Classical and Quantum Optimal Transport problems for Matrices and Tensors

Shmuel Friedland
samuelfriedland in Zoom
Univ. Illinois at Chicago

IPAM workshop IV:
Efficient Tensor Representations for Learning and Computational Complexity
June 8, 2021

Outline

- Matrix Optimal Transport
- Tensor Optimal Transport
- Entropic relaxation of Optimal Transport
- Diagonal scaling of tensors
- Minimization of convex functions
- Matrix Quantum Optimal Transport
- Metrics on density matrices
- The dual problem
- Quantum Optimal Transport for tensors

Part I: Classical Discrete Optimal Transport

$\Pi_n \supset \Pi_n^o$ convex set of probability vectors and its interior in $\mathbb{R}_+^n$

Problem: Given $\mathbf{p}^A, \mathbf{p}^B \in \Pi_n$ define distance $\mathrm{dist}(\mathbf{p}^A, \mathbf{p}^B)$

Many applications:

Comparisons of two histograms

Earth mover's distance: Two different ways of piling up a certain amount of dirt over the region D

Gaspard Monge 1781, in the context of transportation theory

Matrix Optimal Transport

For $\mathbf{p}^A \in \Pi_m, \mathbf{p}^B \in \Pi_n$ let $\Gamma_{cl}(\mathbf{p}^A, \mathbf{p}^B) \subset \mathbb{R}_+^{m\times n}$

the set of coupling matrices $R \in \mathbb{R}_+^{m\times n}$: $R\mathbf{1} = \mathbf{p}^A, R^\top \mathbf{1} = \mathbf{p}^B$

Note: $\mathbf{p}^A(\mathbf{p}^B)^\top \in \Gamma_{cl}(\mathbf{p}^A, \mathbf{p}^B)$

MOT: $\mathrm{T}_C(\mathbf{p}^A, \mathbf{p}^B) = \min\{\mathrm{Tr}\, C^\top R, R \in \Gamma_{cl}(\mathbf{p}^A, \mathbf{p}^B)\}$ for $C \in \mathbb{R}^{m\times n}$

Linear Programming (LP) problem, slow to be solved for big m, n

For $m = n$ and $C = [c_{ij}] \in \mathbb{R}_+^{n\times n}$ distance matrix:

C symmetric, $c_{ii} = 0, 0 < c_{ij} \le c_{ik} + c_{kj}$ for $i \ne j$

$\mathrm{T}_C(\mathbf{p}^A, \mathbf{p}^B)$ is a distance on Π_n

For $t > 0$ $C^{\circ t} = [c_{ij}^t]$

Wasserstein t-distance: $\left(\mathrm{T}_{C^{\circ t}}(\mathbf{p}^A, \mathbf{p}^B)\right)^{1/t}$ for $t \ge 1$

Tensor Optimal Transport I

$\mathbf{p}_j \in \Pi_{n_j}, j \in [d], \mathcal{C} \in \otimes_{j=1}^d \mathbb{R}^{n_j}$ are given

$\Gamma_{cl}(\mathbf{p}_1, \ldots, \mathbf{p}_d) = \{\mathcal{U} \in \otimes_{j=1}^d \mathbb{R}_+^{n_j}, \mathcal{U} \times_k \otimes_{j\in[d]\setminus\{k\}} \mathbf{1}_{n_j} = \mathbf{p}_k, k \in [d]\}$

Note $\otimes_{j=1}^d \mathbf{p}_j \in \Gamma_{cl}(\mathbf{p}_1, \ldots, \mathbf{p}_d)$

$\mathrm{T}_{\mathcal{C}}(\mathbf{p}_1, \ldots, \mathbf{p}_d) := \min\{\langle \mathcal{C}, \mathcal{U}\rangle, \mathcal{U} \in \Gamma_{cl}(\mathbf{p}_1, \ldots, \mathbf{p}_d)\}$

Tensor Optimal Transport II

Assume $d = 2l, n_{l+i} = n_i, i \in [l]$

$\mathcal{C}$ distance tensor: bisymmetric $c_{i_1,\ldots,i_l,i_{l+1},\ldots i_{2l}} = c_{i_{l+1},\ldots,i_{2l},i_1,\ldots i_l}$

Matrix $D(\mathcal{C}) = [d_{(i_1,\ldots,i_l),(i_{l+1},\ldots,i_{2l})}] \in \mathbb{R}_+^{N\times N}$, is a distance matrix

$d_{(i_1,\ldots,i_l),(i_{l+1},\ldots,i_{2l})} = c_{i_1,\ldots,i_{2l}}$, $N = \prod_{j=1}^{l} n_j$

Example: $D(\mathcal{C})$ - uniform: off-diagonal entries are $t > 0$

Thm: $\mathrm{T}_{\mathcal{C}}(\mathbf{p}_1, \ldots, \mathbf{p}_{2l})$ metric on $(\prod_{j=1}^{l} \Pi_{n_j}) \times (\prod_{j=1}^{l} \Pi_{n_j})$

Use the matrix case with $D(\mathcal{C})$ and extra conditions

Tensor Optimal Transport III

Assume $n_i = n, i \in [2l]$ and $D(\mathcal{C})$ a distance matrix

$\mathrm{T}_{\mathcal{C}}(\mathbf{p}_1, \ldots, \mathbf{p}_{2l})$ distance between ordered sets

