Quantum Optimal Transport: Regularization and Algorithms

Pavlo Pelikh PPELIKH@UOTTAWA.CA

University of Ottawa

Augusto Gerolin AGEROLIN@UOTTAWA.CA, GEROLIN@IMPA.BR
University of Ottawa
Instituto Nacional de Matemática Pura e Aplicada

Abstract

Quantum Optimal Transport (QOT) extends optimal transport to quantum data such as states and channels. In this paper, we develop and benchmark computational algorithms for QOT, focusing on the quantum analog of the Sinkhorn algorithm [12]. Applications include the QOT between quantum channels [14] and spin systems, where numerical tests show accurate and efficient performance. Our work bridges quantum information, convex optimization, and statistical physics, providing practical tools for quantum machine learning and machine learning for quantum data.

1. Introduction

Let $N \in \mathbb{N}$ be a natural number and denote by $[N] := \{1, \dots, N\}$ an index set with N elements. Consider N complex finite-dimensional Hilbert spaces $\mathbb{C}^d$ of dimension $d \in \mathbb{N}$ and let $D = d^N$ be the dimension of the composite system. For all $i \in [N]$, let $\gamma_i \in \mathcal{D}_d = \left\{A \in \mathbb{C}^{d\times d} \,\middle|\, A = A^\dagger, A \succcurlyeq 0, \operatorname{Tr}[A] = 1\right\}$ be a density matrix on the i-th Hilbert space $\mathbb{C}^d$. Let $\mathbf{H} \in \mathbb{H}_D = \left\{A \in \mathbb{C}^{D\times D} \,\middle|\, A = A^\dagger\right\}$ be the Hermitian (cost) operator of dimension D.

The main goal of this work is to develop novel optimization methods and computational algorithm for the following Quantum Optimal Transport (QOT) problem [11]

$$\mathfrak{P}^Q_{\mathbf{H}}(\gamma_1, \dots, \gamma_N) = \min_{\Gamma}\{\operatorname{Tr}[\mathbf{H}\Gamma] \,:\, \Gamma \mapsto (\gamma_1, \dots, \gamma_N)\}, \tag{1}$$

where the symbol $\Gamma \mapsto (\gamma_1, \dots, \gamma_N)$ denotes the set of $N-$particle density matrix $\Gamma \in \mathcal{D}_D$ in $\mathbb{C}^D$ having the $i-$th partial traces $\operatorname{Tr}^i[\Gamma] = \gamma_i$, $\forall\, i \in [N]$, i.e. for any $B \in \mathbb{H}_d$ and any $i \in [N]$, i.e.

$$\operatorname{Tr}\left[\Gamma\left(\bigotimes_{j=1}^{i-1} \mathbb{1}_j \otimes B \otimes \bigotimes_{j=i+1}^{N} \mathbb{1}_j\right)\right] = \operatorname{Tr}[B\gamma_i]\,. \tag{2}$$

We say that Γ is a quantum coupling of $\gamma_1, \dots, \gamma_N$, since they can be interpreted as the quantum analog of transport plans or couplings in optimal transport theory. The partial trace conditions in (2) are the quantum analog of marginals in probability theory. The QOT problem (1) admits a (non-unique) minimizer and can be interpreted as a semidefinite program (SDP) [11, Theorem 3.1].

Originating in non-commutative probability and quantum Markov semigroups [8, 13], QOT has been recently applied to quantum information [1, 10, 14, 15, 18], statistical mechanics [5, 6], electronic structure theory [17], and quantum machine learning [9, 20, 21]. Several formulations can be viewed as quantum analogs of Wasserstein distances in optimal transport for probability measures.

Main Contributions. We introduce and benchmark two mathematically rigorous computational algorithms for Quantum Optimal Transport (QOT): (i) *Subgradient ascent for eigenvalue-regularized QOT* (Section 2.1). This method is particularly effective when the primal solution is expected to be rank-1, see Proposition 1 for details; (ii) *L-BFGS for von Neumann entropy regularized QOT* (Section 2.2), which is the quantum analogue of the celebrated Shannon-entropy regularized OT [12]. In Section 3, we present extensive numerical experiments on quantum channels [14], spin systems, and random Hamiltonians, systematically comparing the two approaches.

2. Computational algorithms and Regularized Quantum Optimal Transport

Our algorithms rely on the dual formulation of (1), see Theorem 3.2 in [11].

$$\mathfrak{D}^{Q}_{\mathbf{H}}(\gamma_1,\dots,\gamma_N) = \sup_{U_1,\dots,U_N} \left\{ \sum_{i=1}^{N} \mathrm{Tr}\left[U_i\gamma_i\right] \,:\, U_i \in \mathbb{H}_d \text{ and } \mathbf{H}_{\text{eff}} := \mathbf{H} - \bigoplus_{i=1}^{N} U_i \succcurlyeq 0 \right\}, \tag{3}$$

where $\mathbf{H}_{\text{eff}}$ is the *effective Hamiltonian*.

The dual variables $\{U_i\}_{i=1}^N$ are referred as Kantorovich operators. The symbol $\bigoplus_{i=1}^N U_i$ denotes the Kronecker sum, i.e. $\bigoplus_{i=1}^N U_i = \sum_{i=1}^N \bigotimes_{j=1}^{i-1} \mathbb{1}_j \otimes U_i \otimes \bigotimes_{j=i+1}^N \mathbb{1}_j$. See the Theorem 3.2 in [11] for the proof of strong duality and complementary slackness.

In the following, we briefly introduce the two computational algorithms. For details, we refer to the supplementary material.

2.1. Eigenvalue Penalization.

Our first method penalizes the lowest eigenvalue of $\mathbf{H}_{\text{eff}}$ in (3) to enforce rank-1 primal–dual optimality – see the Prop. 1 in the supplementary material S1 for a rigorous justification

$$D^{reg}_{\mathbf{H}} : \bigtimes_{i=1}^{N} \mathbb{H}_d \to \mathbb{R}, \quad D^{reg}_{\mathbf{H}}(U_1,\dots,U_N) = \sum_{i=1}^{N} \mathrm{Tr}\left[U_i\gamma_i\right] + \lambda_1^{\text{eff}}(U_1,\dots,U_N), \tag{4}$$

where $\lambda_1^{\text{eff}}(U_1,\dots,U_N) = \min\left\{\langle\psi, \mathbf{H}^{eff}\psi\rangle \,:\, \psi\in\mathbb{C}^D \text{ and } \|\psi\|=1\right\}$ is the smallest eigenvalue of $\mathbf{H}_{\text{eff}}$. Then, the problem reduces to solve the following unconstrained maximization problem

$$\sup_{U_i\in\mathbb{H}_d} D^{reg}_{\mathbf{H}}(U_1,\dots,U_N) = \sup_{U_i\in\mathbb{H}_d} \inf_{\substack{\psi\in\mathcal{H}:\\ \|\psi\|=1}} \mathcal{L}(U_1,\dots,U_N,\psi), \tag{5}$$

where $\mathcal{L} : \times_{i=1}^N \mathbb{H}_n \times \mathbb{C}^D \to \mathbb{R}$ is $\mathcal{L}(U_1,\dots,U_N,\psi) = \sum_{i=1}^N \mathrm{Tr}\left[U_i\gamma_i\right] + \langle\psi, (\mathbf{H} - \bigoplus_{i=1}^N U_i)\psi\rangle$.

The functional $D^{reg}_{\mathbf{H}}$ is jointly concave, but its subdifferential is not necessarily a singleton, since the smallest eigenvalue of $\mathbf{H}_{\text{eff}}$ is, in general, degenerate. The subgradient $(G_1,\dots,G_N)$ of $D^{reg}_{\mathbf{H}}$

is given by $G_i = \gamma_i - \mathrm{Tr}^i\left[|\psi_1\rangle\langle\psi_1|\right]$, where $\psi_1 \in \mathrm{argmin}_{\psi\in\mathbb{C}^D,\,\|\psi\|=1}\langle\psi, (\mathbf{H} - \bigoplus_{i=1}^N U_i)\psi\rangle$, see Proposition 2 in the supplementary material S2.

We solve (5) via subgradient ascent with Polyak step sizes [2]. Each iteration requires the smallest eigenpair, computed with stochastic Lanczos [22]. The algorithm stops when the gradient norm is below tolerance or iterations are exhausted. Updates (Prop. 2) and step-size estimates are given in Algorithm 1, implemented in JAX [3].

2.2. Von Neumann Entropy Regularization of Quantum Optimal Transport.

The dual formulation of the von Neumman regularized QOT is given by [7, 17]

$$\mathfrak{D}^Q_{\mathbf{H},\varepsilon} = \sup_{U_i\in\mathbb{H}_d} D^{reg}_{\mathbf{H},\varepsilon}(U_1,\dots,U_N) := \sup_{U_i\in\mathbb{H}_d}\sum_{i=1}^N \mathrm{Tr}\left[U_i\gamma_i\right] - \varepsilon\log\mathrm{Tr}\left[\exp\left(-\frac{\mathbf{H}_{\mathrm{eff}}}{\varepsilon}\right)\right], \quad (6)$$

where $\mathfrak{D}^Q_{\mathbf{H},\varepsilon} = \mathfrak{D}^Q_{\mathbf{H},\varepsilon}(\gamma_1,\dots,\gamma_N)$ is the cost and $\mathbf{H}_{\mathrm{eff}}$ is the effective Hamiltonian defined in (3).

