite operators

An operator $X \in \mathcal{L}(\mathbb{C}^d)$ is *positive semidefinite*, written $X \succeq 0$, when

$$X = X^\dagger \qquad \text{and} \qquad \langle\psi| X |\psi\rangle \geq 0 \quad \text{for every } |\psi\rangle \in \mathbb{C}^d. \tag{5}$$

Equivalently, X is Hermitian and all its eigenvalues are non-negative. The word "semidefinite" allows zero eigenvalues. We write $X \succ 0$ when X is *positive definite*, meaning

$$\langle\psi| X |\psi\rangle > 0 \quad \text{for every non-zero } |\psi\rangle \in \mathbb{C}^d. \tag{6}$$

The use of complex test vectors matters. Positivity of the quadratic form on real vectors alone does not force a real matrix to be symmetric. For example,

$$R = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \qquad \text{satisfies} \qquad x^\mathsf{T} R x = x^\mathsf{T} x \geq 0 \tag{7}$$

for every real vector x, although $R \neq R^\mathsf{T}$. This is why Hermiticity is explicit in Eq. (5).

As a simple example,

$$A = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \succeq 0. \tag{8}$$

Indeed, $A = A^\dagger$, and its eigenvalues are 0 and 2.

1.3 Positivity in quantum theory

A density operator and a positive operator-valued measure (POVM) obey

$$\rho \succeq 0, \qquad \operatorname{tr}(\rho) = 1, \tag{9}$$

and

$$M_i \succeq 0, \qquad \sum_i M_i = \mathbb{1}, \tag{10}$$

respectively. These are already a mixture of positivity and affine equalities, which is precisely the structure that SDP handles.

Quantum channels have the same flavour. Let J_Λ denote the unnormalised Choi operator of a linear map Λ. Then

$$\Lambda \text{ is completely positive} \quad \Longleftrightarrow \quad J_\Lambda \succeq 0, \tag{11}$$

while trace preservation is the affine condition

$$\operatorname{tr}_{\mathrm{out}}(J_\Lambda) = \mathbb{1}_{\mathrm{in}}. \tag{12}$$

Thus states, measurements, and channels all naturally combine linearity, affine normalisation, and positive semidefiniteness.

2 What is an SDP?

An SDP has three ingredients:

$$\text{linear objective} \quad + \quad \text{affine constraints} \quad + \quad \text{positive-semidefinite constraints.} \tag{13}$$

We now put these ingredients into one standard form.

Let

$$A = A^\dagger \in \mathcal{L}(\mathbb{C}^d), \qquad B = B^\dagger \in \mathcal{L}(\mathbb{C}^{d'}), \tag{14}$$

and let

$$\Lambda : \mathcal{L}(\mathbb{C}^d) \longrightarrow \mathcal{L}(\mathbb{C}^{d'}) \tag{15}$$

be a Hermiticity-preserving linear map. This means

$$X = X^\dagger \quad \Longrightarrow \quad \Lambda(X) = \Lambda(X)^\dagger. \tag{16}$$

One standard primal form of an SDP is

$$\begin{aligned} \alpha := \sup_\rho \quad & \operatorname{tr}(A\rho) \\ \text{subject to} \quad & \Lambda(\rho) = B, \\ & \rho \succeq 0. \end{aligned} \tag{17}$$

Here A, B, Λ are fixed data and ρ is the only variable. The objective $\operatorname{tr}(A\rho)$ is linear in ρ, the equality is affine, and the last line is a positive-semidefinite constraint.

An operator ρ satisfying both constraints is a *feasible point*. The SDP is feasible if at least one such operator exists and infeasible otherwise. Every feasible point gives a lower bound on the supremum:

$$\Lambda(\rho) = B, \rho \succeq 0 \quad \Longrightarrow \quad \operatorname{tr}(A\rho) \leq \alpha. \tag{18}$$

We use "supremum" at this stage because feasibility alone does not ensure that an optimal point exists. When the supremum is attained, we may replace sup by max.

The same acronym is used in two closely related ways. *Semidefinite programming* is the field or class of problems; a *semidefinite programme* is one particular optimisation problem. Context normally makes the intended meaning clear.

2.1 Why is it called programming?

The word "programming" here means planning an action under constraints; it does not mean writing computer code. The planning sense predates electronic computers. George Dantzig used it in the title *Programming in a Linear Structure*, published in 1949 [1].

The use of "programming" in linear programming—and hence in semidefinite programming—and its use in computer programming developed independently: neither was named after the other. They nevertheless share the same older root, *programming* in the sense of planning. The slides give earlier computing examples: James W. Bryce's IBM patent, filed in 1937, uses "programing means", and John Mauchly's 1942 memorandum uses "program device". For the historical documents, see the Bryce patent and the Computer History Museum archive.

The terminology aside is useful, but the mathematical point is simpler: an SDP asks us to plan the best feasible choice of a positive-semidefinite matrix.

3 Convex optimisation and the feasible set

For the SDP in Eq. (17), define the feasible set

$$\mathcal{F} := \big\{\rho \in \mathcal{L}(\mathbb{C}^d) \mid \Lambda(\rho) = B,\ \rho \succeq 0\big\}. \tag{19}$$

The problem is simply to maximise the linear function $\rho \mapsto \operatorname{tr}(A\rho)$ over $\mathcal{F}$. Understanding the geometry of this set explains several useful properties of SDP.

3.1 Convex sets and cones

A set $\mathcal{C}$ is convex when

$$X, Y \in \mathcal{C}, \quad t \in [0,1] \quad \Longrightarrow \quad tX + (1-t)Y \in \mathcal{C}. \tag{20}$$

The points $tX + (1-t)Y$ form the line segment joining X and Y. Thus a set is convex precisely when it contains every segment joining two of its points, as illustrated in Fig. 1.

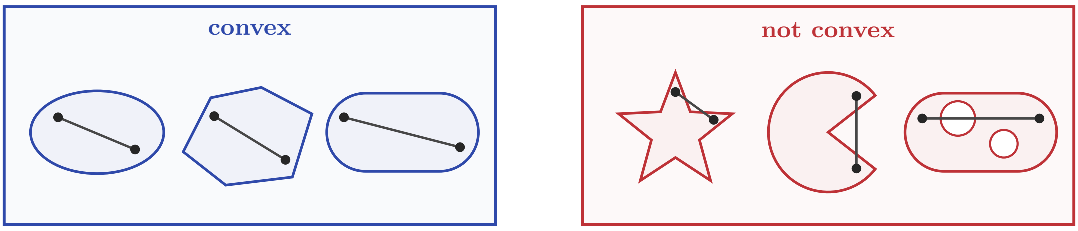

Figure 1: Convex and non-convex sets. The displayed segment remains inside each blue set, but leaves each red set.

The positive-semidefinite operators form the set

$$\mathcal{P}_d := \big\{X \in \mathcal{L}(\mathbb{C}^d) \mid X \succeq 0\big\}. \tag{21}$$

This set is convex. Indeed, if $X, Y \succeq 0$ and $t \in [0,1]$, then

$$\langle\psi|\,\big(tX + (1-t)Y\big)\,|\psi\rangle = t\,\langle\psi|\,X\,|\psi\rangle + (1-t)\,\langle\psi|\,Y\,|\psi\rangle \geq 0 \tag{22}$$

for every $|\psi\rangle$.

It is also a *cone*: if $X \in \mathcal{P}_d$ and $s \geq 0$, then

$$sX \in \mathcal{P}_d. \tag{23}$$

The set of quantum states is a trace-one affine slice of this cone,

$$\mathsf{D}_d = \big\{\rho \in \mathcal{P}_d \mid \operatorname{tr}(\rho) = 1\big\} = \mathcal{P}_d \cap \big\{\rho \mid \operatorname{tr}(\rho) = 1\big\}. \tag{24}$$

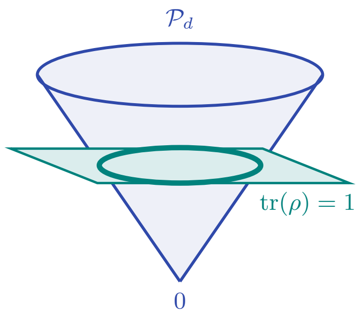

Figure 2: The quantum state space D_d is the trace-one affine slice of the PSD cone $\mathcal{P}_d$.

as illustrated in Fig. 2.

