Semidefinite programming / semidefinite programs (SDPs)

- Many tasks in quantum info theory require optimization
- Many properties in quantum info theory can be phrased as optimizations

Example: • Given a quantum channel Λ and a target state σ, find the input state ρ that maximizes the overlap $\mathrm{tr}(\Lambda[\rho]\sigma)$, i.e.

$$\boxed{\begin{aligned} &\max \quad \mathrm{tr}(\Lambda[\rho]\sigma) \\ &\text{subject to} \quad \mathrm{tr}[\rho]=1 \\ &\qquad\qquad\qquad \rho \geq 0 . \end{aligned}}$$

- The trace norm $\|Z\|_1$ of a Hermitian matrix Z can be phrased as

$$\|Z\|_1 = \boxed{\begin{aligned} &\max \; 2\,\mathrm{tr}(EZ) - \mathrm{tr}(Z) \\ &\text{s.t.} \qquad E \leq \mathbb{1} \\ &\qquad\qquad E \geq 0 \end{aligned}}$$

- Trace distance estimation: measured probabilities $m_x = \mathrm{tr}(\rho M_x)$ on a prepared state ρ. How close is ρ to some target state σ at best.

Task: minimize $\frac{1}{2} \| \rho - \sigma \|_1$
subject to $\mathrm{tr}(M_x \rho) = m_x \quad \forall x = 1, \dots, n$
$\mathrm{tr}(\rho) = 1$
$\rho \geq 0$

$\Rightarrow$ SDPs are particularly important optimization problems

All involved objects are Hermitian matrices
All constraints are affine or positive semidefinite

7.1 What is an SDP?

Disclaimer: From now on all matrices will be assumed to be Hermitian

Def. A semidefinite program is a triple (Φ, A, B) consisting of a linear Hermiticity preserving map $\Phi : L(\mathcal{H}_1) \to L(\mathcal{H}_2)$ and two Hermitian matrices $A \in L(\mathcal{H}_1)$ and $B \in L(\mathcal{H}_2)$.

The corresponding primal and dual problems are

Primal	Dual
objective function	
maximize $\mathrm{tr}(XA)$	minimize $\mathrm{tr}(BY)$

subject to $\quad \Phi(X) = B \quad\quad$ subject to $\quad \Phi^{\dagger}[Y] \geq A$

$X \geq 0$

affine constraint

positivity constraint

$\cdot^{\dagger}$: adjoint / dual map.

Nomenclature: A matrix X that satisfies the constraints of the primal, i.e. $\Phi(X) = B$ and $X \geq 0$ is called primal feasible, while a matrix Y that satisfies $\Phi^{\dagger}[Y] \geq A$ is called dual feasible.

⇒ Let $\mathcal{A}$ and $\mathcal{B}$ be the set of all primal and dual feasible X and Y.

Example: Given A find the largest possible overlap with a quantum state X.

Primal

$B \equiv \mathbb{1} \quad\quad \Phi[\cdot] = \mathrm{tr}[\cdot]$

max. $\quad \mathrm{tr}(XA)$

s.t. $\quad \left.\begin{array}{l} \mathrm{tr}(X) = 1 \\ X \geq 0 \end{array}\right\}$ X is state

↳ Easy to see: computes the largest eigenvalue of A

for the dual, we require $\Phi^{\dagger}$ of $\Phi[\cdot] = \mathrm{tr}[\cdot]$. It is defined through $\mathrm{tr}(\Phi[X]Y) = \mathrm{tr}(X\,\Phi^{\dagger}[Y]) \quad \forall X, Y$

$\Rightarrow \Phi^{\dagger}[Y] = Y\mathbb{1} \quad\quad (Y \in \mathbb{C})$

2) Dual

minimize y

s.t. $y \mathbb{1} \geq A$

⇒ Easy to see: Also computes the largest eigenvalue of A.

⇒ For this example, primal and dual solution coincide.

7.2 Weak and strong duality

What the relevance / meaning of the dual problem?

Thm.: Let $\alpha = \sup_{X \in \mathcal{A}} \operatorname{tr}(XA)$ and $\beta = \inf_{Y \in \mathcal{B}} \operatorname{tr}(YB)$

be the optimal primal and dual values, respectively.