The dual functional $D^{reg}_{\mathbf{H},\varepsilon}$ in (6) is differentiable and strictly concave and the dual problem $\mathfrak{D}^Q_{\mathbf{H},\varepsilon}$ (6) admits a unique maximizer (up to trivial translations) [17], see also S4) and Theorem 4).

The gradient of $D^{reg}_{\mathbf{H},\varepsilon}$ with respect to U_i is given by $\nabla_{U_i} D^{reg}_{\mathbf{H},\varepsilon} = \gamma_i - Z_\varepsilon^{-1}\,\mathrm{Tr}^i\left[\exp\left(-\frac{\mathbf{H}_{\mathrm{eff}}}{\varepsilon}\right)\right]$, with $Z_\varepsilon = \mathrm{Tr}\left[\exp\left(-\frac{\mathbf{H}_{\mathrm{eff}}}{\varepsilon}\right)\right]$ (Prop. 3). We solve (6) using the L-BFGS algorithm [23] from Optax [16]. The stopping criteria matches the subgradient method.

Finally, notice the right-hand side in equation (6) can be interpreted as a softmin regularization (log-sum-exp term). Indeed, any sequence $(U_i^\varepsilon)_{i\in\mathbb{N}}$ of maximizing the entropy-regularized QOT (6) $\mathfrak{D}^Q_{\mathbf{H},\varepsilon}(\gamma_1,\dots,\gamma_N)$ converges to $(U_i)_{i\in\mathbb{N}}$ a maximizer of the QOT problem (3) when $\varepsilon\to 0^+$ [7]. In particular, the following limit

$$\lim_{\varepsilon\to 0^+}\varepsilon\log\mathrm{Tr}\left[\exp\left(-\frac{\mathbf{H}_{\mathrm{eff}}}{\varepsilon}\right)\right] = \lim_{\varepsilon\to 0^+}\varepsilon\log\sum_{k=1}^D\exp\left(-\frac{\lambda_k^{\mathrm{eff}}}{\varepsilon}\right) = \lambda_1^{\mathrm{eff}}, \quad (7)$$

converges to the smallest eigenvalue of the effective Hamiltonian $\mathbf{H}_{\mathrm{eff}}$, where in (7) $\lambda_1^{\mathrm{eff}} \le \cdots \le \lambda_D^{\mathrm{eff}}$ are the eigenvalues of $\mathbf{H}_{\mathrm{eff}}$. The smaller ε, the closer entropy regularized QOT approaches unregularized QOT - see Figure 1.

3. Numerical experiments

We investigate three distinct classes of QOT problems: (i) QOT between quantum channels governed by a Hamiltonian formulated in terms of quadrature operators [14], see eq. (8); (ii) QOT in spin systems described by the Ising model Hamiltonian; (iii) QOT with random Hamiltonians. A performance comparison between the two algorithms introduced in the

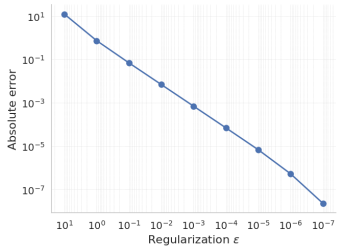

Figure 1: Absolute error $|\mathfrak{D}^Q_{\mathbf{H}} - \mathfrak{D}^Q_{\mathbf{H},\varepsilon}|$ between the QOT problem (3) and the entropy regularized QOT (6) as a function of ε, for the Ising Hamiltonian example in Section 3.2.

previous sections for (i)–(ii) is presented in Table 1. Numerical results for (iii) are given in the supplementary material S8. Closed-form solutions and SDP solvers are used as benchmarks for evaluating the performance of our computational algorithms. All experiments were conducted on a server equipped with an NVIDIA H100 NVL Tensor Core GPU.

3.1. Quantum Optimal Transport between quantum channels.

The first set of experiments focuses on the problem of Quantum Optimal Transport between quantum channels, as introduced by G. De Palma and D. Trevisan in [14] – see also [25].

Consider $N = 2$ identical Hilbert spaces $L^2(\mathbb{R}^m)$, where $m \in \mathbb{N}$. Let $\{Q_j\}_{j\in[m]}$ and $\{P_j\}_{j\in[m]}$ be the position and momentum operators, respectively. These act on functions $\psi \in L^2(\mathbb{R}^m)$ as

$$(Q_j\psi)(q) = q_j\psi(q), \text{ and } (P_j\psi)(q) = -\mathbf{i}\frac{\partial}{\partial q_j}\psi(q), \quad \forall\, j \in [m],\, q = (q_1, \dots, q_m) \in \mathbb{R}^m.$$

The Hamiltonian (cost) operator is defined through the quadrature operators as

$$\mathbf{H} = \sum_{i=1}^{m} \left(Q_i \otimes \mathbb{1} - \mathbb{1} \otimes Q_i^T\right)^2 + \sum_{i=1}^{m} \left(P_i \otimes \mathbb{1} - \mathbb{1} \otimes P_i^T\right)^2. \tag{8}$$

To obtain a finite-dimensional discretization of $L^2(\mathbb{R}^m)$, we adopt the Fock basis $\{|1\rangle, \dots, |d\rangle\}$, corresponding to the first d eigenstates of the harmonic oscillator Hamiltonian $H_{\text{ho}} = \frac{1}{2}(Q^2 + P^2)$. The Hilbert space of each subsystem is then spanned by these d basis vectors, and is thus identified with $\mathbb{C}^d$. Thus, the Hilbert space of the composite system is $\mathbb{C}^{d^2}$, with canonical basis $\{|i\rangle \otimes |j\rangle\}_{i,j\in[d]}$. The finite-dimensional representation of the cost is given by $\mathbf{H}_{(ij)(kl)} = \langle i \otimes j, \mathbf{H}\, k \otimes l\rangle, \forall\, i, j, k, l \in [d]$, resulting in a Hermitian matrix $\mathbf{H} \in \mathbb{H}_{d^2}$.

For our simulations, we fix $d = 50$ and consider the following cases for the marginals $\gamma_1, \gamma_2 \in \mathcal{D}_d$, see example 1 in the supplementary material for the definition of a thermal state.

- **WP**: Both marginals are thermal states at temperature zero, i.e. $\gamma_1 = \gamma_2 = |0\rangle\langle 0|$.
- **WM**: γ_1 is a thermal state at temperature zero; $\gamma_2 = \frac{\mathbb{1}}{d}$ is the maximally mixed state.
- **WC**: The marginals $\gamma_1 = \gamma_2$ are cat states, i.e. $\gamma_i = \text{Tr}^i\left[|\psi_{\text{cat}}(2)\rangle\langle\psi_{\text{cat}}(2)|\right],\ i \in [2]$, where $|\psi_{\text{cat}}(\alpha)\rangle = |\alpha\rangle \otimes |\alpha\rangle + |-\alpha\rangle \otimes |-\alpha\rangle, |\alpha\rangle = e^{-\frac{|\alpha|^2}{2}} \sum_{k=0}^{d-1} \frac{\alpha^k}{\sqrt{k!}} |k\rangle$ normalized to 1.
- **WG**: The marginals $\gamma_1 = \gamma_2$ are taken from a GHZ state, i.e. $\gamma_i = \text{Tr}^i\left[|\psi_{\text{GHZ}}\rangle\langle\psi_{\text{GHZ}}|\right],\ i \in [2]$, where $|\psi_{\text{GHZ}}\rangle = \frac{|0,0\rangle + |1,1\rangle}{\sqrt{2}}$.

Notice that, when γ_1 and γ_2 are thermal states, the QOT coupling is also a thermal state [14], see also Example 1 in the supplementary material. The results are summarized next page on Table 1. The convergence plots are given in the Figure 2.

3.2. Multi-marginal Quantum Optimal Transport for the Ising Model

Consider a one-dimensional lattice with N sites, each occupied by an electron with spin $s_i \in \{-1, +1\}$, $i \in [N]$. Each site is associated with a two-dimensional Hilbert space $\mathbb{C}^2$, and the Hilbert space of the composite system is the tensor product $(\mathbb{C}^2)^{\otimes N}$. The Hamiltonian is given by

$$\mathbf{H} = -J \sum_{i=1}^{N-1} \sigma_i^Z \sigma_{i+1}^Z - h \sum_{i=1}^{N} \sigma_i^Z, \quad \sigma_i^Z = \bigotimes_{j=1}^{i-1} \mathbb{1} \otimes \sigma^Z \otimes \bigotimes_{j=i+1}^{N} \mathbb{1}, \text{ with } \sigma^Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{9}$$

Test	Rank	Tol.	M	Time	Iters	Ach.	Err.
WP	1	1e-3	S	0.057	97	Yes	1.6e-06
		1e-4	S	0.067	121	Yes	1.4e-08
		1e-5	S	0.078	149	Yes	2.0e-09
		1e-6	S	0.095	196	Yes	2.6e-11
		1e-3	L	0.187	2	Yes	3.6e-03
		1e-4	L	0.190	2	Yes	3.6e-03
		1e-5	L	0.191	2	Yes	3.6e-03
		1e-6	L	0.194	2	Yes	3.6e-03
WG	1	1e-3	S	0.194	449	Yes	3.7e-04
		1e-4	S	0.274	658	Yes	1.4e-07
		1e-5	S	0.322	781	Yes	8.0e-10
		1e-6	S	0.379	931	Yes	2.9e-11
		1e-3	L	0.182	2	Yes	3.6e-03
		1e-4	L	0.182	2	Yes	3.6e-03
		1e-5	L	0.184	2	Yes	3.6e-03
		1e-6	L	0.184	2	Yes	3.6e-03
IG	1	1e-3	S	0.185	198	Yes	2.0e-05
		1e-4	S	0.426	543	Yes	5.0e-07
		1e-5	S	0.857	1149	Yes	5.9e-09
		1e-6	S	1.281	1813	Yes	7.9e-11
		1e-3	L	1.597	2	Yes	2.1e-11
		1e-4	L	1.636	2	Yes	2.1e-11
		1e-5	L	1.654	2	Yes	2.1e-11
		1e-6	L	1.624	2	Yes	2.1e-11
IE	1	1e-3	S	0.150	148	Yes	4.5e-06
		1e-4	S	0.480	625	Yes	7.7e-07
		1e-5	S	1.186	1647	Yes	4.3e-08
		1e-6	S	1.251	1741	Yes	5.1e-10
		1e-3	L	1.212	4	Yes	1.8e-02
		1e-4	L	1.358	5	Yes	2.0e-05
		1e-5	L	1.373	5	Yes	2.0e-05
		1e-6	L	1.984	7	Yes	4.1e-10