The same picture applies to every SDP. If

$$\mathcal{A}_B := \{\rho \mid \Lambda(\rho) = B\}, \tag{25}$$

then

$$\mathcal{F} = \mathcal{P}_d \cap \mathcal{A}_B. \tag{26}$$

The set $\mathcal{A}_B$ is an affine subspace when it is non-empty. Hence the SDP feasible set is an affine slice of the PSD cone and is convex; see Fig. 3.

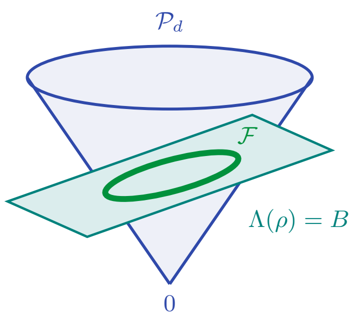

Figure 3: The feasible set $\mathcal{F}$ is the intersection of the PSD cone with the affine subspace $\Lambda(\rho) = B$.

Sometimes there is no objective at all. We may simply ask whether

$$\Lambda(\rho) = B, \qquad \rho \succeq 0 \tag{27}$$

has a solution. This is a *semidefinite feasibility problem* and is usually treated as an SDP as well.

3.2 Linear objectives on convex sets

Convexity removes suboptimal local maxima. Suppose $X \in \mathcal{C}$, the set $\mathcal{C}$ is convex, and there exists $Y \in \mathcal{C}$ with $f(Y) > f(X)$ for a linear function f. For every $t \in (0, 1]$,

$$X_t = (1-t)X + tY \in \mathcal{C}, \qquad f(X_t) = (1-t)f(X) + tf(Y) > f(X). \tag{28}$$

Since $X_t \to X$ as $t \to 0$, the point X cannot be a local maximum. Consequently, every local maximum is global.

There may nevertheless be many different global maximisers. If the maximum exists, define

$$\alpha = \max_{\rho \in \mathcal{C}} \mathrm{tr}(A\rho), \qquad \mathcal{O}^\star := \{\rho \in \mathcal{C} \mid \mathrm{tr}(A\rho) = \alpha\}. \tag{29}$$

The value α is one number, while $\mathcal{O}^\star$ can be a whole face of the feasible set, as in Fig. 4.

Finally, if $\mathcal{C}$ is non-empty, compact, and convex, continuity ensures that a linear objective attains its maximum, and at least one extreme point of $\mathcal{C}$ is optimal. Compactness is essential here: on an open or unbounded convex set, a finite supremum need not be attained.

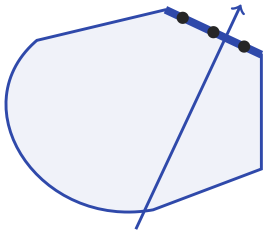

Figure 4: A linear objective can attain its optimum on a whole face. The highlighted edge is $\mathcal{O}^\star$; its endpoints are extreme points.

4 State discrimination

Let

$$\mathcal{E} = \{p_i, \rho_i\}_{i=1}^N \tag{30}$$

be an ensemble of quantum states on $\mathbb{C}^d$, where $p_i \geq 0$, $\sum_{i=1}^N p_i = 1$, and every ρ_i is a density operator. Alice draws a label $i \in \{1, \ldots, N\}$ with probability p_i, prepares ρ_i, and sends it to Bob. Bob knows the ensemble but not the chosen label. He makes a POVM $\{M_j\}_{j=1}^N$ and uses outcome j as his guess.

By the Born rule,

$$\Pr(j \mid i) = \mathrm{tr}(\rho_i M_j). \tag{31}$$

Bob succeeds when $j = i$. Therefore

$$\begin{aligned} p_{\mathrm{succ}}(\{M_j\}) &= \sum_{i=1}^N p_i \sum_{j=1}^N \mathrm{tr}(\rho_i M_j)\delta_{ij} \\ &= \sum_{i=1}^N p_i \, \mathrm{tr}(\rho_i M_i). \end{aligned} \tag{32}$$

Optimising over all POVMs gives

$$\begin{aligned} p^\star_{\mathrm{succ}} = \max_{\{M_i\}_{i=1}^N} \quad & \sum_{i=1}^N p_i \, \mathrm{tr}(\rho_i M_i) \\ \text{subject to} \quad & \sum_{i=1}^N M_i = \mathbb{1}, \\ & M_i \succeq 0 \quad \text{for every } i. \end{aligned} \tag{33}$$

The variable is the measurement, not the ensemble. The objective is linear, the normalisation is affine, and the remaining constraints are PSD.

4.1 Two simple examples

First take two orthogonal states with equal prior probabilities,

$$p_1 = p_2 = \frac{1}{2}, \qquad \rho_1 = |0\rangle\langle 0| \,, \qquad \rho_2 = |1\rangle\langle 1| \,. \tag{34}$$

The computational-basis POVM $M_1 = |0\rangle\langle 0|$, $M_2 = |1\rangle\langle 1|$ succeeds with probability one. Hence $p^\star_{\mathrm{succ}} = 1$.

Now consider the non-orthogonal pair

$$p_1 = p_2 = \frac{1}{2}, \qquad \rho_1 = |0\rangle\langle 0| \,, \qquad \rho_2 = |+\rangle\langle +| \,, \qquad |+\rangle := \frac{|0\rangle + |1\rangle}{\sqrt{2}}. \tag{35}$$

Since $|\langle 0|+\rangle|^2 = 1/2$, the states are not orthogonal and cannot be perfectly discriminated. The computational-basis measurement gives

$$p_{\text{succ}} = \frac{1}{2}(1) + \frac{1}{2}\left(\frac{1}{2}\right) = \frac{3}{4}. \tag{36}$$

This proves the lower bound $p^\star_{\text{succ}} \geq 3/4$, but we can do better.

Define

$$|m_1\rangle := \cos\left(\frac{\pi}{8}\right)|0\rangle - \sin\left(\frac{\pi}{8}\right)|1\rangle = \frac{|0\rangle + |-\rangle}{\sqrt{2+\sqrt{2}}}, \tag{37}$$

$$|m_2\rangle := \sin\left(\frac{\pi}{8}\right)|0\rangle + \cos\left(\frac{\pi}{8}\right)|1\rangle = \frac{|1\rangle + |+\rangle}{\sqrt{2+\sqrt{2}}}, \tag{38}$$

where $|-\rangle := (|0\rangle - |1\rangle)/\sqrt{2}$. The two vectors are orthonormal, so $M_i := |m_i\rangle\langle m_i|$ defines a POVM. Its success probability is

$$\begin{aligned} p_{\text{succ}}(\{M_1, M_2\}) &= \frac{1}{2}|\langle 0|m_1\rangle|^2 + \frac{1}{2}|\langle +|m_2\rangle|^2 \\ &= \cos^2\left(\frac{\pi}{8}\right) = \frac{2+\sqrt{2}}{4} \approx 0.8536. \end{aligned} \tag{39}$$

For the moment, this proves only

$$p^\star_{\text{succ}} \geq \frac{2+\sqrt{2}}{4}. \tag{40}$$

An upper bound is still needed before we can claim optimality.

