Lecture 15: Semidefinite programming (§1.2.3)

Semidefinite programming lets you optimize a linear function subject to semidefinite constraints (as well as linear equalities and inequalities). Notation shorthand: $C \geq D \iff C-D \geq 0$.

Let $\Phi \in T(\mathcal{X},\mathcal{Y})$ be a linear Hermitian-preserving map, and let $A \in \mathrm{Herm}(\mathcal{X})$ and $B \in \mathrm{Herm}(\mathcal{Y})$.

Recall that $\langle M, N\rangle = \mathrm{Tr}(M^* N)$, for any $M, N \in L(\mathcal{X})$, and the <u>adjoint superoperator</u> $\Phi^* \in T(\mathcal{Y},\mathcal{X})$ is defined by the identity

$$\langle Y, \Phi(X)\rangle = \langle \Phi^*(Y), X\rangle$$

which must hold for all $X \in L(\mathcal{X})$ and $Y \in L(\mathcal{Y})$.

<u>Def (p. 54)</u>: A <u>semidefinite program</u> (SDP) is a triple (Φ, A, B) to which we associate a pair of optimization problems:

<u>Primal problem</u>

maximize: $\langle A, X\rangle$

subject to: $\Phi(X) = B$
$X \in \mathrm{Pos}(\mathcal{X})$

<u>Dual problem</u>

minimize: $\langle B, Y\rangle$

subject to: $\Phi^*(Y) \geq A$
$Y \in \mathrm{Herm}(\mathcal{Y})$

Some terminology:

- primal / dual feasible sets: $\mathcal{A} = \{X \in \mathrm{Pos}(\mathcal{X}) : \Phi(X) = B\}$
  $\mathcal{B} = \{Y \in \mathrm{Herm}(\mathcal{Y}) : \Phi^*(Y) \geq A\}$
- primal / dual objective functions: $X \mapsto \langle A, X\rangle$
  $Y \mapsto \langle B, Y\rangle$
- primal / dual optimum values: $\alpha = \sup\{\langle A, X\rangle : X \in \mathcal{A}\}$
  $\beta = \inf\{\langle B, Y\rangle : Y \in \mathcal{B}\}$

<u>Example 1.20</u>: Let $\Phi \in T(\mathcal{X} \oplus \mathcal{Y})$ be the following Hermitian-preserving map:

$$\Phi\begin{pmatrix} X & \cdot \\ \cdot & Y \end{pmatrix} = \begin{pmatrix} X & 0 \\ 0 & Y \end{pmatrix} \qquad \begin{matrix} \forall X \in L(\mathcal{X}) \\ \forall Y \in L(\mathcal{Y}) \end{matrix}$$

and let $A, B \in \text{Herm}(\mathcal{X} \oplus \mathcal{Y})$ be defined as follows:

$$A = \frac{1}{2}\begin{pmatrix} 0 & K^* \\ K & 0 \end{pmatrix} \qquad B = \begin{pmatrix} I_X & 0 \\ 0 & I_Y \end{pmatrix}$$

for some $K \in L(\mathcal{X}, \mathcal{Y})$.

After some simplifications, the primal and dual problems can be expressed as follows:

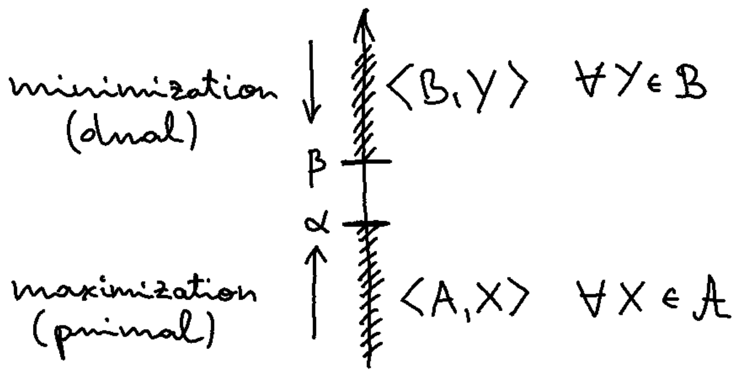
<u>Primal</u>	<u>Dual</u>
maximize: $\text{Re}(\langle K, Z \rangle)$	minimize: $\frac{1}{2}\text{Tr}(X) + \frac{1}{2}\text{Tr}(Y)$
subject to: $\begin{pmatrix} I_X & Z^* \\ Z & I_Y \end{pmatrix} \geq 0$	subject to: $\begin{pmatrix} X & -K^* \\ -K & Y \end{pmatrix} \geq 0$
$Z \in L(\mathcal{X}, \mathcal{Y})$	$X \in \text{Pos}(\mathcal{X})$, $Y \in \text{Pos}(\mathcal{Y})$