Test	Rank	Tol.	M	Time	Iters	Ach.	Err.
WC	2	1e-3	S	0.206	493	Yes	1.7e-04
		1e-4	S	0.256	616	Yes	2.3e-05
		1e-5	S	0.378	931	Yes	4.4e-09
		1e-6	S	0.390	964	Yes	2.1e-09
		1e-3	L	0.181	2	Yes	3.6e-03
		1e-4	L	0.988	6	Yes	1.3e-09
		1e-5	L	2.133	27	Yes	2.0e-09
		1e-6	L	54.985	296	Yes	2.1e-09
WM	50	1e-3	S	194.245	500000	No	1.9e+00
		1e-4	S	193.805	500000	No	1.9e+00
		1e-5	S	194.505	500000	No	1.9e+00
		1e-6	S	193.867	500000	No	1.9e+00
		1e-3	L	628.555	10000	No	8.0e-01
		1e-4	L	628.508	10000	No	8.0e-01
		1e-5	L	628.046	10000	No	8.0e-01
		1e-6	L	628.109	10000	No	8.0e-01
IM	1	1e-3	S	0.160	157	Yes	8.7e-06
		1e-4	S	0.263	306	Yes	1.7e-07
		1e-5	S	0.400	504	Yes	1.7e-09
		1e-6	S	0.541	711	Yes	8.7e-11
		1e-3	L	1.086	4	Yes	1.7e-03
		1e-4	L	1.270	5	Yes	3.9e-04
		1e-5	L	1.417	6	Yes	9.0e-07
		1e-6	L	2.055	7	Yes	3.5e-10
IT	233	1e-3	S	11.708	16735	Yes	1.1e-05
		1e-4	S	17.761	25457	Yes	2.8e-07
		1e-5	S	25.883	37097	Yes	2.8e-10
		1e-6	S	32.273	46262	Yes	7.8e-12
		1e-3	L	2840.952	10000	No	3.1e-02
		1e-4	L	2837.323	10000	No	3.1e-02
		1e-5	L	2822.841	10000	No	3.1e-02
		1e-6	L	2827.996	10000	No	3.1e-02

Table 1: Performance comparison between the subgradient and L-BFGS (with von Neumann entropy regularization) algorithms on QOT problems. The columns are as follows: **Test** – identifier of the test instance; **Rank** – rank of the QOT coupling in 1 returned by the ground-truth solution (MOSEK [24]); **Tol.** – gradient norm threshold; **M** – method used (`S` for subgradient, `L` for L-BFGS); **Time** – wall-clock runtime in seconds; **Iters** – number of iterations; **Ach.** – whether the method met the target tolerance; **Err.** – final absolute difference between the objective value and the ground truth.

where $J \in \mathbb{R}$ controls the nearest-neighbor coupling and $h \in \mathbb{R}$ the external field strength. These operators correspond to a spin measurement in the Z-direction at site $i \in [N]$.

For our numerical experiments, we fix $N = 12$ and consider the following test cases for the marginals, derived from low-energy eigenstates $\psi_1, \ldots, \psi_l$ of the Ising Hamiltonain (9), with $l \leq 2^N$ (arranged in ascending eigenvalue order), and thermal mixtures:

- **IG**: The marginals $\gamma_k = \mathrm{Tr}^k \left[|\psi_1\rangle \langle\psi_1|\right]$, for all $k \in [N]$.
- **IE**: The marginals $\gamma_k = \mathrm{Tr}^k \left[|\psi_2\rangle \langle\psi_2|\right]$, for all $k \in [N]$.
- **IM**: The marginals $\gamma_k = \mathrm{Tr}^k \left[\frac{1}{2} |\psi_1\rangle \langle\psi_1| + \frac{1}{2} |\psi_2\rangle \langle\psi_2|\right]$, for all $k \in [N]$.

- **IT**: $\gamma_k = e^{-\frac{H_{\text{ho}}}{k}}/Z_k$ is a thermal state at temperature $k \in [N]$.

The corresponding results are summarized in Table 1 in the next page.

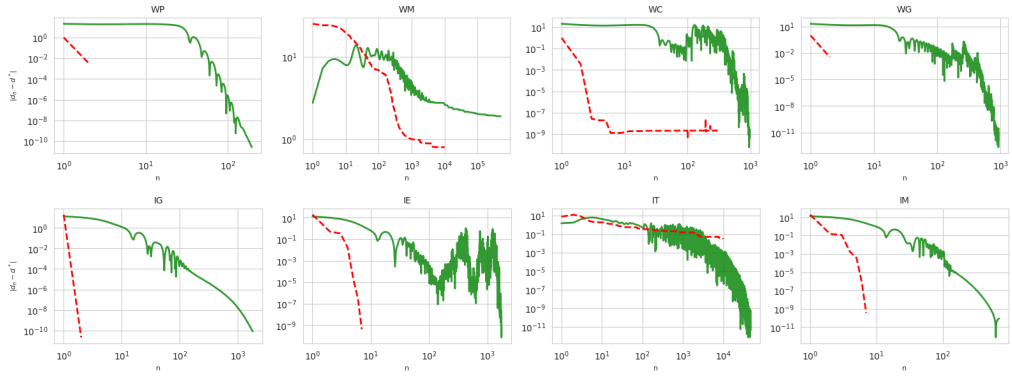

Figure 2: Optimality gap $|d_n - d^*|$ versus iteration for the subgradient method (green) and L-BFGS on the tests in Sections 3.1, 3.2. Here d_n is the objective at iteration n, d^* the MOSEK ground truth, and $\varepsilon = 10^{-12}$.

4. Discussion, conclusion and future work

We propose two computational approaches based on the dual formulation of Quantum Optimal Transport: a subgradient method for the eigenvalue-regularized formulation, and a von Neumann entropy–regularized solved via L-BFGS. Our experiments demonstrate that von Neumann entropy regularization is broadly applicable and generally converges in fewer iterations, though at the expense of a significantly higher per-iteration computational cost. In contrast, the subgradient method is particularly effective when the optimal coupling is (or can be well-approximated by) a rank-1 operator, providing an efficient alternative in such settings. Future work will address scalability in high-dimensional regimes and exploit problem structure in applications such as electronic structure and machine learning. Furthermore, we will concentrate on generalizing our regularization methodology to quantum optimization problems that are commonly formulated as semi-definite programs.

Future applications includes machine learning, for instance, quantum Wasserstein generative adversarial networks [9], where QOT defines the training objective. The von Neumann–regularized QOT provides a smooth, differentiable loss between quantum data, while the subgradient formulation offers an efficient approximation scheme suitable for large-scale or iterative learning settings.

5. Acknowledgments

PP and AG thank the Natural Sciences and Engineering Research Council of Canada (NSERC) for financial support under the Multi-Marginal Optimal Transport grant (Grant No. RGPIN-2022-05207), the Mitacs Accelerate Program, and the Canada Research Chairs Program (Grant No. CRC-2021-00234). AG also thanks the NSERC Alliance grant (Grant No. ALLRP/592521-2023) and the National Research Council of Canada (Grant No. AQC-208-1). Part of this research was conducted while the authors were visiting the Institute for Pure and Applied Mathematics (IPAM), which is supported by the National Science Foundation (Grant No. DMS-1925919).

References

[1] Emily Beatty and Daniel Stilck França. Order p quantum Wasserstein distances from couplings. In *Annales Henri Poincaré*, pages 1–59. Springer, 2025.

[2] Stephen Boyd. Subgradient methods. Lecture notes for EE364b, Stanford University, Accessed 2025. URL https://web.stanford.edu/class/ee364b/lectures/subgrad_method_notes.pdf.

[3] James Bradbury, Roy Frostig, Peter Hawkins, Matthew James Johnson, Chris Leary, Dougal Maclaurin, George Necula, Adam Paszke, Jake VanderPlas, Skye Wanderman-Milne, and Qiao Zhang. JAX: composable transformations of Python+NumPy programs, 2018. URL http://github.com/jax-ml/jax.

[4] Jonatan Bohr Brask. Gaussian states and operations – a quick reference. arXiv preprint arXiv:2102.05748 [quant-ph], 2022. URL https://arxiv.org/abs/2102.05748.

[5] E. Caglioti, F. Golse, and T. Paul. Quantum optimal transport is cheaper. *Journal of Statistical Physics*, 181(1):149–162, 2020. doi: 10.1007/s10955-020-02571-7. URL https://doi.org/10.1007/s10955-020-02571-7.