4.2 Putting the problem into standard form

Although Eq. (33) contains N matrix variables, we can combine them into one block-diagonal operator. Introduce a classical label space $E \simeq \mathbb{C}^N$ with basis $\{|i\rangle_E\}_{i=1}^N$, and set

$$M := \sum_{i=1}^N M_i \otimes |i\rangle\langle i|_E = \bigoplus_{i=1}^N M_i. \tag{41}$$

Then

$$M \succeq 0 \iff M_i \succeq 0 \quad \text{for every } i. \tag{42}$$

Similarly, collect the fixed data into

$$P := \sum_{i=1}^N p_i \rho_i \otimes |i\rangle\langle i|_E = \bigoplus_{i=1}^N p_i \rho_i. \tag{43}$$

The objective becomes

$$\operatorname{tr}(PM) = \sum_{i=1}^N p_i \operatorname{tr}(\rho_i M_i). \tag{44}$$

The partial trace over E is determined by

$$\operatorname{tr}_E(X \otimes Y_E) := X \operatorname{tr}(Y_E). \tag{45}$$

Applying it to Eq. (41) gives

$$\operatorname{tr}_E(M) = \sum_{i=1}^N M_i. \tag{46}$$

Consequently, state discrimination has the standard form

$$p^\star_{\text{succ}} = \max_M \quad \operatorname{tr}(PM) \tag{47}$$

$$\text{subject to} \quad \mathrm{tr}_E(M) = \mathbb{1}, \quad M \succeq 0.$$

The correspondence with Eq. (17) is

$$A = P, \qquad \rho \longleftrightarrow M, \qquad \Lambda = \mathrm{tr}_E, \qquad B = \mathbb{1}. \tag{48}$$

This direct-sum construction is general: finitely many PSD variables can be placed on the diagonal of one larger PSD variable, and finitely many affine equalities can be collected into one linear map.

4.3 Linear objective functions

The trace form of the objective is not restrictive. Let $\mathcal{H}$ be a finite-dimensional Hilbert space over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. The Riesz representation theorem says that, for every linear functional

$$\ell : \mathcal{H} \longrightarrow \mathbb{F}, \tag{49}$$

there is a unique fixed vector $a \in \mathcal{H}$ such that

$$\ell(x) = \langle a, x \rangle \qquad \text{for every } x \in \mathcal{H}. \tag{50}$$

Thus every linear functional is an inner product with a fixed vector.

For matrices, the natural inner product is the Hilbert–Schmidt inner product,

$$\langle A, B \rangle_{\mathrm{HS}} := \mathrm{tr}\big(A^\dagger B\big)\,. \tag{51}$$

The Hermitian operators form a real Hilbert space under this inner product. Consequently, every real-linear functional ℓ on a Hermitian matrix variable ρ has the form

$$\ell(\rho) = \langle A, \rho \rangle_{\mathrm{HS}} = \mathrm{tr}(A\rho) \in \mathbb{R} \tag{52}$$

for a fixed Hermitian operator $A = A^\dagger$. Hence writing the objective as $\mathrm{tr}(A\rho)$ represents every real-linear objective needed here.

5 Duality theory

Every feasible measurement gives an attainable value and hence a lower bound on $p^\star_{\mathrm{succ}}$. Achievability is only half of an optimality proof: we also need a bound that applies to *every* feasible measurement. Duality provides such upper bounds.

Return to the general primal SDP

$$\begin{aligned} \alpha = \sup_{\rho} \quad & \mathrm{tr}(A\rho) \\ \text{subject to} \quad & \Lambda(\rho) = B, \\ & \rho \succeq 0. \end{aligned} \tag{P}$$

Associate a Hermitian multiplier $Y \in \mathcal{L}(\mathbb{C}^{d'})$ with the equality constraint and a PSD multiplier $\sigma \in \mathcal{L}(\mathbb{C}^d)$ with the positivity constraint. We use the Lagrangian

$$L(\rho, Y, \sigma) := \mathrm{tr}(A\rho) + \mathrm{tr}\big[Y(B - \Lambda(\rho))\big] + \mathrm{tr}(\sigma\rho). \tag{53}$$

If ρ is primal feasible and $\sigma \succeq 0$, then the equality term vanishes and

$$\mathrm{tr}(\sigma\rho) \geq 0. \tag{54}$$

The latter follows, for example, from $\mathrm{tr}(\sigma\rho) = \mathrm{tr}(\sigma^{1/2}\rho\sigma^{1/2})$ and positivity of $\sigma^{1/2}\rho\sigma^{1/2}$. Hence

$$\mathrm{tr}(A\rho) \leq L(\rho, Y, \sigma). \tag{55}$$

5.1 The adjoint map and the dual

The adjoint $\Lambda^\dagger$ is defined by the Hilbert–Schmidt pairing:

$$\operatorname{tr}\big[Y\Lambda(\rho)\big] = \operatorname{tr}\big[\Lambda^\dagger(Y)\rho\big]. \tag{56}$$

Using this identity, we can collect every term containing ρ:

$$L(\rho, Y, \sigma) = \operatorname{tr}(BY) + \operatorname{tr}\big[\big(A - \Lambda^\dagger(Y) + \sigma\big)\rho\big]. \tag{57}$$

To obtain a finite upper bound independent of ρ, impose

$$A - \Lambda^\dagger(Y) + \sigma = 0. \tag{58}$$

The resulting dual problem is

$$\begin{aligned} \beta = \inf_{Y,\sigma} \quad & \operatorname{tr}(BY) \\ \text{subject to} \quad & A - \Lambda^\dagger(Y) + \sigma = 0, \\ & \sigma \succeq 0, \qquad Y = Y^\dagger. \end{aligned} \tag{59}$$

The variable σ is redundant. Indeed,

$$\sigma = \Lambda^\dagger(Y) - A \succeq 0 \quad \Longleftrightarrow \quad \Lambda^\dagger(Y) \succeq A. \tag{60}$$

Eliminating it gives the standard dual

$$\begin{aligned} \beta = \inf_{Y=Y^\dagger} \quad & \operatorname{tr}(BY) \\ \text{subject to} \quad & \Lambda^\dagger(Y) \succeq A. \end{aligned} \tag{D}$$

This is itself an SDP.

5.2 Weak duality

Let ρ be primal feasible and Y dual feasible. Then

$$\begin{aligned} \operatorname{tr}(A\rho) &\le \operatorname{tr}\big(\Lambda^\dagger(Y)\rho\big) \\ &= \operatorname{tr}\big(Y\Lambda(\rho)\big) \\ &= \operatorname{tr}(BY). \end{aligned} \tag{61}$$

Taking the supremum over primal points and the infimum over dual points gives

$$\boxed{\alpha \le \beta}. \tag{62}$$

This statement is called *weak duality.* A primal feasible point certifies an achievable lower bound, and a dual feasible point certifies a universal upper bound.

Weak duality does not yet say that $\alpha = \beta$, nor does it guarantee that either optimum is attained. Those are the strong-duality questions of Lecture 2.

Lecture 1 in brief

Quantum theory is built from linear maps and positive operators. An SDP has a linear objective, affine constraints, and positive-semidefinite constraints. Minimum-error state discrimination has exactly this form. Primal feasible points describe attainable strategies; dual feasible points give rigorous upper bounds. Lecture 2 makes this certificate viewpoint concrete.

Part II

Lecture 2: Dual certificates, strong duality, and numerical SDPs

Recap

For the general primal–dual pair introduced in Lecture 1, primal feasible points give lower bounds and dual feasible points give upper bounds. Weak duality guarantees that the former never exceed the latter. We now return to the state-discrimination SDP and derive its dual directly.

6 Duality theory 2.0

We now derive the dual of state discrimination directly from its operational form. This gives a particularly transparent interpretation: one Hermitian operator will upper-bound the success probability of every measurement.

6.1 The state-discrimination dual

Start from the primal problem

$$\begin{aligned} p^\star_{\mathrm{succ}} = \max_{\{M_i\}} \quad & \sum_{i=1}^{N} p_i \operatorname{tr}(M_i \rho_i) \\ \text{subject to} \quad & M_i \succeq 0 \quad \text{for every } i, \\ & \sum_{i=1}^{N} M_i = \mathbb{1}. \end{aligned} \tag{63}$$

Associate a PSD multiplier $\sigma_i \succeq 0$ with each positivity constraint and a Hermitian multiplier $Y = Y^\dagger$ with the normalisation constraint. The Lagrangian is

$$\begin{aligned} L(\{M_i\}, Y, \{\sigma_i\}) :=& \sum_i p_i \operatorname{tr}(M_i \rho_i) + \operatorname{tr}\left[Y\left(\mathbb{1} - \sum_i M_i\right)\right] + \sum_i \operatorname{tr}(\sigma_i M_i) \\ =& \operatorname{tr}(Y) + \sum_i \operatorname{tr}[M_i(p_i\rho_i - Y + \sigma_i)]\,. \end{aligned} \tag{64}$$

Once the constraints have entered the Lagrangian, the matrices M_i are otherwise unrestricted Hermitian matrices. The supremum over them is finite only if

$$p_i \rho_i - Y + \sigma_i = 0 \qquad \text{for every } i. \tag{65}$$

Consequently,

$$\sigma_i = Y - p_i \rho_i \succeq 0 \quad \Longleftrightarrow \quad Y \succeq p_i \rho_i. \tag{66}$$

Eliminating all slack variables gives

$$\begin{aligned} \beta = \min_{Y = Y^\dagger} \quad & \operatorname{tr}(Y) \\ \text{subject to} \quad & Y \succeq p_i \rho_i \quad \text{for every } i. \end{aligned} \tag{D-SD}$$

The certificate interpretation can also be seen without the Lagrangian. If $Y \succeq p_i \rho_i$ for every i, then every POVM obeys

$$\begin{aligned} \sum_i p_i \operatorname{tr}(\rho_i M_i) \le& \sum_i \operatorname{tr}(Y M_i) \\ =& \operatorname{tr}\left(Y \sum_i M_i\right) = \operatorname{tr}(Y). \end{aligned} \tag{67}$$

Thus a feasible Y is a universal upper-bound certificate.