Then we have $\boxed{\alpha \leq \beta}$, i.e. "the dual upper bounds the primal"

Proof: For arbitrary feasible $X \in \mathcal{A}$ and $Y \in \mathcal{B}$, we have

$$\underbrace{\operatorname{tr}(YB)}_{\text{objective of the dual}} = \operatorname{tr}(Y\Phi[X]) = \operatorname{tr}(\Phi^{\dagger}[Y]X) \geq \underbrace{\operatorname{tr}(AX)}_{\text{objective fct. of the primal}}$$

$$\Rightarrow \inf_{Y \in \mathcal{B}} \operatorname{tr}(YB) = \beta \geq \alpha = \sup_{X \in \mathcal{A}} \operatorname{tr}(AX)$$

$\boxtimes$

When do primal and dual solutions actually coincide?

Thm: (Strong duality, Slater condition):
Let $\mathcal{A}$ and $\mathcal{B}$ be non-empty. Then the optimal values for the dual and primal problem coincide, i.e. $\alpha = \beta$, if there exists either

- a Hermitian matrix $Y \in \mathcal{B}$ such that $\Phi^*[Y] > A$ (strict dual feasibility)
- or a matrix $X \in \mathcal{A}$ with $X > 0$, (strict primal feasibility)

Proof: See, for example, John Watrous' lecture notes. $\boxtimes$

7.3 Numerical solutions of SDPs

Recall:

$$\begin{aligned} &\max && \operatorname{tr}(XA) && \rightsquigarrow \text{linear objective fct.} \\ &\text{s.t.} && \Phi[X] = B && \rightsquigarrow \text{affine constraint} \\ & && X \geq 0 && \rightsquigarrow \text{positivity constraint} \end{aligned}$$

$\Rightarrow$ SDPs constitute the optimization of objective over a convex set.

$\hookrightarrow$ Efficient, faithful and precise numerical solutions via "interior point methods"

Practically: $\exists$ open source numerical solvers like SeDuMi and SDP3, as well the freely available solver MOSEK for SDPs

$\Rightarrow$ Require input of the SDP in "standard form"

Rule of thumb: Every optimization of a linear function under affine and positivity constraints is an SDP (see Watrous' lecture notes).

Luckily, there exist numerical packages that "communicate" with the solvers (e.g., cvx and YALMIP for matlab, JuMP for Julia, PICOS and cvxpy for Python.)

Example: Computation of the trace norm.

7.4. Dual problem via the Lagrangian

Let

$$\boxed{\begin{array}{ll} \max & \mathrm{tr}(XA) \\ \text{s.t.} & \Phi(X) = B \\ & X \geq 0 \end{array}}$$

be a primal SDP with opt. α.

$\Rightarrow$ Lagrangian multiplier Y

$\Rightarrow$ Lagrangian multiplier $S \geq 0$

To find the dual, we introduce a Lagrangian multiplier for each constraint and set

$$L(X,Y,S) = \mathrm{tr}(XA) + \mathrm{tr}\left(Y(B-\Phi[X])\right) + \mathrm{tr}(SX)$$

$= \mathrm{tr}(X(A - \Phi^{\dagger}[Y] + S)) + \mathrm{tr}(YB)$

For $S \succeq 0$, we see that $\sup_X L(X, Y, S) =: g(Y,S) \geq \alpha$

The <u>best</u> upper bound is achieved through $\min_{Y,S} g(Y,S)$.

However, if $A - \Phi^{\dagger}[Y] + S \neq 0$, then $g(Y,S) = \infty$

Consequently, we arrive at the constrained optimization problem:

$$\boxed{\begin{aligned} &\text{minimize} && \mathrm{tr}(YB) \\ &\text{s.t.} && A - \Phi^{\dagger}[Y] + S = 0 \\ &&& S \succeq 0 \end{aligned}}$$

This is equivalent to

$$\boxed{\begin{aligned} &\min && \mathrm{tr}(YB) \\ &\text{s.t.} && \Phi^{\dagger}[Y] \succeq A \end{aligned}}$$

↳ This is exactly the dual problem.