[6] Emanuele Caglioti, François Golse, and Thierry Paul. Towards optimal transport for quantum densities. *Annali della Scuola Normale Superiore di Pisa, Classe di Scienze*, XXIV(4):1981–2045, 2023. ISSN 2036-2145. doi: 10.2422/2036-2145.202106_011. URL https://doi.org/10.2422/2036-2145.202106_011. Submitted: Jun 9, 2021; Accepted: May 10, 2022; Published: Dec 29, 2023.

[7] Emanuele Caputo, Augusto Gerolin, Nataliia Monina, and Lorenzo Portinale. Quantum optimal transport with convex regularization. arXiv preprint arXiv:2409.03698 [math-ph], 2024. URL https://arxiv.org/abs/2409.03698.

[8] Eric A. Carlen and Jan Maas. An analog of the 2-Wasserstein metric in non-commutative probability under which the fermionic Fokker–Planck equation is gradient flow for the entropy. *Communications in Mathematical Physics*, 331(3):887–926, 2014. doi: 10.1007/s00220-014-2124-8. URL https://doi.org/10.1007/s00220-014-2124-8.

[9] Shouvanik Chakrabarti, Huang Yiming, Tongyang Li, Soheil Feizi, and Xiaodi Wu. Quantum Wasserstein generative adversarial networks. *Advances in Neural Information Processing Systems*, 32, 2019.

[10] Yongxin Chen, Tryphon T Georgiou, and Allen Tannenbaum. Matrix optimal mass transport: a quantum mechanical approach. *IEEE Transactions on Automatic Control*, 63(8):2612–2619, 2017.

[11] Sam Cole, Michał Eckstein, Shmuel Friedland, and Karol Życzkowski. On quantum optimal transport. *Mathematical Physics, Analysis and Geometry*, 26(2):14, 2023. doi: 10.1007/s11040-023-09456-7. URL https://doi.org/10.1007/s11040-023-09456-7.

[12] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems*, 26, 2013.

[13] Francesco D'andrea, Pierre Martinetti, et al. A view on optimal transport from noncommutative geometry. *SIGMA. Symmetry, Integrability and Geometry: Methods and Applications*, 6: 057, 2010.

[14] Giacomo De Palma and Dario Trevisan. Quantum optimal transport with quantum channels. *Annales Henri Poincaré*, 22(10):3199–3234, 2021. doi: 10.1007/s00023-021-01042-3. URL https://doi.org/10.1007/s00023-021-01042-3.

[15] Giacomo De Palma and Dario Trevisan. The wasserstein distance of order 1 for quantum spin systems on infinite lattices. In *Annales Henri Poincaré*, volume 24, pages 4237–4282. Springer, 2023.

[16] DeepMind, Igor Babuschkin, Kate Baumli, Alison Bell, Surya Bhupatiraju, Jake Bruce, Peter Buchlovsky, David Budden, Trevor Cai, Aidan Clark, Ivo Danihelka, Antoine Dedieu, Claudio Fantacci, Jonathan Godwin, Chris Jones, Ross Hemsley, Tom Hennigan, Matteo Hessel, Shaobo Hou, Steven Kapturowski, Thomas Keck, Iurii Kemaev, Michael King, Markus Kunesch, Lena Martens, Hamza Merzic, Vladimir Mikulik, Tamara Norman, George Papamakarios, John Quan, Roman Ring, Francisco Ruiz, Alvaro Sanchez, Laurent Sartran, Rosalia Schneider, Eren Sezener, Stephen Spencer, Srivatsan Srinivasan, Miloš Stanojević, Wojciech Stokowiec, Luyu Wang, Guangyao Zhou, and Fabio Viola. The DeepMind JAX Ecosystem, 2020. URL http://github.com/google-deepmind.

[17] Dario Feliciangeli, Augusto Gerolin, and Lorenzo Portinale. A non-commutative entropic optimal transport approach to quantum composite systems at positive temperature. *Journal of Functional Analysis*, 285(4):1–39, August 2023. ISSN 0022-1236. doi: 10.1016/j.jfa.2023.109963.

[18] Tryphon T Georgiou and Michele Pavon. Positive contraction mappings for classical and quantum Schrödinger systems. *Journal of Mathematical Physics*, 56(3), 2015.

[19] J Robert Johansson, Paul D Nation, and Franco Nori. Qutip: An open-source python framework for the dynamics of open quantum systems. *Computer physics communications*, 183(8): 1760–1772, 2012.

[20] Wiktor Jurasz and Christian B Mendl. Quantum Wasserstein gans for state preparation at unseen points of a phase diagram. *arXiv preprint arXiv:2309.09543*, 2023.

[21] Bobak Toussi Kiani, Giacomo De Palma, Milad Marvian, Zi-Wen Liu, and Seth Lloyd. Learning quantum data with the quantum earth mover's distance. *Quantum Science and Technology*, 7(4):045002, 2022.

[22] J. Kuczyński and H. Woźniakowski. Estimating the largest eigenvalue by the power and Lanczos algorithms with a random start. *SIAM Journal on Matrix Analysis and Applications*, 13 (4):1094–1122, 2025/04/30 1992.

[23] Dong C. Liu and Jorge Nocedal. On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, 45(1):503–528, 1989.

[24] MOSEK ApS. *MOSEK Optimization Suite*, 2025. URL https://www.mosek.com/. Version 10.1.

[25] Dario Trevisan. Quantum optimal transport: an invitation. *Bollettino dell'Unione Matematica Italiana*, 18(1):347–360, 2025. doi: 10.1007/s40574-024-00428-5. URL https://doi.org/10.1007/s40574-024-00428-5.

Supplementary Material

S1 Saddle Point of the Quantum Optimal Transport Lagrangian

Proposition 1 *Let $d, N \in \mathbb{N}$ be the natural numbers, $D = d^N$, $\gamma_1, \dots, \gamma_N \in \mathcal{D}_d$ be density matrices, $\mathbf{H} \in \mathbb{H}_d$ be a Hermitian cost operator. Let $U_i^* \in \mathbb{H}_d,\ i \in [N]$ and $\psi^* \in \mathbb{C}^D$ be the saddle point of the Lagrangian:*

$$\mathcal{L} : \bigtimes_{i=1}^{N} \mathbb{H}_d \times \mathbb{C}^d \to \mathbb{R}, \quad \mathcal{L}(U_1, \dots, U_N, \psi) = \sum_{i=1}^{N} \operatorname{Tr}\left[U_i \gamma_i\right] + \langle \psi, (\mathbf{H} - \bigoplus_{i=1}^{N} U_i)\psi \rangle. \tag{S10}$$

Then $|\psi^*\rangle \langle\psi^*|$ is the solution of (1), and $U_1^*, \dots, U_N^*$ is the solution of (3).

Proof Let $U_i^* \in \mathbb{H}_d,\ i \in [N]$ and $\psi^* \in \mathbb{C}^D$ be the saddle point of (S10). Then, for any $U_i \in \mathbb{H}_d,\ i \in [N]$ and for any unit vector $\psi \in \mathbb{C}^D$ holds

$$\mathcal{L}(U_1, \dots, U_N, \psi^*) \leq \mathcal{L}(U_1^*, \dots, U_N^*, \psi^*) \leq \mathcal{L}(U_1^*, \dots, U_N^*, \psi).$$

However, we need to show that for any $U_i \in \mathbb{H}_d,\ i \in [N]$ and for any $\Gamma \in \mathcal{D}_D$ holds

$$\mathcal{L}^{QOT}(U_1, \dots, U_N, |\psi^*\rangle \langle\psi^*|) \leq \mathcal{L}^{QOT}(U_1^*, \dots, U_N^*, |\psi^*\rangle \langle\psi^*|) \leq \mathcal{L}^{QOT}(U_1^*, \dots, U_N^*, \Gamma),$$

where

$$\begin{aligned} &\mathcal{L}^{QOT}(U_1, \dots, U_N, \Gamma) : \bigtimes_{i=1}^{N} \mathbb{H}_d \times \mathcal{D}_D \\ &\mathcal{L}^{QOT}(U_1, \dots, U_N, \Gamma) = \sum_{i=1}^{N} \operatorname{Tr}\left[U_i \gamma_i\right] + \operatorname{Tr}\left[(\mathbf{H} - \bigoplus_{i=1}^{N} U_i)\Gamma\right] \end{aligned} \tag{S11}$$

is a Lagrangian of QOT problem (1). The left inequality we already have due to the fact that $\mathcal{L}(U_1, \dots, U_N, \psi) = \mathcal{L}^{QOT}(U_1, \dots, U_N, |\psi\rangle \langle\psi|)$. So, it suffices to show only the right one.