6.2 The binary example: matching certificates

Return to the equal-prior ensemble in Eq. (35). The POVM defined by Eqs. (37) and (38) achieves

$$p_{\text{succ}}(\{M_1, M_2\}) = \frac{2+\sqrt{2}}{4}. \tag{68}$$

Consider the Hermitian operator

$$Y^\star := \frac{1}{4}\left(\rho_1 + \rho_2 + \frac{\mathbb{1}}{\sqrt{2}}\right) = \frac{1}{8}\begin{pmatrix} 3+\sqrt{2} & 1 \\ 1 & 1+\sqrt{2} \end{pmatrix}. \tag{69}$$

Its two dual slacks have exact rank-one factorisations:

$$Y^\star - \frac{1}{2}\rho_1 = \frac{1}{2\sqrt{2}} M_2 \succeq 0, \qquad Y^\star - \frac{1}{2}\rho_2 = \frac{1}{2\sqrt{2}} M_1 \succeq 0. \tag{70}$$

Hence $Y^\star$ is dual feasible. Its objective value is

$$\operatorname{tr}(Y^\star) = \frac{1}{4}\left(1 + 1 + \frac{2}{\sqrt{2}}\right) = \frac{2+\sqrt{2}}{4}. \tag{71}$$

Weak duality now sandwiches the optimum between equal numbers:

$$\frac{2+\sqrt{2}}{4} = p_{\text{succ}}(\{M_1, M_2\}) \le p^\star_{\text{succ}} \le \operatorname{tr}(Y^\star) = \frac{2+\sqrt{2}}{4}. \tag{72}$$

We have therefore proved

$$\boxed{p^\star_{\text{succ}} = \frac{2+\sqrt{2}}{4}}. \tag{73}$$

Notice the logic: feasibility of the measurement gives the lower bound, feasibility of $Y^\star$ gives the upper bound, and equality proves optimality. No general strong-duality theorem was needed for this particular certificate.

6.3 Strong duality and Slater's theorem

For the general pair, weak duality gives $\alpha \le \beta$. When

$$\alpha = \beta, \tag{74}$$

we say that *strong duality* holds. If both values are finite, the quantity $\beta - \alpha \ge 0$ is the duality gap.

The following finite-dimensional version of Slater's theorem is the one we will use; see, for example, [2].

Theorem 6.1 (Slater condition)**.** *Consider the primal SDP in Eq. (P). If its value is bounded above and there exists a strictly feasible point*

$$\rho \succ 0, \qquad \Lambda(\rho) = B, \tag{75}$$

then $\alpha = \beta$ and the dual optimum is attained.

Theorem 6.2 (Slater conditions 2)**.** *If both the primal and the dual SDPs are strictly feasible, then strong duality holds. Equivalently,*

$$\alpha = \beta. \tag{76}$$

State discrimination satisfies these conditions explicitly. On the primal side,

$$M_i = \frac{\mathbb{1}}{N} \succ 0, \qquad \sum_{i=1}^{N} M_i = \mathbb{1}. \tag{77}$$

On the dual side, choose

$$Y = \lambda \mathbb{1}, \qquad \lambda > \max_i p_i. \tag{78}$$

Every density operator obeys $\rho_i \preceq \mathbb{1}$, so

$$Y - p_i \rho_i \succeq (\lambda - p_i)\mathbb{1} \succ 0 \qquad \text{for every } i. \tag{79}$$

The primal value is also bounded above by one. We conclude that

$$\max_{\substack{M_i \succeq 0 \\ \sum_i M_i = \mathbb{1}}} \sum_i p_i \operatorname{tr}(\rho_i M_i) = \min_{\substack{Y = Y^\dagger \\ Y \succeq p_i \rho_i \ \forall i}} \operatorname{tr}(Y), \tag{80}$$

and both optima are attained.

7 Analytical applications

Dual feasible points are not merely formal variables. A simple choice can already prove a useful theorem.

7.1 A universal dimension bound

Theorem 7.1. *For every ensemble $\{p_i, \rho_i\}_{i=1}^N$ on a d-dimensional Hilbert space,*

$$p^\star_{\mathrm{succ}} \le d \max_i p_i. \tag{81}$$

Proof. Since every density operator satisfies $0 \preceq \rho_i \preceq \mathbb{1}$,

$$p_i \rho_i \preceq p_i \mathbb{1} \preceq \left(\max_j p_j\right) \mathbb{1}. \tag{82}$$

Therefore

$$Y := \left(\max_j p_j\right) \mathbb{1} \tag{83}$$

is feasible for the dual in Eq. (D-SD). Its trace is $d \max_j p_j$, which proves the claim. □

Of course, a success probability is also at most one. The slightly sharper way to record the conclusion is

$$p^\star_{\mathrm{succ}} \le \min\left\{1, d \max_i p_i\right\}. \tag{84}$$

For uniform priors $p_i = 1/N$, the dual certificate is $Y = \mathbb{1}/N$, and

$$p^\star_{\mathrm{succ}} \le \frac{d}{N}. \tag{85}$$

7.2 Uniform pure-state tight frames

We can identify exactly when the bound in Eq. (85) is attained by an important family of pure states.

Theorem 7.2. *Let $\{|\psi_i\rangle\}_{i=1}^N \subset \mathbb{C}^d$ be pure states with uniform priors. If*

$$\frac{1}{N} \sum_{i=1}^N |\psi_i\rangle\langle\psi_i| = \frac{\mathbb{1}}{d}, \tag{86}$$

then

$$p^\star_{\mathrm{succ}} = \frac{d}{N}. \tag{87}$$

Proof. Write $\rho_i = |\psi_i\rangle\langle\psi_i|$ and define

$$M_i := \frac{d}{N}\rho_i. \tag{88}$$

Every M_i is PSD, and Eq. (86) gives $\sum_i M_i = \mathbb{1}$. Hence $\{M_i\}$ is a POVM. Its success probability is

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}\operatorname{tr}(\rho_i M_i) &= \frac{d}{N^2}\sum_{i=1}^{N}\operatorname{tr}(\rho_i^2) \\ &= \frac{d}{N},\end{aligned} \tag{89}$$

where the last equality uses purity. This lower bound matches the dual upper bound from Eq. (85). □

The purity assumption is important: for mixed states, $\operatorname{tr}(\rho_i^2)$ need not equal one.

One can ask the same question when k identical copies of the unknown state are available. Which sets of states are the most discriminable in this multicopy regime? The symmetric-subspace argument in Appendix C gives a simple first answer. Readers interested in this question may wish to consult Kvashchuk et al., *The most discriminable quantum states in the multicopy regime* [3].