<u>NB</u>: This approach can also be used to go from a minimization to a maximization problem.

<u>Why is Lagrangian approach useful?</u>

It does not require the SDP to be in standard form.

Ex.: Computation of the trace norm:

$$\|Z\|_1 = \boxed{\begin{array}{ll} \max & 2\,\mathrm{tr}(EZ) - \mathrm{tr}(Z) \\ \text{s.t.} & E \succeq 0 \\ & E \preceq \mathbb{1} \end{array}}$$

$\geq$, $R \succeq 0$

$\geq$, $S \succeq 0$

$$L(E,R,S) = 2\,\mathrm{tr}(EZ) - \mathrm{tr}(Z) + \mathrm{tr}(RE) + \mathrm{tr}(S(\mathbb{1}-E))$$

$$= \mathrm{tr}(E(2Z + R - S)) + \mathrm{tr}(S - Z)$$

Dual SDP:

$$\begin{array}{ll} \min & \mathrm{tr}(S-Z) \\ \text{s.t.} & 2Z + R - S = 0 \\ & R, S \succeq 0 \end{array}$$

Setting $\tilde{S} := S - Z$, this can be further simplified

to

$$\underbrace{\begin{array}{ll} \min & \mathrm{tr}(\tilde{S}) \\ \text{s.t.} & \tilde{S} - Z \succeq 0 \\ & Z + \tilde{S} \succeq 0 \end{array}}$$

$$\Downarrow$$

$$\|Z\|_1 = \boxed{\begin{array}{ll} \min & \mathrm{tr}(\tilde{S}) \\ \text{s.t.} & -\tilde{S} \preceq Z \preceq \tilde{S} \end{array}}$$

Example: State estimation problem: What is the closest state to a target σ that fits with an observed measurement record

$$\{ m_x := \mathrm{tr}(M_x \rho) \} \qquad x = 1, \ldots, N$$

$$\Rightarrow \quad \begin{array}{ll} \text{minimize} & \frac{1}{2} \|\rho - \sigma\|_1 \\ \text{s.t.} & \mathrm{tr}(M_x \rho) = m_x \quad \forall\, x = 1, \ldots, N \qquad (*) \\ & \mathrm{tr}(\rho) = 1 \\ & \rho \geq 0 \end{array}$$

A priori, $\frac{1}{2} \|\rho - \sigma\|_1$ is not linear in ρ $\Rightarrow$ not an SDP?

But: we know that $\|\rho - \sigma\|_1 = \min \mathrm{tr}(\tilde{\rho})$ s.t. $-\tilde{\rho} \leq \rho - \sigma \leq \tilde{\rho}$

Inserting this into $(*)$

$$\begin{array}{ll} \text{minimize} & \frac{1}{2} \mathrm{tr}(\tilde{\rho}) \\ \text{s.t.} & \mathrm{tr}(M_x \rho) = m_x \quad \forall\, x = 1, \ldots, N \\ & \mathrm{tr}(\rho) = 1 \\ & \rho \geq 0 \\ & -\tilde{\rho} \leq \rho - \sigma \\ & \rho - \sigma \leq \tilde{\rho} \end{array}$$

This problem is indeed an SDP.

(NB: Identification of what problems are actually SDPs is ongoing research, but there are many ingenious "tricks" to obtain SDP formulations - see Watrous' lecture notes)

Waterloo, Quantum Info lecture.

See below for additional example

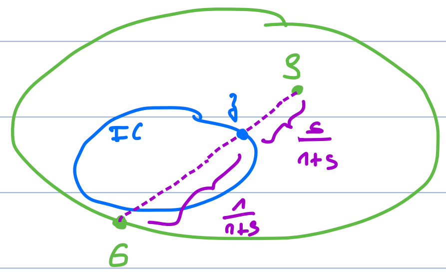
2. Robustness of coherence (how coherent is a quantum state)

Idea: measure coherence by robustness to noise

Let IC be the set of all incoherent states, i.e. the set of all states ξ that satisfy $\Delta[\xi] = \xi$, where Δ is the completely dephasing map. The robustness of coherence is defined as:

$$R_c[S] = \min_{s,G} \left\{ s \geq 0 \;\middle|\; \frac{S + sG}{1+s} = \xi \in IC \right\}$$

graphically:



⇝ Random is the resource theory of coherence.