<u>Useful observation</u>: if $M = \begin{pmatrix} X & \cdot \\ \cdot & Y \end{pmatrix} \geq 0$ then $X \geq 0$ and $Y \geq 0$ because $\langle \psi | M | \psi \rangle \geq 0$ for all $|\psi\rangle$, including ones that have support (i.e., non-zero entries) only in the chosen block.

<u>Homework</u>: the optimum value of the above SDP is $\|K\|_1$.

SDP duality

Weak duality: it is always the case that $\alpha \leq \beta$.



Hence every dual feasible operator $Y \in \mathcal{B}$ provides an upper bound on α:

$$\alpha \leq \langle B, Y \rangle, \quad \forall Y \in \mathcal{B}.$$

Similarly, every primal feasible $X \in \mathcal{A}$ provides a lower bound on β:

$$\langle A, X \rangle \leq \beta, \quad \forall X \in \mathcal{A}$$

In practice, it is often the case that the primal and dual optimum values agree, i.e.,

$$\alpha = \beta.$$

This is called strong duality. Conditions for when this happens are known (see Theorem 1.18).

Discriminating quantum states of an ensemble (§3.1.2)

Recall that an *ensemble* of quantum states is a map $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X})$ such that

$$\sum_{a\in\Sigma} \mathrm{Tr}(\eta(a)) = 1.$$

You can think of this as having state $\rho_a = \frac{\eta(a)}{\mathrm{Tr}(\eta(a))}$ with probability $p_a = \mathrm{Tr}(\eta(a))$. We can represent η by a classical-quantum state

$$\sum_{a\in\Sigma} |a\rangle\langle a|_Y \otimes \eta(a)_X$$

where Alice has the classical register Y and Bob has the quantum register X. Bob's goal is to guess the state of Y by measuring X. If his measurement is $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$ then the probability of success is

$$\sum_{a\in\Sigma} \langle \mu(a), \eta(a)\rangle.$$

Let us denote the optimal success probability (maximized over all possible measurements) by

$$\mathrm{opt}(\eta) = \max_{\mu} \sum_{a\in\Sigma} \langle \mu(a), \eta(a)\rangle.$$

Can we compute this for a given ensemble η?

Theorem 3.4 (Helstrom–Holevo): If $|\Sigma| = 2$ $(\Sigma = \{0,1\})$ then

$$\mathrm{opt}(\eta) = \frac{1}{2} + \frac{1}{2}\|\eta(0) - \eta(1)\|_1.$$

Note that $\|\eta(0) - \eta(1)\|_1 \in [0,1]$. The extremes correspond to identical and orthogonal states, respectively.

What about $|\Sigma| \geq 3$? There is no simple formula known in this case. However, finding the optimal measurement μ can be stated as an SDP:

Primal

$$\text{maximize:} \quad \sum_{a\in\Sigma} \langle \mu(a), \eta(a) \rangle$$

$$\text{subject to:} \quad \mu(a) \in \mathrm{Pos}(\mathcal{X}), \quad \forall a \in \Sigma$$

$$\sum_{a\in\Sigma} \mu(a) = I_X.$$

The optimum value α of this SDP gives the best success probability for discriminating states from the ensemble η. The optimal values of the variables $\mu(a)$ (or X in the standard form) specify the optimal measurement.

Exercise: formulate this SDP in the standard form and find its dual.

Is there any simpler way to find a good measurement that may not be optimal but is sufficiently good? I.e., can we trade optimality for simplicity?