7.3 The trine states

Let $\omega = e^{2\pi i/3}$ and define three qubit states

$$|\psi_i\rangle := \frac{|0\rangle + \omega^{i-1}|1\rangle}{\sqrt{2}}, \qquad \rho_i := |\psi_i\rangle\langle\psi_i|, \qquad i = 1,2,3. \tag{90}$$

Their Bloch vectors are separated by 120° in the equatorial plane, and

$$\frac{1}{3}\sum_{i=1}^{3}\rho_i = \frac{\mathbb{1}}{2}. \tag{91}$$

The tight-frame theorem gives the exact primal–dual pair

$$M_i = \frac{2}{3}\rho_i, \qquad Y = \frac{\mathbb{1}}{3}, \qquad \boxed{p^\star_{\text{succ}} = \frac{2}{3}}. \tag{92}$$

7.4 The six Pauli eigenstates

Let $\sigma_x, \sigma_y, \sigma_z$ be the Pauli matrices. For $a \in \{x, y, z\}$ and $s \in \{+1, -1\}$, define

$$\rho_{a,s} := \frac{1}{2}(\mathbb{1} + s\sigma_a). \tag{93}$$

These are the six Pauli eigenstates. Their Bloch vectors are the vertices of an octahedron. Each antipodal pair sums to the identity, and hence

$$\frac{1}{6}\sum_{a\in\{x,y,z\}}\sum_{s\in\{+1,-1\}}\rho_{a,s} = \frac{\mathbb{1}}{2}. \tag{94}$$

Again applying the tight-frame theorem,

$$M_{a,s} = \frac{1}{3}\rho_{a,s}, \qquad Y = \frac{\mathbb{1}}{6}, \qquad \boxed{p^\star_{\text{succ}} = \frac{1}{3}}. \tag{95}$$

The displayed POVM is optimal, but it is not the only optimal POVM.

8 Numerical solvers

SDPs admit systematic numerical methods. A solver receives a finite representation of the matrix data and an accuracy request, then attempts to return approximate primal and dual candidates, objective values, and diagnostics. It may instead report infeasibility, unboundedness, or numerical difficulty.

The word *efficient* needs some care. Under standard regularity and finite-precision assumptions, theoretical SDP algorithms have polynomial scaling in the encoded input size and in $\log(1/\varepsilon)$, where ε is the requested accuracy. This does not mean that every SDP is fast in practice. The matrix-block sizes, number and structure of affine constraints, sparsity, conditioning, requested accuracy, and choice of algorithm all matter.

8.1 Languages, modelling layers, and solvers

Languages and modelling layers can be paired as follows:

Language	Modelling layer
Julia	JuMP
MATLAB	CVX/YALMIP
Python	CVXPY

Solvers form a separate layer. They are not tied to a single programming language, although a modelling layer needs an interface for the chosen solver. Some common examples are:

Solver	Availability
Clarabel	open source
SCS	open source
MOSEK	commercial; free academic licences

The language stores data and runs the programme. The modelling layer lets us write the optimisation problem close to its mathematical form and converts it to a representation accepted by the solver. The solver implements the numerical algorithm. The lecture demonstrations use Julia, JuMP [4], and Clarabel.

Quantum-information libraries can simplify common tasks such as partial traces, partial transposition, state discrimination, and entanglement tests. Examples are Ket.jl for Julia, QETLAB for MATLAB, and toqito for Python. This is not an exhaustive list, but rather a few libraries I have used and liked.

8.2 A short JuMP model

The state-discrimination primal and dual can be written as two short functions. The following is the code shown in the lecture.

```
using LinearAlgebra
using JuMP
import Clarabel

function state_discrimination(rho, p)
    N, d = length(rho), size(rho[1], 1)
    model = Model(Clarabel.Optimizer)
    M = [@variable(model, [1:d, 1:d] in HermitianPSDCone())
         for i in 1:N]
    @constraint(model, sum(M) == I)
    @objective(model, Max,
        sum(p[i] * real(tr(rho[i] * M[i])) for i in 1:N))
    optimize!(model)
    return objective_value(model), value.(M)
end

function state_discrimination_dual(rho, p)
```

```
    N, d = length(rho), size(rho[1], 1)
    model = Model(Clarabel.Optimizer)
    @variable(model, Y[1:d, 1:d] in HermitianPSDCone())
    @constraint(model, [i = 1:N],
        Hermitian(Y - p[i] * rho[i]) in HermitianPSDCone())
    @objective(model, Min, real(tr(Y)))
    optimize!(model)
    return objective_value(model), value.(Y)
end
```

Declaring $Y \succeq 0$ in the code is redundant but harmless: the constraints $Y \succeq p_i\rho_i \succeq 0$ already imply it. The calls to $\operatorname{Re}\operatorname{tr}(\cdot)$ suppress negligible imaginary round-off in a quantity that is mathematically real.

The functions above are deliberately minimal for presentation. In a working script, also inspect the termination status, feasibility status, residuals, and solver tolerances before using the returned values.

8.3 Two small computations

For the equal-prior $|0\rangle, |+\rangle$ ensemble, the tested environment Julia 1.12.6, JuMP 1.31.1, and Clarabel 0.11.1 returns

quantity	numerical value
primal candidate objective	0.8535533693557011
dual candidate objective	0.8535533897770226
exact target	0.8535533905932737

The exact target is $(2+\sqrt{2})/4$, proved in Eq. (73). The small difference between the two numerical objectives is about 2.04×10^{-8}.

For the six uniformly distributed Pauli eigenstates, the same environment returns

quantity	numerical value
primal candidate objective	0.3333333325055029
dual candidate objective	0.3333333329815097
exact target	0.3333333333333333

Similarly, after a careful analysis of the numerical results, it is possible to guess an analytical form for the variables. For instance, in this case, we may have

$$M_{a,s} = \frac{1}{3}\rho_{a,s}, \qquad Y = \frac{\mathbb{1}}{6}.$$

These candidates are proved optimal in Eq. (95). A solver need not return this particular POVM because the optimal measurement is not unique.

Numerical arithmetic is powerful and reliable. Nevertheless, a displayed decimal is not by itself a mathematical proof. The next section explains how to turn useful numerical structure into a certificate.

9 Computer-assisted proofs

A floating-point calculation can provide excellent evidence, but the logical conclusion depends on what has actually been verified. This section gives a simple proof-oriented workflow for SDP.

9.1 What floating-point numbers forget

A normal binary floating-point number has the schematic form

$$(-1)^s(1.b_1b_2\dots b_k)_2\, 2^e. \tag{96}$$

Only finitely many bit patterns are available. Consequently, only finitely many real numbers can be represented exactly, and most inputs are rounded.

For IEEE binary64 arithmetic, the two distinct integers

$$x = 2^{53} = 9\,007\,199\,254\,740\,992, \qquad y = x + 1 \tag{97}$$

are converted to the same `Float64` value. Distinct exact inputs can therefore become indistinguishable.

Floating-point addition is also not associative. With

$$a = 10^{16}, \qquad b = -10^{16}, \qquad c = 1, \tag{98}$$

exact arithmetic gives both grouped sums equal to one, whereas binary64 arithmetic gives

$$(a + b) + c = 1.0, \qquad a + (b + c) = 0.0. \tag{99}$$

These examples are not an argument against numerical optimisation. They are a reminder to separate an approximate candidate from a verified theorem.

9.2 The solver returns useful matrices

For the binary example, the solver returns not only objective values but also matrices with a recognisable pattern:

$$M_1^{\mathrm{num}} \approx \begin{pmatrix} 0.853551 & -0.353555 \\ -0.353555 & 0.146449 \end{pmatrix}, \qquad M_2^{\mathrm{num}} \approx \begin{pmatrix} 0.146449 & 0.353555 \\ 0.353555 & 0.853551 \end{pmatrix}, \tag{100}$$

and

$$Y^{\mathrm{num}} \approx \begin{pmatrix} 0.551772 & 0.124996 \\ 0.124996 & 0.301781 \end{pmatrix}. \tag{101}$$

The entries suggest the exact objects from Eqs. (37), (38) and (69). Suggestion is not certification: every relevant constraint must still be checked.

9.3 A recipe for a dual upper bound

Suppose the problem data are exact, or have rigorous enclosures. A robust dual workflow is:

1. round or reconstruct an exact Hermitian candidate $\widetilde{Y} = \widetilde{Y}^\dagger$;
2. obtain an exact or controlled-error number $\varepsilon \geq 0$ such that
$$\lambda_{\min}(\widetilde{Y} - p_i\rho_i) \geq -\varepsilon \quad \text{for every } i; \tag{102}$$
3. set
$$Y^{\mathrm{cert}} := \widetilde{Y} + \varepsilon\mathbb{1}; \tag{103}$$
4. evaluate its objective exactly.