For any $\Gamma = \sum_{j=1}^{D} \lambda_j |\phi_j\rangle \langle\phi_j| \in \mathcal{D}_D$ and $U_i \in \mathbb{H}_d,\ i \in [N]$, the (S11) reads

$$\begin{aligned} &\mathcal{L}^{QOT}(U_1, \dots, U_N, \Gamma) \\ &= \sum_{i=1}^{N} \operatorname{Tr}\left[U_i \gamma_i\right] + \operatorname{Tr}\left[(\mathbf{H} - \bigoplus_{i=1}^{N} U_i)\Gamma\right] \\ &= \sum_{i=1}^{N} \operatorname{Tr}\left[U_i \gamma_i\right] + \sum_{j=0}^{D-1} \lambda_j \langle \phi_j, (\mathbf{H} - \bigoplus_{i=1}^{N} U_i)\phi_j \rangle \\ &= \sum_{j=1}^{D} \lambda_j \left(\sum_{i=1}^{N} \operatorname{Tr}\left[U_i \gamma_i\right] + \langle \phi_j, (\mathbf{H} - \bigoplus_{i=1}^{N} U_i)\phi_j \rangle \right) \qquad \left(\text{Since } \sum_{j=1}^{D} \lambda_j = 1\right) \\ &= \sum_{j=1}^{D} \lambda_j \mathcal{L}(U_1, \dots, U_N, \phi_j). \end{aligned}$$

Thus, we have

$$\begin{aligned}\mathcal{L}^{QOT}(U_1^*,\dots,U_N^*,|\psi^*\rangle\langle\psi^*|) &= \sum_{j=1}^{D}\lambda_j\mathcal{L}(U_1^*,\dots,U_N^*,\psi^*)\\ &\le \sum_{j=1}^{D}\lambda_j\mathcal{L}(U_1^*,\dots,U_N^*,\phi_j)\\ &= \mathcal{L}^{QOT}(U_1^*,\dots,U_N^*,\Gamma),\end{aligned}$$

which proves that $(U_1^*,\dots,U_N^*,|\psi^*\rangle\langle\psi^*|)$ is a saddle point for $\mathcal{L}^{QOT}$. Thus, $|\psi^*\rangle\langle\psi^*|$ minimizes (1), and $U_1^*,\dots,U_N^*$ maximizes (3). ■

S2 Properties of the Eigenvalue Regularized Dual Functional

Proposition 2 *Let $d, N \in \mathbb{N}$ be the natural numbers, $D = d^N$, $\gamma_1,\dots,\gamma_N \in \mathcal{D}_d$ be density matrices, $\mathbf{H} \in \mathbb{H}_D$ be a Hermitian operator, and $D_{\mathbf{H}}^{reg}$ be the dual functional defined in (4). Then*

1. *$D_{\mathbf{H}}^{reg}$ is a jointly concave function and bounded from above by the primal problem (1).*
2. *Let $G_i = \gamma_i - \mathrm{Tr}^i[|\psi_1\rangle\langle\psi_1|]$, where $\psi_1 \in \mathrm{Argmin}_{\psi\in\mathbb{C}^D:\,\|\psi\|=1}\langle\psi,(\mathbf{H}-\bigoplus_{i=1}^N U_i)\psi\rangle$. Then, $(G_1,\dots,G_N)$ is a subgradient of $D_{\mathbf{H}}^{reg}$ at $(U_1,\dots,U_N)$. Moreover, for every $i\in[N]$, G_i is traceless, that is $\mathrm{Tr}[G_i]=0$, and its the norm of $(G_1,\dots G_N)$ is bounded by*

$$\sum_{i=1}^{N}\left\|\Lambda_i-\overline{\Lambda}\right\|_2^2 \le \sum_{i=1}^{N}\|G_i\|_2^2 \le 4N,$$

where $\Lambda_i = (\lambda_1^i,\dots,\lambda_d^i)$, $\overline{\Lambda} = 1/N\sum_{i=1}^N\Lambda_i$, and $\lambda_1^i \le \dots \le \lambda_d^i$ are the eigenvalues of γ_i, for all $i\in[N]$.

Proof Let's prove the first statement. Note that since $D_{\mathbf{H}}^{reg}(U_1,\dots,U_N) = D(\bigoplus_{i=1}^N U_i)$, where

$$D:\mathbb{H}_D\to\mathbb{R},\quad D(B) = \mathrm{Tr}\left[B\bigotimes_{i=1}^{N}\gamma_i\right] + \min_{\substack{\psi\in\mathbb{C}^D\\ \|\psi\|=1}}\langle\psi,(\mathbf{H}-B)\psi\rangle.$$

It is enough to show only the concavity of $D(B)$. Let $B_1, B_2 \in \mathbb{H}_D,\ \alpha \in [0,1]$, then

$$\begin{aligned}
D(\alpha B_1 + (1-\alpha)B_2) &= \operatorname{Tr}\left[(B_1 + (1-\alpha)B_2)\bigotimes_{i=1}^{N}\gamma_i\right] + \min_{\substack{\psi\in\mathbb{C}^D\\ \|\psi\|=1}} \langle\psi, (\mathbf{H} - (\alpha B_1 + (1-\alpha)B_2))\,\psi\rangle \\
&= \alpha \operatorname{Tr}\left[B_1\bigotimes_{i=1}^{N}\gamma_i\right] + (1-\alpha)\operatorname{Tr}\left[B_2\bigotimes_{i=1}^{N}\gamma_i\right] \\
&\quad + \min_{\substack{\psi\in\mathbb{C}^D\\ \|\psi\|=1}} \langle\psi, (\alpha(\mathbf{H}-B_1) + (1-\alpha)(\mathbf{H}-B_2))\,\psi\rangle \\
&\geq \alpha \operatorname{Tr}\left[B_1\bigotimes_{i=1}^{N}\gamma_i\right] + (1-\alpha)\operatorname{Tr}\left[B_2\bigotimes_{i=1}^{N}\gamma_i\right] \\
&\quad + \alpha \min_{\substack{\psi\in\mathbb{C}^D\\ \|\psi\|=1}} \langle\psi, (\mathbf{H}-B_1)\psi\rangle + (1-\alpha)\min_{\substack{\psi\in\mathbb{C}^D\\ \|\psi\|=1}} \langle\psi, (\mathbf{H}-B_2)\psi\rangle \\
&= \alpha D(B_1) + (1-\alpha)D(B_2).
\end{aligned}$$

This follows from the linearity of the trace and the fact that the sum of minima is less than or equal to the minimum of the sum.

Consequently, for any $\alpha \in [0,1]$, $U_i^1, U_i^2 \in \mathbb{H}_d, U_i^\alpha = \alpha U_i^1 + (1-\alpha)U_i^2,\ i \in [N]$,

$$D_{\mathbf{H}}^{reg}(U_1^\alpha, \ldots, U_N^\alpha) \geq \alpha D_{\mathbf{H}}^{reg}(U_1^1, \ldots U_N^1) + (1-\alpha)D_{\mathbf{H}}^{reg}(U_1^2, \ldots U_N^2).$$

Secondly, to show that $D_{\mathbf{H}}^{reg}$ is bounded from above, we may simply notice that

$$D_{\mathbf{H}}^{reg}(U_1, \ldots, U_N) = \sum_{i=1}^{N} \operatorname{Tr}[U_i\gamma_i] + \lambda_1^{eff} = \sum_{i=1}^{N} \operatorname{Tr}\left[(U_i + \frac{\lambda_1^{eff}}{N}\mathbb{1}_i)\gamma_i\right] \leq \mathfrak{P}_{\mathbf{H}}^{Q},$$

since $(U_1 + \frac{\lambda_1^{eff}}{N}\mathbb{1}_1, \ldots, U_N + \frac{\lambda_1^{eff}}{N}\mathbb{1}_N)$ is a feasible point for the dual problem (3), that is $\mathbf{H} - \bigoplus_{i=1}^{N}(U_i + \frac{\lambda_1^{eff}}{N}\mathbb{1}_i) \succcurlyeq 0$.

To prove the third statement, for simplicity, consider the case when the last $N-1$ variables are fixed. Then, for fixed $U_2, \ldots, U_N \in \mathbb{H}_d$ consider a function

$$d : \mathbb{H}_d \to \mathbb{R}, \quad d(U) = D_{\mathbf{H}}^{reg}(U, U_2, \ldots, U_N).$$

Then,

$$\begin{aligned}
d(U) &= \operatorname{Tr}[U\gamma_1] + \min_{\substack{\psi\in\mathbb{C}^D\\ \|\psi\|=1}} \langle\psi, (\mathbf{H} - \bigoplus_{i=2}^{N} U_i - U\otimes\bigotimes_{i=2}^{N}\mathbb{1}_i)\psi\rangle + \sum_{i=2}^{N}\operatorname{Tr}[U_i\gamma_i] \\
&= \operatorname{Tr}[U\gamma_1] + \min_{\substack{\psi\in\mathbb{C}^D\\ \|\psi\|=1}} \langle\psi, (\tilde{\mathbf{H}} - U\otimes\bigotimes_{i=2}^{N}\mathbb{1}_i)\psi\rangle + C,
\end{aligned}$$

where $\tilde{\mathbf{H}} = \mathbf{H} - \bigoplus_{i=2}^{N} U_i$. Then, for any $U' \in \mathbb{H}_d$

$$\begin{aligned}
d(U') - d(U) &= \operatorname{Tr}\left[(U'-U)\gamma_1\right] + \min_{\substack{\psi\in\mathbb{C}^D \\ \|\psi\|=1}} \langle \psi, (\tilde{\mathbf{H}} - U' \otimes \bigotimes_{i=2}^{N} \mathbb{1}_i)\psi\rangle - \min_{\substack{\psi\in\mathbb{C}^D \\ \|\psi\|=1}} \langle \psi, (\tilde{\mathbf{H}} - U \otimes \bigotimes_{i=2}^{N} \mathbb{1}_i)\psi\rangle \\
&= \operatorname{Tr}\left[(U'-U)\gamma_1\right] + \min_{\substack{\psi\in\mathbb{C}^D \\ \|\psi\|=1}} \langle \psi, (\tilde{\mathbf{H}} - U' \otimes \bigotimes_{i=2}^{N} \mathbb{1}_i)\psi\rangle - \langle \psi_1, (\tilde{\mathbf{H}} - U \otimes \bigotimes_{i=2}^{N} \mathbb{1}_i)\psi_1\rangle \\
&\leq \operatorname{Tr}\left[(U'-U)\gamma_1\right] + \langle \psi_1, (\tilde{\mathbf{H}} - U' \otimes \bigotimes_{i=2}^{N} \mathbb{1}_i)\psi_1\rangle - \langle \psi_1, (\tilde{\mathbf{H}} - U \otimes \bigotimes_{i=2}^{N} \mathbb{1}_i)\psi_1\rangle \\
&= \operatorname{Tr}\left[(U'-U)\gamma_1\right] + \langle \psi_1, \tilde{\mathbf{H}}\psi_1\rangle - \operatorname{Tr}\left[\operatorname{Tr}^1\left[|\psi_1\rangle\langle\psi_1|\right](U'-U)\right] - \langle \psi_1, \tilde{\mathbf{H}}\psi_1\rangle \\
&= \operatorname{Tr}\left[(U'-U)(\gamma_1 - \operatorname{Tr}^1\left[|\psi_1\rangle\langle\psi_1|\right])\right].
\end{aligned}$$