The pretty good measurement (PGM)

Let $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X})$ be an ensemble. Let $\rho = \sum_{a\in\Sigma} \eta(a)$ be the averaged state (assume ρ is full rank for simplicity) and set

$$\mu(a) = \rho^{-\frac{1}{2}} \eta(a) \rho^{-\frac{1}{2}}.$$

This is called the pretty good measurement for ensemble η. Let us denote its success probability by

$$\mathrm{pgm}(\eta) = \sum_{a\in\Sigma} \langle \mu(a), \eta(a) \rangle.$$

Theorem 3.10 (Barnum-Knill): For any ensemble η,

$$\mathrm{opt}(\eta) \geq \mathrm{pgm}(\eta) \geq \mathrm{opt}(\eta)^2.$$

where $\mathrm{opt}(\eta)$ is the success probability of the optimal measurement.

Proof: The first inequality is by definition of $\mathrm{opt}(\eta)$. Let $\nu : \Sigma \to \mathrm{Pos}(\mathcal{X})$ be any measurement. Then

$$\langle \nu(a), \eta(a) \rangle = \langle \rho^{\frac{1}{4}} \nu(a) \rho^{\frac{1}{4}}, \rho^{-\frac{1}{4}} \eta(a) \rho^{-\frac{1}{4}} \rangle$$
$$\leq \| \rho^{\frac{1}{4}} \nu(a) \rho^{\frac{1}{4}} \|_2 \cdot \| \rho^{-\frac{1}{4}} \eta(a) \rho^{-\frac{1}{4}} \|_2$$

by cyclicity of Tr and Cauchy-Schwarz inequality for operators. Applying the same inequality for real vectors, i.e., $\vec{a}\cdot\vec{b} \leq \|\vec{a}\|_2 \cdot \|\vec{b}\|_2$, we get

$$\sum_{a\in\Sigma} \langle \nu(a), \eta(a) \rangle \leq \sqrt{\sum_{a\in\Sigma} \| \rho^{\frac{1}{4}} \nu(a) \rho^{\frac{1}{4}} \|_2^2} \cdot \sqrt{\sum_{a\in\Sigma} \| \rho^{-\frac{1}{4}} \eta(a) \rho^{-\frac{1}{4}} \|_2^2}.$$

Let us upper bound the two terms separately. Note that

$$\| \rho^{\frac{1}{4}} \nu(a) \rho^{\frac{1}{4}} \|_2^2 = \langle \rho^{\frac{1}{4}} \nu(a) \rho^{\frac{1}{4}}, \rho^{\frac{1}{4}} \nu(a) \rho^{\frac{1}{4}} \rangle$$
$$= \langle \nu(a), \sqrt{\rho}\, \nu(a) \sqrt{\rho} \rangle \leq \mathrm{Tr}\left[\sqrt{\rho}\, \nu(a) \sqrt{\rho}\right]$$

where we used the cyclic property of trace and the inequalities $\nu(a) \leq I$ (since $\sum_{a\in\Sigma} \nu(a) = I$) and $\sqrt{\rho}\, \nu(a) \sqrt{\rho} \geq 0$. Hence

$$\sum_{a\in\Sigma} \| \rho^{\frac{1}{4}} \nu(a) \rho^{\frac{1}{4}} \|_2^2 \leq \sum_{a\in\Sigma} \mathrm{Tr}(\sqrt{\rho}\, \nu(a) \sqrt{\rho}) = \mathrm{Tr}\, \rho = 1.$$

The second term can be upper bounded in a similar way:

$$\|\rho^{-\frac{1}{4}}\eta(a)\rho^{-\frac{1}{4}}\|_2^2 = \langle \underbrace{\rho^{-\frac{1}{2}}\eta(a)\rho^{-\frac{1}{2}}}_{\mu(a)\text{ — the PGM operator}}, \eta(a)\rangle = \langle \mu(a), \eta(a)\rangle,$$

so

$$\sum_{a\in\Sigma} \|\rho^{-\frac{1}{4}}\eta(a)\rho^{-\frac{1}{4}}\|_2^2 = \sum_{a\in\Sigma} \langle \mu(a), \eta(a)\rangle = \mathrm{pgm}(\eta).$$

Putting both bounds together:

$$\sum_{a\in\Sigma} \langle \nu(a), \eta(a)\rangle \leqslant \sqrt{1}\cdot\sqrt{\mathrm{pgm}(\eta)}.$$

Since this bound holds for any measurement ν, it also holds for the optimal measurement. If we maximize the left-hand side over all measurements, we get

$$\mathrm{opt}(\eta) \leqslant \sqrt{\mathrm{pgm}(\eta)}$$

which is the desired inequality. $\square$