By Eq. (102),

$$Y^{\mathrm{cert}} - p_i\rho_i \succeq 0 \quad \text{for every } i, \tag{104}$$

so Y^{cert} is genuinely dual feasible and

$$p^\star_{\mathrm{succ}} \leq \mathrm{tr}(Y^{\mathrm{cert}}) = \mathrm{tr}(\widetilde{Y}) + d\varepsilon. \tag{105}$$

Sometimes, after a careful analysis of the numerical results, it is possible to guess an analytical form for the variables. For instance, in this case, we may have

$$Y^{\mathrm{exact}} := \frac{1}{8}\begin{pmatrix} 3 + \sqrt{2} & 1 \\ 1 & 1 + \sqrt{2} \end{pmatrix}. \tag{106}$$

The rank-one factorisations in Eq. (70) prove exact dual feasibility, and therefore

$$p^\star_{\mathrm{succ}} \leq p^\star_{\mathrm{upper}} := \mathrm{tr}(Y^{\mathrm{exact}}) = \frac{2 + \sqrt{2}}{4}. \tag{107}$$

9.4 A recipe for a primal lower bound

The analogous primal workflow is:

1. reconstruct exact Hermitian matrices $\widetilde{M}_i = \widetilde{M}_i^\dagger$;
2. repair the affine normalisation by defining

$$R := \mathbb{1} - \sum_j \widetilde{M}_j, \qquad \widehat{M}_i := \widetilde{M}_i + \frac{R}{N}. \tag{108}$$

   Then $\sum_i \widehat{M}_i = \mathbb{1}$ exactly;

3. prove, with exact or controlled-error calculations, that

$$\lambda_{\min}(\widehat{M}_i) \geq -\delta \quad \text{for every } i \tag{109}$$

   for some exact $\delta \geq 0$;

4. choose an exact $\eta \in [0, 1]$ satisfying

$$\eta \geq \frac{N\delta}{1 + N\delta} \tag{110}$$

   and mix with the uniform POVM:

$$M_i^{\text{cert}} := (1 - \eta)\widehat{M}_i + \eta \frac{\mathbb{1}}{N}; \tag{111}$$

5. evaluate the primal objective exactly.

The matrices in Eq. (111) still sum to the identity, and

$$\lambda_{\min}(M_i^{\text{cert}}) \geq -(1 - \eta)\delta + \frac{\eta}{N} \geq 0. \tag{112}$$

Thus they form a genuine POVM and their exact objective value is a rigorous lower bound. This repair pattern is a state-discrimination specialisation of the more general certification methods described in [5, Appendix C].

For our binary example, the exact reconstruction is

$$M_1^{\text{exact}} := \frac{1}{4}\begin{pmatrix} 2+\sqrt{2} & -\sqrt{2} \\ -\sqrt{2} & 2-\sqrt{2} \end{pmatrix}, \qquad M_2^{\text{exact}} := \frac{1}{4}\begin{pmatrix} 2-\sqrt{2} & \sqrt{2} \\ \sqrt{2} & 2+\sqrt{2} \end{pmatrix}. \tag{113}$$

These matrices equal $|m_1\rangle\langle m_1|$ and $|m_2\rangle\langle m_2|$, so they are PSD, and direct addition gives

$$M_1^{\text{exact}} + M_2^{\text{exact}} = \mathbb{1}. \tag{114}$$

Their objective value is

$$p^\star_{\text{lower}} := \frac{1}{2}\operatorname{tr}(\rho_1 M_1^{\text{exact}}) + \frac{1}{2}\operatorname{tr}(\rho_2 M_2^{\text{exact}}) = \frac{2+\sqrt{2}}{4}. \tag{115}$$

Combining Eqs. (107) and (115),

$$p^\star_{\text{lower}} \leq p^\star_{\text{succ}} \leq p^\star_{\text{upper}} \quad \Longrightarrow \quad p^\star_{\text{succ}} = \frac{2+\sqrt{2}}{4}. \tag{116}$$

Reconstruction does not always work. Exact-looking entries remain conjectural until all constraints and objective values are checked. The problem data, and the bounds δ and ε, must themselves be exact or rigorously controlled. Ordinary floating-point eigenvalues and residuals are not enough for this final step.

10 Outlook

Let us collect the main messages.

- A primal feasible point describes an attainable strategy and gives a lower bound.
- A dual feasible point is an upper-bound certificate.
- Matching primal and dual values prove optimality; Slater conditions often guarantee that the optimal values match.
- Numerical solvers provide useful approximate candidates and diagnostics.
- Numerical structure can sometimes be reconstructed and verified as an exact computer-assisted proof.

State discrimination is visibly an SDP once the measurement is recognised as the variable. In other problems the SDP structure may be hidden. One may need to introduce auxiliary variables, use block matrices, take a dual, exploit a Schur complement, or otherwise reformulate the problem. When no single exact SDP formulation is available, a hierarchy of SDPs may still give systematically improving upper or lower bounds.

SDP methods occur across quantum information, including:

- entanglement detection and steering [6, 7];
- measurement compatibility [8];
- Bell nonlocality and correlations in networks [9, 10];
- quantum marginal problems [11];
- state and channel discrimination [2, 5];
- bounds on quantum channel capacities [12];
- quantum metrology [13];
- higher-order quantum computation and indefinite causal order [14, 15].

These lectures provide the common SDP language. Jessica Bavaresco's lectures develop several of these applications, together with hierarchy and see-saw techniques.

For a focused continuation in quantum information, we recommend Skrzypczyk and Cavalcanti's *Semidefinite Programming in Quantum Information Science* [2]. For a broader introduction to convex optimisation, we recommend Boyd and Vandenberghe's *Convex Optimization* [16], which is freely available from the authors' website.

Appendix

A A stronger argument for the dual constraint

In Section 5, we moved rather quickly from the Lagrangian to the condition

$$A - \Lambda^{\dagger}(Y) + \sigma = 0. \tag{117}$$

This appendix derives that condition explicitly. The key point is simple: after the primal constraints have been encoded in the Lagrangian, the dual function evaluates the same expression over every Hermitian operator $H = H^{\dagger}$. A non-zero coefficient of H would therefore make this maximisation unbounded.

A.1 Recall the primal and dual problems

Let

$$A = A^{\dagger} \in \mathcal{L}(\mathbb{C}^{d}), \qquad B = B^{\dagger} \in \mathcal{L}(\mathbb{C}^{d'}), \tag{118}$$

and let

$$\Lambda : \mathcal{L}(\mathbb{C}^{d}) \longrightarrow \mathcal{L}(\mathbb{C}^{d'}) \tag{119}$$

be a Hermiticity-preserving linear map. The primal SDP is

$$\begin{aligned} \alpha = \sup_{\rho} \quad & \operatorname{tr}(A\rho) \\ \text{subject to} \quad & \Lambda(\rho) = B, \\ & \rho \succeq 0. \end{aligned} \tag{P}$$

Thus ρ is the variable, while A, B, and Λ are fixed data. Every primal feasible ρ gives an attainable value $\operatorname{tr}(A\rho)$, hence a lower bound on α.

The dual problem derived in the main text is

$$\begin{aligned} \beta = \inf_{Y=Y^{\dagger}} \quad & \operatorname{tr}(BY) \\ \text{subject to} \quad & \Lambda^{\dagger}(Y) \succeq A. \end{aligned} \tag{D}$$

Every dual feasible Y gives an upper bound on every primal feasible value. Our aim is to see explicitly why the constraint $\Lambda^{\dagger}(Y) \succeq A$ appears.

A.2 Build the Lagrangian

We introduce one multiplier for each primal constraint:

- a Hermitian operator $Y = Y^{\dagger}$ for the equality $\Lambda(\rho) = B$;
- a positive-semidefinite operator $\sigma \succeq 0$ for the cone constraint $\rho \succeq 0$.