The resulting subgradient is traceless

$$\operatorname{Tr}\left[\gamma_1 - \operatorname{Tr}^1\left[|\psi_1\rangle\langle\psi_1|\right]\right] = \operatorname{Tr}\left[\gamma_1\right] - \operatorname{Tr}\left[|\psi_1\rangle\langle\psi_1|\right] = 0.$$

Now, consider the subgradient $(G_1, \ldots, G_N)$, where $G_i = \gamma_i - \operatorname{Tr}^i\left[|\psi_1\rangle\langle\psi_1|\right], i \in [N]$. Upper bound for the l^2-norm of each component G_i follows from the triangle inequality and from the fact that l^1-norm of matrix is bigger than l^2-norm

$$\left\|\gamma_i - \operatorname{Tr}^i\left[|\psi_1\rangle\langle\psi_1|\right]\right\|_2 \leq \|\gamma_i\|_2 + \left\|\operatorname{Tr}^i\left[|\psi_1\rangle\langle\psi_1|\right]\right\|_2 \leq \|\gamma_i\|_1 + \left\|\operatorname{Tr}^i\left[|\psi_1\rangle\langle\psi_1|\right]\right\|_1 = 2.$$

The lower bound on $(G_1, \ldots, G_N)$ follows from the Wielandt-Hoffman theorem

$$\sum_{i=1}^{N} \left\|\gamma_i - \operatorname{Tr}^i\left[|\psi_1\rangle\langle\psi_1|\right]\right\|_2^2 \geq \sum_{i=1}^{N} \|\Lambda_i - \Lambda\|_2^2,$$

where $\Lambda_i = (\lambda_1^i, \ldots, \lambda_d^i)$, and $\lambda_1^i \leq \cdots \leq \lambda_d^i$ are the eigenvalues of γ_i, for all $i \in [N]$. $\Lambda = (\lambda_1, \ldots, \lambda_d)$ denotes the spectrum of $\operatorname{Tr}^i\left[|\psi_1\rangle\langle\psi_1|\right], i \in [N]$; since the partial traces are obtained from a pure state, their spectrum $\lambda_1 \leq \cdots \leq \lambda_d$ is the same.

Finally, by taking infimum of RHS w.r.t. all possible values of spectrum Λ,

$$\inf_{\substack{\Lambda\in\mathbb{R}_+^d \\ \sum_{j=1}^d \Lambda_j = 1}} \sum_{i=1}^{N} \|\Lambda_i - \Lambda\|_2^2.$$

We observe that minimum is achieved at $\overline{\Lambda} = 1/N \sum_{i=1}^{N} \Lambda_i$, which is an attainable value. Thus, we have the bounds

$$\sum_{i=1}^{N} \left\|\Lambda_i - \overline{\Lambda}\right\|_2^2 \leq \sum_{i=1}^{N} \|G_i\|_2^2 \leq 4N.$$

■

S3 Subgradient Ascent Algorithm

The algorithm we used to find the optimal point of (5) is given in the Algorithm 1.

Algorithm 1: Subgradient Ascent for the eigenvalue regularization in (5)

Data:
$U_1^0, \dots, U_N^0 \in \mathbb{H}_d$ (initial traceless Hermitian duals);
$\mathbf{H} \in \mathbb{H}_D$ (cost Hamiltonian);
$\gamma_1, \dots, \gamma_N \in \mathcal{D}_d$ (marginals)
Result:
Maximizing sequence for (3) $U_1^*, \dots, U_N^* \in \mathbb{H}_d$

$\overline{f} \leftarrow 0$ `// Initialize estimated optimal value`

for $n \leftarrow 0, 1, 2, \dots$ **do**
  $\mathbf{H}_n^{\text{eff}} \leftarrow \mathbf{H} - \bigoplus_{i=1}^N U_i^n$
  $\psi_1^n \leftarrow \arg\min_\psi \langle \psi, \mathbf{H}_n^{\text{eff}} \psi \rangle$
  $f_n \leftarrow \sum_{i=1}^N \operatorname{Tr}[U_i^n \gamma_i] + \langle \psi_1^n, \mathbf{H}_n^{\text{eff}} \psi_1^n \rangle$
  if $f_n > \overline{f}$ **then**
    $\overline{f} \leftarrow f_n$
  end
  $\tau_n \leftarrow \dfrac{|\overline{f}|}{n+1}$ `// Correction term for` $\overline{f}$
  $\eta_n \leftarrow \dfrac{\overline{f} + \tau_n - f_n}{4N}$ `// Approx. Polyak step size`
  for $i \leftarrow 1$ **to** N **do**
    $U_i^{n+1} \leftarrow U_i^n + \eta_n \big(\gamma_i - \operatorname{Tr}^i [|\psi_1^n\rangle \langle \psi_1^n|]\big)$
  end
end

S4 Properties of the Entropy Regularized Dual Functional

Proposition 3 *Let $d, N \in \mathbb{N}$ be the natural numbers, $D = d^N$, $\gamma_1, \dots, \gamma_N \in \mathcal{D}_d$ be density matrices, $\mathbf{H} \in \mathbb{H}_D$ be a Hermitian operator, and $D^{reg}_{\mathbf{H},\varepsilon}$ be the dual QOT functional*

$$D^{reg}_{\mathbf{H},\varepsilon} : \bigtimes_{i=1}^N \mathbb{H}_d \to \mathbb{R}, \quad D^{reg}_{\mathbf{H},\varepsilon}(U_1, \dots, U_N) = \sum_{i=1}^N \operatorname{Tr}[U_i \gamma_i] - \varepsilon \log \operatorname{Tr}\left[\exp\left(-\frac{\mathbf{H}_{\text{eff}}}{\varepsilon}\right)\right] \tag{S12}$$

for given regularization parameter $\varepsilon > 0$. Then

1. *The dual functional $D^{reg}_{\mathbf{H},\varepsilon}$ defined as in (S12) is differentiable and strictly concave function;*
2. *The gradient of $D^{reg}_{\mathbf{H},\varepsilon}$ with respect to U_i is $\nabla_{U_i} D^{reg}_{\mathbf{H},\varepsilon} = \gamma_i - \dfrac{\operatorname{Tr}^i\left[\exp\left(-\frac{\mathbf{H}_{\text{eff}}}{\varepsilon}\right)\right]}{\operatorname{Tr}\left[\exp\left(-\frac{\mathbf{H}_{\text{eff}}}{\varepsilon}\right)\right]}$.*

Proof The fact that it is continuous follows from the fact that $D^{reg}_{\mathbf{H}}$ is a sum of the linear function, which is continuous, and the logarithm of strictly positive continuous function, so the whole expression is continuous. To prove strict concavity, it suffices to show that $\log \operatorname{Tr}\left[\exp\left(-\frac{\mathbf{H}_{\text{eff}}}{\varepsilon}\right)\right]$ is strictly convex. Let $\alpha \in (0,1)$, $U_i^1, U_i^2 \in \mathbb{H}_d,\ i \in [N]$. Then we have

$$\begin{aligned}
&\operatorname{Tr}\left[\exp\left(-\frac{\mathbf{H}-\bigoplus_{i=1}^N(\alpha U_i^1+(1-\alpha)U_i^2)}{\varepsilon}\right)\right]\\
&=\operatorname{Tr}\left[\exp\left(-\frac{\alpha(\mathbf{H}-\bigoplus_{i=1}^N U_i^1)+(1-\alpha)(\mathbf{H}-\bigoplus_{i=1}^N U_i^2)}{\varepsilon}\right)\right]\\
&=\operatorname{Tr}\left[\exp\left(-\alpha\frac{\mathbf{H}-\bigoplus_{i=1}^N U_i^1}{\varepsilon}-(1-\alpha)\frac{\mathbf{H}-\bigoplus_{i=1}^N U_i^2}{\varepsilon}\right)\right],
\end{aligned}$$

so we will show that $f(A)=\log\operatorname{Tr}[\exp(A)]$ is strictly convex. For two hermitian matrices $A, B \in \mathbb{H}_d$, let $v_1,\dots,v_d$ be the eigenvectors of $\alpha A+(1-\alpha)B$, then

$$\begin{aligned}
&\log\operatorname{Tr}\left[\exp(\alpha A+(1-\alpha)B)\right]\\
&=\log\sum_{i=1}^d e^{\langle v_i,(\alpha A+(1-\alpha)B)v_i\rangle}\\
&=\log\sum_{i=1}^d e^{\alpha\langle v_i,Av_i\rangle+(1-\alpha)\langle v_i,Bv_i\rangle}\\
&\le \alpha\log\sum_{i=1}^d e^{\langle v_i,Av_i\rangle}+(1-\alpha)\log\sum_{i=1}^d e^{\langle v_i,Bv_i\rangle} && \text{(log-sum-exp convexity)}\\
&< \alpha\log\sum_{i=1}^d\langle v_i,e^A v_i\rangle+(1-\alpha)\log\sum_{i=1}^d\langle v_i,e^B v_i\rangle && \text{(strict Jensen's inequality)}\\
&=\alpha\log\operatorname{Tr}\left[e^A\right]+(1-\alpha)\log\operatorname{Tr}\left[e^B\right].
\end{aligned}$$

So finally we have proved the strong concavity of $D^{reg}_{\mathbf{H},\varepsilon}$.