Can this be computed via SDP?

$$\begin{aligned} R_c[S] = \min \quad & s \\ \text{s.t.} \quad & S + sG = (1+s)\xi \\ & S \geq 0, \; G \geq 0, \; \xi \geq 0 \\ & \mathrm{tr}(G) = \mathrm{tr}(\xi) = 1 \\ & \Delta[\xi] - \xi = 0 \end{aligned}$$

⇝ non-linear constraint.

<u>But</u>: We can set $\tilde{\xi} := (1+s)\xi$ to re-write this optimization:

$$\left.\begin{aligned} R_c[S] = \min \quad & \mathrm{tr}(\tilde{\xi}) - 1 \\ \text{s.t.} \quad & S \leq \tilde{\xi} \\ & \tilde{\xi} \geq 0 \\ & \Delta[\tilde{\xi}] - \tilde{\xi} = 0 \end{aligned}\right\}$$

This is indeed an SDP.

What can we learn from its dual?

$$L(\tilde{\varsigma}, R, S, T) = \mathrm{tr}(\tilde{\varsigma}) - 1 - \mathrm{tr}(R(\tilde{\varsigma} - \varsigma)) - \mathrm{tr}(S\tilde{\varsigma}) -$$

$$+ \mathrm{tr}[T(\underbrace{\Delta}_{\text{self-dual map}}[\tilde{\varsigma}] - \tilde{\varsigma})]$$

$$= \mathrm{tr}(\tilde{\varsigma}(\mathbb{1} - R - S - \Delta[T] + T)) - 1 + \mathrm{tr}(R\varsigma)$$

$$\Rightarrow R_c[\varsigma] = \left.\begin{array}{ll} \max & \mathrm{tr}(R\varsigma) - 1 \\ \text{s.t.} & \mathbb{1} - R - S + \Delta[T] - T = 0 \\ & R, S \geq 0 \end{array}\right\} \begin{array}{l}\text{dual}\\ \text{SDP}\end{array}$$

$\Rightarrow$ This can be further massaged:

$$R_c[\varsigma] = \begin{array}{ll} \max & \mathrm{tr}(R\varsigma) - 1 \\ \text{s.t.} & R \leq \underbrace{\mathbb{1} + T - \Delta[T]}_{\substack{\text{arbitrary Hermitian matrix } \tilde{T} \\ \text{with ones on the diagonal}}} \\ & R \geq 0 \end{array}$$

Finally:

$$R_c[\varsigma] = \begin{array}{ll} \max & \mathrm{tr}(R\varsigma) - 1 \\ \text{s.t.} & 0 \leq R \leq \tilde{T} \\ & \mathrm{diag}[\tilde{T}] = \mathbb{1} \end{array}$$

We can even remove the inequality: If R_* is feasible, then

$\mathrm{diag}[R_*] \leq \mathbb{1}$. $\Rightarrow$ Completing it to $\tilde{R}_*$ that satisfies $\mathrm{diag}[\tilde{R}_*] = \mathbb{1}$ can only improve the value

$$\Rightarrow \quad R_c[\rho] = \max \ \mathrm{tr}(\tilde{R}\rho) - 1$$
$$\text{s.t.} \quad \tilde{R} \geq 0$$
$$\mathrm{diag}[\tilde{R}] = \mathbb{1}$$

Setting $W := \tilde{R} - \mathbb{1}$, this allows for a nice interpretation in terms of witnesses of coherence:

$$\boxed{\begin{aligned} R_c[\rho] = \ &\max \ \mathrm{tr}(W\rho) \\ &\text{s.t.} \quad W + \mathbb{1} \geq 0 \\ &\qquad\ \ \mathrm{diag}[W] = 0 \end{aligned}}$$

Any such W will yield $\mathrm{tr}(W\rho) = 0$ for incoherent states, but positive values for any coherent state ρ ($\Rightarrow$ witness of coheren.)