With the signs appropriate for our maximisation problem, the Lagrangian is

$$L(\rho, Y, \sigma) := \operatorname{tr}(A\rho) + \operatorname{tr}\bigl[Y(B - \Lambda(\rho))\bigr] + \operatorname{tr}(\sigma\rho). \tag{120}$$

Suppose for a moment that ρ is primal feasible. The equality term then vanishes. Moreover, $\rho \succeq 0$ and $\sigma \succeq 0$ imply

$$\operatorname{tr}(\rho\sigma) \geq 0. \tag{121}$$

Consequently,

$$\operatorname{tr}(A\rho) \leq L(\rho, Y, \sigma) \qquad \text{for every primal feasible } \rho \text{ and every } \sigma \succeq 0. \tag{122}$$

The Lagrangian is therefore a candidate upper bound.

To expose its dependence on ρ, recall that the adjoint map is defined by

$$\operatorname{tr}\bigl[Y\Lambda(\rho)\bigr] = \operatorname{tr}\bigl[\Lambda^{\dagger}(Y)\rho\bigr]. \tag{123}$$

Substituting this identity into Eq. (120) gives

$$L(\rho, Y, \sigma) = \operatorname{tr}(BY) + \operatorname{tr}\left[\bigl(A - \Lambda^{\dagger}(Y) + \sigma\bigr)\rho\right]. \tag{124}$$

A.3 Why the dual cannot depend on the primal variable

At this stage, Y and σ are the proposed dual variables. A fixed dual candidate should give one number that bounds the objective value of *every* primal feasible point. This number must therefore depend only on Y and σ, not on a still-undetermined primal variable.

We remove the primal variable by taking the worst-case value of the Lagrangian. To distinguish this maximisation variable from the actual primal variable ρ, denote it by H and define

$$g(Y, \sigma) := \sup_{H=H^{\dagger}} L(H, Y, \sigma). \tag{125}$$

The operator H ranges over all Hermitian operators and need not be PSD. This is deliberate: both primal restrictions have already been represented by multiplier terms in the Lagrangian. Only the actual primal feasible operator ρ_f must be PSD. Enlarging the set over which we maximise cannot spoil the upper-bound property. Indeed,

$$\mathrm{tr}(A\rho_\mathrm{f}) \le L(\rho_\mathrm{f}, Y, \sigma) \le g(Y, \sigma) \qquad \text{for every primal feasible } \rho_\mathrm{f}. \tag{126}$$

Thus, whenever $g(Y, \sigma)$ is finite, it is the desired single upper bound depending only on the dual variables.

Set

$$C := A - \Lambda^\dagger(Y) + \sigma. \tag{127}$$

The operator C is Hermitian, and for every Hermitian H,

$$L(H, Y, \sigma) = \mathrm{tr}(BY) + \mathrm{tr}(CH). \tag{128}$$

If $C \ne 0$, choose the allowed Hermitian operator

$$H = tC, \qquad t > 0. \tag{129}$$

Then

$$L(H, Y, \sigma) = \mathrm{tr}(BY) + t\,\mathrm{tr}(C^2). \tag{130}$$

Since C is Hermitian,

$$\mathrm{tr}(C^2) = \|C\|_{\mathrm{HS}}^2 > 0 \qquad \text{whenever } C \ne 0. \tag{131}$$

Letting t tend to infinity therefore gives

$$L(H, Y, \sigma) \longrightarrow +\infty. \tag{132}$$

Hence $g(Y, \sigma)$ is finite only if $C = 0$, which is precisely Eq. (117). Conversely, if $C = 0$, all dependence on H disappears and

$$g(Y, \sigma) = \mathrm{tr}(BY). \tag{133}$$

We have proved the complete expression

$$g(Y, \sigma) = \begin{cases} \mathrm{tr}(BY), & A - \Lambda^\dagger(Y) + \sigma = 0, \\ +\infty, & \text{otherwise.} \end{cases} \tag{134}$$

A.4 Recover the dual SDP

We now minimise the finite upper bounds supplied by the dual function. The finite-value condition can be rearranged as

$$\sigma = \Lambda^\dagger(Y) - A. \tag{135}$$

Because $\sigma \succeq 0$, this is equivalent to

$$\Lambda^\dagger(Y) \succeq A. \tag{136}$$

Therefore

$$\begin{aligned} \inf_{\substack{Y=Y^\dagger \\ \sigma\succeq 0}} g(Y, \sigma) &= \inf_{\substack{Y=Y^\dagger,\, \sigma\succeq 0 \\ A-\Lambda^\dagger(Y)+\sigma=0}} \mathrm{tr}(BY) \\ &= \inf_{\substack{Y=Y^\dagger \\ \Lambda^\dagger(Y)\succeq A}} \mathrm{tr}(BY) = \beta. \end{aligned} \tag{137}$$

This is exactly the dual problem in Eq. (D).

Finally, Eq. (126) shows directly that every finite dual value is an upper bound on every primal value. Taking the supremum over primal feasible points and then the infimum over dual feasible points gives

$$\alpha \le \beta, \tag{138}$$

which is weak duality. Equality $\alpha = \beta$ is a separate question; it requires the strong-duality conditions discussed in Section 6.3.

The essential point is worth repeating. A dual candidate must give a bound that depends only on the dual variables. Taking the supremum over every Hermitian H removes the primal variable from the problem. Any non-zero coefficient of H can be amplified along its own direction, making the proposed upper bound infinite. The finite-value condition is exactly what makes the Lagrangian independent of the primal variable.

B Linear programming as a special case of SDP

A linear programme (LP) is an optimisation problem with a real vector variable, a linear objective, and linear equality and inequality constraints. After changing the sign of a minimisation objective, introducing slack variables, and splitting any unrestricted variable into the difference of two non-negative variables, every finite-dimensional LP can be written in the standard form

$$\begin{aligned} \ell := \sup_{x\in\mathbb{R}^n} \quad & c^{\mathsf{T}}x \\ \text{subject to} \quad & Gx = b, \\ & x \geq 0. \end{aligned} \tag{139}$$

Here $c \in \mathbb{R}^n$, $G \in \mathbb{R}^{m\times n}$, and $b \in \mathbb{R}^m$ are fixed data, while $x \in \mathbb{R}^n$ is the variable. The last inequality is understood componentwise:

$$x \geq 0 \quad \Longleftrightarrow \quad x_i \geq 0 \quad \text{for every } i. \tag{140}$$

B.1 The diagonal case

We now write Eq. (139) as an SDP. Define the diagonal operators

$$A := \mathrm{Diag}(c_1, \dots, c_n), \qquad B := \mathrm{Diag}(b_1, \dots, b_m). \tag{141}$$

Thus both A and B are diagonal. For a Hermitian operator $\rho \in \mathcal{L}(\mathbb{C}^n)$, let

$$\mathrm{diag}(\rho) := (\rho_{11}, \dots, \rho_{nn})^{\mathsf{T}} \tag{142}$$

and define the Hermiticity-preserving linear map

$$\Lambda(\rho) := \mathrm{Diag}\big(G\,\mathrm{diag}(\rho)\big). \tag{143}$$

The general SDP

$$\begin{aligned} \sup_{\rho} \quad & \mathrm{tr}(A\rho) \\ \text{subject to} \quad & \Lambda(\rho) = B, \\ & \rho \succeq 0 \end{aligned} \tag{144}$$

now depends only on the diagonal entries of ρ. Indeed, if $x := \mathrm{diag}(\rho)$, then

$$\mathrm{tr}(A\rho) = c^{\mathsf{T}}x, \qquad \Lambda(\rho) = B \quad \Longleftrightarrow \quad Gx = b. \tag{145}$$

Moreover, positivity implies that every diagonal entry is non-negative:

$$\rho \succeq 0 \quad \Longrightarrow \quad x_i = \langle i|\rho|i\rangle \geq 0 \quad \text{for every } i. \tag{146}$$

Conversely, if $x \geq 0$, then

$$\rho = \mathrm{Diag}(x_1, \dots, x_n) \succeq 0. \tag{147}$$

Thus every feasible point of the SDP gives a feasible point of the LP with the same objective value, and every feasible point of the LP gives a diagonal feasible point of the SDP with the same value. In fact, the off-diagonal entries of ρ play no role: replacing ρ by $\mathrm{Diag}(\mathrm{diag}(\rho))$ changes neither the constraints nor the objective.