The form of the gradient we could prove via the Fréchet derivative, since it will give clear and intuitive proof. Let us prove the result for U_1 since others can be shown equivalently. Again, the linear term does not cause any complication, so we merely need to work on the log-trace-exp. for the function $f(A)=\log\operatorname{Tr}[\exp(A)]$ the Fréchet derivative $Df(A)$ in the direction A' is

$$Df(A)=\frac{\operatorname{Tr}\left[A'\exp(A)\right]}{\operatorname{Tr}\left[\exp(A)\right]}.$$

Thus, is we perturb U_1 by U_1', we basically perturb the log-trace-exp term by $U_1'\otimes\bigotimes_{i=2}^N\mathbb{1}_i$, so the Fréchet derivative is:

$$\frac{\operatorname{Tr}\left[U_1'\otimes\bigotimes_{i=2}^N\mathbb{1}_i\exp\left(-\frac{\mathbf{H}-\bigoplus_{i=1}^N U_i}{\varepsilon}\right)\right]}{\varepsilon\operatorname{Tr}\left[\exp\left(-\frac{\mathbf{H}-\bigoplus_{i=1}^N U_i}{\varepsilon}\right)\right]}=\operatorname{Tr}\left[U_1'\frac{\operatorname{Tr}^1\left[\exp\left(-\frac{\mathbf{H}-\bigoplus_{i=1}^N U_i}{\varepsilon}\right)\right]}{\varepsilon\operatorname{Tr}\left[\exp\left(-\frac{\mathbf{H}-\bigoplus_{i=1}^N U_i}{\varepsilon}\right)\right]}\right].$$

Thus, the Fréchet derivative of $D^{reg}_{\mathbf{H},\varepsilon}$ w.r.t. U_1 in the direction U_1' is

$$D_{U_1} D^{reg}_{\mathbf{H},\varepsilon} = \operatorname{Tr}\left[U_1'\gamma_1\right] - \operatorname{Tr}\left[U_1' \frac{\operatorname{Tr}^1\left[\exp\left(-\frac{\mathbf{H}-\bigoplus_{i=1}^N U_i}{\varepsilon}\right)\right]}{\operatorname{Tr}\left[\exp\left(-\frac{\mathbf{H}-\bigoplus_{i=1}^N U_i}{\varepsilon}\right)\right]}\right],$$

which proves the form of the gradient, considering the continuity of $D^{reg}_{\mathbf{H},\varepsilon}$.

■

S5 Von Neumman-regularized Optimal Transport

The primal formulation of von Neumman-regularized Optimal Transport is given by

$$\mathfrak{P}^Q_{\mathbf{H},\varepsilon}(\gamma_1,\dots,\gamma_N) = \min_{\Gamma}\{\operatorname{Tr}\left[\mathbf{H}\Gamma\right] + \varepsilon \operatorname{Tr}\left[\Gamma\left(\log\Gamma - \mathbb{1}\right)\right] \,:\, \Gamma \mapsto (\gamma_1,\dots,\gamma_N)\}. \tag{S13}$$

The von Neumann entropy-regularization term enforces the optimal quantum coupling be a Gibbs state [17]: if $(\overline{U}_i^\varepsilon)_{i\in[N]}$ is a optimal dual entropy-potentials in (6) then

$$\Gamma^\varepsilon = \exp\left(\frac{\bigoplus_{i=1}^N \overline{U}_i^\varepsilon - \mathbf{H}}{\varepsilon}\right), \tag{S14}$$

is a feasible solution $\Gamma^\varepsilon \mapsto (\gamma_1,\dots,\gamma_N)$ and solves the primal formulation of the regularized QOT problem in eq. (6) [17].

The next theorem state and fully characterizes the primal-dual solutions for the von Neumann entropy-regularized Quantum optimal Transport problem in (6) and (S13), see [17].

Theorem 4 *(Theorem 2.1 in [17]) Fix $d, N \in \mathbb{N}$, $D = d^N$ and $\varepsilon > 0$. Let $\mathbf{H} \in \mathbb{H}_D$ be a Hermitian cost operator and $\gamma_1,\dots,\gamma_N \in \mathcal{D}_d$ be density matrices in $\mathbb{C}^d$. Then*

1. (Strong duality) $\mathfrak{P}^Q_{\mathbf{H},\varepsilon}(\gamma_1,\dots,\gamma_N) = \mathfrak{D}^Q_{\mathbf{H},\varepsilon}(\gamma_1,\dots,\gamma_N)$.
2. (Existence of maximizer) *The dual problem $\mathfrak{D}^Q_{\mathbf{H},\varepsilon}(\gamma_1,\dots,\gamma_N)$ in (6) admits a maximizer $\overline{U}_1^\varepsilon,\dots,\overline{U}_N^\varepsilon \in \mathbb{H}_d$ that is unique up to a trivial shift.*
3. (Existence of minimizer) *The primal problem $\mathfrak{P}^Q_{\mathbf{H},\varepsilon}(\gamma_1,\dots,\gamma_N)$ admits a unique minimizer $\Gamma^\varepsilon \in \mathcal{D}_D$ which is given by (S14).*

S6 Wigner Function and Gaussian States in Continuous Variables

Let $\mathbf{q}, \mathbf{p} \in \mathbb{R}^m$ be the position and momentum vectors, and define the phase space vector

$$\mathbf{r} = (q_1,\dots,q_m,p_1,\dots,p_m)^T \in \mathbb{R}^{2m}.$$

Let $|q\rangle$ denote an eigenvector of the position operator Q with eigenvalue $q \in \mathbb{R}$, and define the tensor product $|\mathbf{q}\rangle = |q_1\rangle \otimes \cdots \otimes |q_m\rangle$. Let ρ be a density operator on the Hilbert space $\mathcal{H} = L^2(\mathbb{R}^m)$. The *Wigner transform* of ρ is the function $W(\mathbf{q}, \mathbf{p})$ on phase space $\mathbb{R}^{2m}$ defined by

$$W(\mathbf{q}, \mathbf{p}) = \frac{1}{(2\pi)^m} \int_{\mathbb{R}^m} \left\langle \mathbf{q} + \frac{1}{2}\mathbf{z} \middle| \rho \middle| \mathbf{q} - \frac{1}{2}\mathbf{z} \right\rangle e^{i\mathbf{p}\cdot\mathbf{z}}\, d\mathbf{z}.$$

This function represents the quantum state in phase space, though it is generally not pointwise positive. A particularly important class of states are *Gaussian states*, for which the Wigner function is a multivariate normal distribution.

Define the canonical operators $R_{2k-1} = Q_k$, $R_{2k} = P_k$ for $k \in [m]$. Then

- The *mean vector* is $\overline{\mathbf{r}} \in \mathbb{R}^{2m}$, with entries $\overline{r}_i = \operatorname{Tr}[R_i \rho]$, for all $i \in [2m]$.
- The *covariance matrix* $\Sigma \in \mathbb{R}^{2m \times 2m}$ is given by

$$\Sigma_{ij} = \frac{1}{2} \operatorname{Tr}[\{R_i, R_j\}\rho] - \overline{r}_i \overline{r}_j, \quad \text{for all } i, j \in [2m].$$

  where $\{R_i, R_j\} = R_i R_j + R_j R_i$ is the anticommutator.

Then the Wigner function of a Gaussian state is

$$W(\mathbf{r}) = \frac{1}{(2\pi)^m \sqrt{\det \Sigma}} \exp\left(-\frac{1}{2} (\mathbf{r} - \overline{\mathbf{r}})^T \Sigma^{-1} (\mathbf{r} - \overline{\mathbf{r}}) \right).$$

Numerical Computation from the Fock Basis: if ρ is represented in the Fock basis $\{|n\rangle\}_{n \geq 0}$, the Wigner function at phase space point $\alpha = \frac{q + \mathrm{i}p}{\sqrt{2}} \in \mathbb{C}$ can be expressed as

$$W(\alpha) = \frac{2}{\pi} \operatorname{Tr}\left[D(\alpha) \rho D^\dagger(\alpha) P \right],$$

where $D(\alpha) = \exp(\alpha a^\dagger - \alpha^* a), a = \frac{Q + \mathrm{i}P}{\sqrt{2}}$ is the displacement operator, and $P = \sum_n (-1)^n |n\rangle \langle n|$ is the parity operator. In the Fock basis, this becomes

$$W(\alpha) = \frac{2}{\pi} \sum_{m,n} \rho_{mn} \langle n | D(2\alpha) | m \rangle (-1)^m,$$

where matrix elements of $D(2\alpha)$ can be computed analytically or numerically, for instance, using `QuTiP` [19].