We have therefore recovered exactly the linear programme Eq. (139). In short,

$$\begin{aligned} x \geq 0 \quad & \longleftrightarrow \quad \mathrm{Diag}(x) \succeq 0, \\ c^{\mathsf{T}}x \quad & \longleftrightarrow \quad \mathrm{tr}\big(\mathrm{Diag}(c)\,\mathrm{Diag}(x)\big), \\ Gx = b \quad & \longleftrightarrow \quad \Lambda(\mathrm{Diag}(x)) = \mathrm{Diag}(b). \end{aligned} \tag{148}$$

Linear programming is therefore the diagonal, or commuting, special case of semidefinite programming. An LP optimises over non-negative vectors; an SDP replaces those vectors by positive-semidefinite matrices.

C Multiple copies and the symmetric subspace

Let $|\psi_i\rangle\langle\psi_i|$, for $i = 1, \ldots, N$, be pure states on $\mathbb{C}^d$, prepared with uniform priors. Suppose that k identical copies of the unknown state are available. The dual state-discrimination problem is then

$$\begin{aligned} \inf_{Y} \quad & \operatorname{tr}(Y) \\ \text{subject to} \quad & Y \succeq \frac{1}{N} |\psi_i\rangle\langle\psi_i|^{\otimes k} \quad \text{for every } i. \end{aligned} \tag{149}$$

The k-fold tensor-product space $(\mathbb{C}^d)^{\otimes k}$ has one tensor factor for each copy. For a permutation $\pi \in S_k$, let U_π denote the unitary that permutes these factors. The symmetric subspace is the subspace of permutation-invariant vectors:

$$\operatorname{Sym}^k(\mathbb{C}^d) := \left\{ |\phi\rangle \in (\mathbb{C}^d)^{\otimes k} : U_\pi |\phi\rangle = |\phi\rangle \text{ for every } \pi \in S_k \right\}.$$

Equivalently,

$$\operatorname{Sym}^k(\mathbb{C}^d) = \operatorname{span}\left\{ |\psi\rangle^{\otimes k} : |\psi\rangle \in \mathbb{C}^d \right\}.$$

In particular, every k-copy vector $|\psi_i\rangle^{\otimes k}$ in the discrimination problem is permutation invariant. Let $\Pi_{\mathrm{sym}}^{(d,k)}$ denote the projector onto the symmetric subspace. Its dimension is

$$d_{\mathrm{sym}}(d,k) := \operatorname{tr}\left(\Pi_{\mathrm{sym}}^{(d,k)}\right) = \binom{d+k-1}{k}. \tag{150}$$

Definition C.1 (Weighted state k-design)**.** We say that the states contain a *weighted state k-design* if there are numbers $q_i \geq 0$, with $\sum_i q_i = 1$, such that

$$\sum_{i=1}^{N} q_i |\psi_i\rangle\langle\psi_i|^{\otimes k} = \frac{\Pi_{\mathrm{sym}}^{(d,k)}}{d_{\mathrm{sym}}(d,k)}. \tag{151}$$

Thus the weighted average of the k-copy states is uniform on the symmetric subspace. Equal weights give an unweighted state k-design. For $k = 1$, this is precisely the tight-frame condition used in Eq. (86).

Theorem C.2 (Multicopy dimension bound)**.** *For N uniformly likely pure states on $\mathbb{C}^d$, given in k copies,*

$$p_{\mathrm{succ}}^{\star,(k)} \leq \min\left\{1, \frac{d_{\mathrm{sym}}(d,k)}{N}\right\}. \tag{152}$$

If the states contain a weighted state k-design, then the $d_{\mathrm{sym}}(d,k)/N$ bound is attained.

Proof. Every vector $|\psi_i\rangle^{\otimes k}$ lies in the symmetric subspace. Consequently,

$$|\psi_i\rangle\langle\psi_i|^{\otimes k} \preceq \Pi_{\mathrm{sym}}^{(d,k)} \quad \text{for every } i. \tag{153}$$

Therefore the scaled symmetric projector

$$Y_{\mathrm{sym}} := \frac{1}{N} \Pi_{\mathrm{sym}}^{(d,k)} \tag{154}$$

is feasible for Eq. (149). Weak duality gives

$$p_{\mathrm{succ}}^{\star,(k)} \leq \operatorname{tr}(Y_{\mathrm{sym}}) = \frac{d_{\mathrm{sym}}(d,k)}{N}. \tag{155}$$

Combining this with the trivial upper bound of one proves Eq. (152).

Now suppose that Eq. (151) holds. Define

$$M_i := d_{\mathrm{sym}}(d,k) q_i |\psi_i\rangle\langle\psi_i|^{\otimes k} + \frac{1}{N}\left(\mathbb{1}_{d^k} - \Pi_{\mathrm{sym}}^{(d,k)}\right). \tag{156}$$

Every M_i is PSD, and the design identity gives

$$\sum_{i=1}^{N} M_i = \Pi_{\text{sym}}^{(d,k)} + \mathbb{1}_{d^k} - \Pi_{\text{sym}}^{(d,k)} = \mathbb{1}_{d^k}. \tag{157}$$

Hence $\{M_i\}_i$ is a POVM. Since each $|\psi_i\rangle\langle\psi_i|^{\otimes k}$ is supported on the symmetric subspace, this POVM attains

$$\frac{1}{N}\sum_{i=1}^{N} \operatorname{tr}\Big(|\psi_i\rangle\langle\psi_i|^{\otimes k}\, M_i\Big) = \frac{d_{\text{sym}}(d,k)}{N}\sum_{i=1}^{N} q_i = \frac{d_{\text{sym}}(d,k)}{N}. \tag{158}$$

This matches the dual upper bound. □

Purity is essential here. A tensor power of a mixed state need not be supported on the symmetric subspace, so the same projector need not be dual feasible. For a more complete study of the pure-state multicopy problem and its relation to state designs, see [3].

References

[1] G. B. Dantzig, Econometrica **17**, 73 (1949).
[2] P. Skrzypczyk and D. Cavalcanti, *Semidefinite Programming in Quantum Information Science* (IOP Publishing, 2023) arXiv:2306.11637 .
[3] M. Kvashchuk, P. Chernyshova, L. E. A. Porto, T.-A. Ohst, L. B. Vieira, and M. T. Quintino, The most discriminable quantum states in the multicopy regime (2026), arXiv:2604.26927 .
[4] I. Dunning, J. Huchette, and M. Lubin, SIAM Review **59**, 295 (2017), arXiv:1508.01982 .
[5] J. Bavaresco, M. Murao, and M. T. Quintino, Phys. Rev. Lett. **127**, 200504 (2021), arXiv:2011.08300 .
[6] A. C. Doherty, P. A. Parrilo, and F. M. Spedalieri, Phys. Rev. A **69**, 022308 (2004), arXiv:quant-ph/0308032 .
[7] D. Cavalcanti and P. Skrzypczyk, Rep. Prog. Phys. **80**, 024001 (2017), arXiv:1604.00501 .
[8] M. M. Wolf, D. Pérez-García, and C. Fernández, Phys. Rev. Lett. **103**, 230402 (2009), arXiv:0905.2998 .
[9] M. Navascués, S. Pironio, and A. Acín, New J. Phys. **10**, 073013 (2008), arXiv:0803.4290 .
[10] E. Wolfe, A. Pozas-Kerstjens, M. Grinberg, D. Rosset, A. Acín, and M. Navascués, Phys. Rev. X **11**, 021043 (2021), arXiv:1909.10519 .
[11] W. Hall, Phys. Rev. A **75**, 032102 (2007), arXiv:quant-ph/0610031 .
[12] X. Wang and R. Duan, in *2016 IEEE International Symposium on Information Theory (ISIT)* (2016) pp. 1690–1694, arXiv:1601.06888 .
[13] G. Chiribella, New J. Phys. **14**, 125008 (2012), arXiv:1207.6172 .
[14] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Phys. Rev. Lett. **101**, 060401 (2008), arXiv:0712.1325 .
[15] M. Araújo, C. Branciard, F. Costa, A. Feix, C. Giarmatzi, and Č. Brukner, New J. Phys. **17**, 102001 (2015), arXiv:1506.03776 .
[16] S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, 2004).