S7 Thermal states

Example 1 (Thermal state) *One important example of quantum channel is a so-called thermal state. A thermal state γ at inverse temperature $\beta > 0$ state is defined as*

$$\gamma_\beta = \frac{e^{-\beta H_{\text{ho}}}}{Z_\beta} = \sum_{n=0}^{\infty} \frac{e^{-\beta(n+\frac{1}{2})}}{Z_\beta} |n\rangle\langle n|, \; Z_\beta = \operatorname{Tr}\left[e^{-\beta H_{\text{ho}}} \right] = \sum_{n=0}^{\infty} e^{-\beta(n+\frac{1}{2})} = \frac{e^{-\beta/2}}{1 - e^{-\beta}}, \quad \text{(S15)}$$

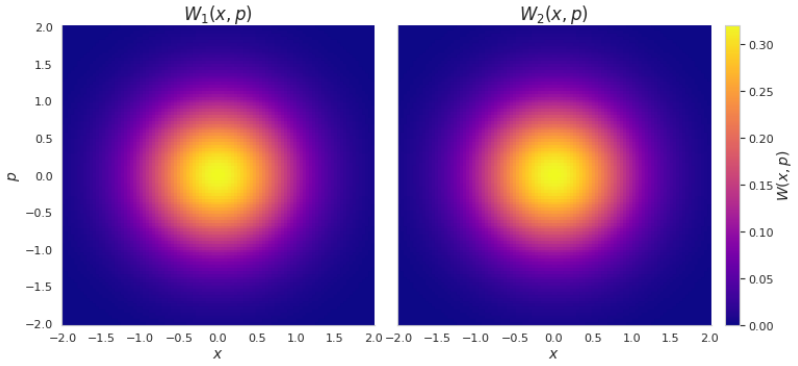

Figure S3: Wigner transforms of the (partial traces of) the quantum optimal coupling $\Gamma^* \in \mathbb{C}^D$ solving the QOT problem 1 in example 1 (see also Proposition 5). The space $\mathbb{C}^D$, with $D = 2500$, corresponds to a truncation of the Fock basis to the first $d = 50$ elements. The Wigner transforms $W_1(x_1, p_1)$ and $W_2(x_2, p_2)$ (see Appendix S6) depict the marginals of the partial traces of Γ^*. In this case, Γ^* is a thermal state at zero temperature.

where $H_{\text{ho}} = \frac{1}{2}(Q^2 + P^2)$ *is the Hamiltonian of the quantum harmonic oscillator, with* Q *and* P *denoting the position and momentum operators, respectively. The eigenstates of* H_{ho} *are the Fock states* $\{|n\rangle\}_{n\in\mathbb{N}}$*, with corresponding energies* $E_n = n + \frac{1}{2}$ *(*$\hbar = 1$*). In the zero-temperature limit* $\beta \to \infty$*, the thermal state approaches the* vacuum state $|0\rangle$*):* $\gamma_\infty = \lim_{\beta\to\infty} \gamma_\beta = |0\rangle\langle 0|$.

Thermal states belong to a special class of quantum states known as Gaussian states, whose Wigner transforms are Gaussian distributions in phase space, see [4] for details. Gaussian states are completely characterized by their first moments (means) and second moments (covariance matrices). They are relevant because closed-form solutions to (1) were derived in [14] for thermal states with covariance matrices proportional to the identity, where the cost Hamiltonian is given by (8). These analytical solutions are used to benchmark our algorithms in for the cost operator (**??**).

Proposition 5 (Closed form solutions, [14]) *Let* H *be the quadratic Hamiltonian on* $L^2(\mathbb{R}^m) \otimes L^2(\mathbb{R}^m)$ *defined in* (8)*. If* γ_{β_1} *and* γ_{β_2} *are thermal states with covariance matrices, respectively, given by* $\nu_1 \mathbb{1}$ *and* $\nu_2 \mathbb{1}$ *such that* $1/2 \leq \nu_1 \leq \nu_2$*, then the QOT problem* (1) *is given by [14]*

$$\mathfrak{P}^Q_{\mathbf{H}}(\gamma_{\beta_1}, \gamma_{\beta_2}) = \sqrt{m}\left(\sqrt{\nu_2 + \frac{1}{2}} - \sqrt{\nu_1 - \frac{1}{2}}\right).$$

S8 Multi-marginal Quantum Optimal Transport with Random Hamiltonian

In this final set of experiments, we also consider N Hilbert spaces $\mathbb{C}^2$, so that the Hilbert space of the composite system is defined on $(\mathbb{C}^2)^{\otimes N}$. The Hamiltonian is drawn from a normal distribution, i.e. it is constructed by sampling 2^{2N} independent standard normal variables, reshaping them into a $2^N \times 2^N$ matrix, and symmetrizing to ensure Hermiticity.

As in the previous section, we fix $N = 12$ and define the marginals using the low-energy eigenvectors of the random Hamiltonian. Let $\psi_1, \ldots, \psi_l$, with $l \leq 2^N$, denote the eigenvectors corresponding to the first l eigenvalues in ascending order. The test cases are as follows and reported in Table S2 below and on the convergence plots in the Figure S4.

- **RG**: The marginals $\gamma_k = \mathrm{Tr}^k\left[|\psi_1\rangle\langle\psi_1|\right]$, for all $k \in [N]$.
- **RE**: The marginals $\gamma_k = \mathrm{Tr}^k\left[|\psi_2\rangle\langle\psi_2|\right]$, for all $k \in [N]$.
- **RM**: The marginals $\gamma_k = \mathrm{Tr}^k\left[\frac{1}{2}|\psi_1\rangle\langle\psi_1| + \frac{1}{2}|\psi_2\rangle\langle\psi_2|\right]$, for all $k \in [N]$.
- **RT**: $\gamma_k = e^{-\frac{H_{\mathrm{ho}}}{k}}/Z_k$ is a thermal state at temperature $k \in [N]$, see equation (S15).

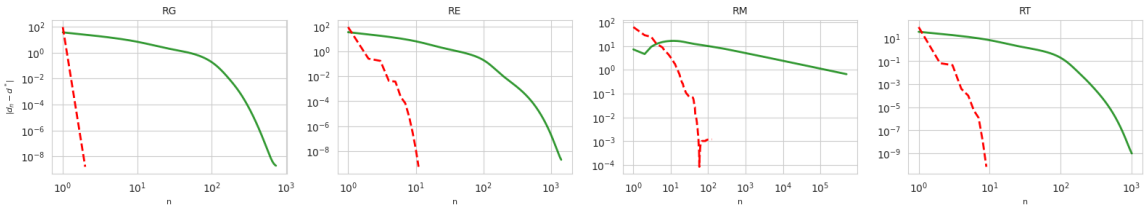

Figure S4: Convergence of the optimality gap $|d_n - d^*|$ versus number of iteration for the subgradient method (green line) and L-BFGS for the tests in Section S8. The variable d_n denotes the objective value at iteration n for each algorithm, d^* is the ground-truth optimal value computed via using MOSEK solver. The regularization parameter $\varepsilon = 10^{-12}$.

Test	Rank	Tol.	M	Time	Iters	Ach.	Err.
RG	1	1e-3	S	0.237	283	Yes	1.8e-04
		1e-4	S	0.333	414	Yes	2.4e-06
		1e-5	S	0.437	564	Yes	3.2e-08
		1e-6	S	0.560	734	Yes	2.0e-09
		1e-3	L	1.286	2	Yes	1.6e-09
		1e-4	L	1.337	2	Yes	1.6e-09
		1e-5	L	1.293	2	Yes	1.6e-09
		1e-6	L	1.321	2	Yes	1.6e-09
RE	1	1e-3	S	0.270	337	Yes	1.0e-03
		1e-4	S	0.522	692	Yes	9.8e-06
		1e-5	S	0.773	1051	Yes	9.8e-08
		1e-6	S	1.019	1409	Yes	2.0e-09
		1e-3	L	1.601	6	Yes	2.2e-04
		1e-4	L	1.848	8	Yes	5.4e-06
		1e-5	L	1.965	9	Yes	2.9e-07
		1e-6	L	2.267	11	Yes	5.3e-10

Test	Rank	Tol.	M	Time	Iters	Ach.	Err.
RT	1	1e-3	S	153.545	218685	Yes	8.6e-01
		1e-4	S	350.928	500000	No	6.6e-01
		1e-5	S	349.184	500000	No	6.6e-01
		1e-6	S	349.416	500000	No	6.6e-01
		1e-3	L	9.076	60	Yes	3.0e-04
		1e-4	L	10.444	69	Yes	1.0e-03
		1e-5	L	11.005	72	Yes	1.0e-03
		1e-6	L	27.138	125	Yes	1.2e-03
RM	1	1e-3	S	0.250	299	Yes	4.2e-04
		1e-4	S	0.400	510	Yes	6.2e-06
		1e-5	S	0.575	759	Yes	6.2e-08
		1e-6	S	0.749	1009	Yes	9.8e-10
		1e-3	L	1.263	4	Yes	3.9e-04
		1e-4	L	1.584	6	Yes	5.2e-06
		1e-5	L	1.713	7	Yes	1.2e-06
		1e-6	L	1.973	9	Yes	6.8e-11

Table S2: Comparison of L-BFGS and subgradient method for randomly generated Hamiltonian. The columns are as follows: **Test** – identifier of the test instance; **Rank** – rank of the QOT coupling in 1 returned by the ground-truth solution (MOSEK [24]); **Tol.** – (sub-)gradient norm threshold; **M** – method used (`S` for subgradient, `L` for L-BFGS); **Time** – wall-clock runtime in seconds; **Iters** – number of iterations; **Ach.** – whether the method met the target tolerance; **Err.** – final absolute difference between the objective value and the ground truth.