$(\mathbf{p}_1, \ldots, \mathbf{p}_l), (\mathbf{p}_{l+1}, \ldots, \mathbf{p}_{2l}) \in (\Pi_n)^l$

Need conditions for this metric to be metric on sets

$\{\mathbf{p}_1, \ldots, \mathbf{p}_l\}, \{\mathbf{p}_{l+1}, \ldots, \mathbf{p}_{2l}\}$

1. $\mathrm{dist}((\mathbf{p}_1, \ldots, \mathbf{p}_l), (\mathbf{p}_{l+1}, \ldots, \mathbf{p}_{2l})) =$
$\min\{\mathrm{T}_{\mathcal{C}}(\mathbf{p}_{\sigma(1)}, \ldots, \mathbf{p}_{\sigma(l)}, \mathbf{p}_{l+1}, \ldots, \mathbf{p}_{2l})\}$

$\sigma \in \Sigma_l$-all permutations of $[l]$

2. Symmetrize $D(\mathcal{C})$ on $[l], [2l] \setminus \{[l]\}$ and declare

$\mathrm{dist}((\mathbf{p}_1, \ldots, \mathbf{p}_l), (\mathbf{p}_1, \ldots, \mathbf{p}_l)) = 0$ to obtain noncontinuous distance

Entropic relaxation of TOT I

Linear programming problems are efficiently solved by interior methods

by adding a strict convex term to minimize the linear function

For Matrix Optimal Transport was done by Marco Cuturi 2013

For $\mathbf{x} \in \Pi_n$, $H(\mathbf{x}) = -\sum_{i=1}^n x_i \log x_i$ stricly concave function

$0 \le H(\mathbf{x}) \le \log n$, equality holds for uniform probability vector

For $\mathcal{U} \in \mathbb{R}_+^{\mathbf{n}}$ probability tensor: $\langle \mathcal{U}, \otimes_{j=1}^d \mathbf{1}_{n_j} \rangle = 1$:

$$H(\mathcal{U}) = -\sum_{i_1,\ldots,i_d} u_{i_1,\ldots,i_d} \log u_{i_1,\ldots,i_d} \in [0, \sum_{j=1}^d \log n_j]$$

$$f_\lambda(\mathcal{U}) = \langle \mathcal{C}, \mathcal{U} \rangle - \frac{1}{\lambda} H(\mathcal{U}),\ \lambda > 0$$

$$f_\lambda(\mathcal{U}) \le \langle \mathcal{C}, \mathcal{U} \rangle \le f_\lambda(\mathcal{U}) + \frac{\sum_{j=1}^d \log n_j}{\lambda}$$

Entropic relaxation of TOT II

$\exp(\mathcal{T}) := [\exp(t_{i_1,\ldots,i_d})] \in \otimes_{j=1}^{d} \mathbb{R}_+^{n_j} = \mathbb{R}_+^{\mathbf{n}}$

$min\{\langle \mathcal{C}, \mathcal{U} \rangle - \frac{1}{\lambda} H(\mathcal{U}), U \in \Gamma_{cl}(\mathbf{p}_1, \ldots, \mathbf{p}_d)\} = \langle \mathcal{C}, \mathcal{U}_\lambda \rangle - \frac{1}{\lambda} H(\mathcal{U}_\lambda)$

For $\lambda \gg 1$ $\mathcal{U}_\lambda$ a good approximation to optimal solution for TOT

Theorem I: $\mathcal{U}_\lambda$ is unique diagonally equivalent tensor to $\exp(-\lambda\mathcal{C})$

Cuturi 2013 for matrices

Diagonal scaling of tensors I

$\mathcal{T} = [t_{i_1,\ldots,i_d}] \in \mathbb{R}_+^{\mathbf{n}}$ is diagonally scaled to a tensor in $\Gamma_{cl}(\mathbf{p}_1, \ldots, \mathbf{p}_d)$:

$\mathcal{T}' = \mathcal{T}(\mathbf{x}) := [\exp(x_{i_1,1} + \cdots + x_{i_d,d}) t_{i_1,\ldots,i_d}] \in \Gamma_{cl}(\mathbf{p}_1, \ldots, \mathbf{p}_d)$

for some $\mathbf{x}_j \in \mathbb{R}^{n_j}, j \in [d]$

For matrices $T' = D_1 T D_2$ D_1, D_2 diagonal with positive diagonal entries

Existence theorem: Diagonal scaling of $\mathcal{T} \in \mathbb{R}_+^{\mathbf{n}}$ exists iff

$\Gamma_{cl}(\mathbf{p}_1, \ldots, \mathbf{p}_d)$ contains a tensor with the same zero pattern as $\mathcal{T}$

(It is enough to consider the case $\mathbf{p}_1, \ldots, \mathbf{p}_d > \mathbf{0}$)

Sinkhorn 1964 positive matrices and uniform distribution- doub. stoch.

Menon 1968 for matrices, Bapat 1982, Raghavan 1984 for pos. ten.

Bapat-Raghavan, Franklin-Lorentz 1989, Friedland 2011

Diagonal scaling of tensors II

Franklin-Lorentz: $\Gamma_{cl}(\mathbf{p}_1,\ldots,\mathbf{p}_d,\mathcal{T})=\{u_{i_1,\ldots,i_d}=0 \text{ if } t_{i_1,\ldots,i_d}=0\}$

$g(\mathcal{U}):=\sum_{i_j\in[n_j]} u_{i_1,\ldots,i_d}\log\frac{u_{i_1,\ldots,i_d}}{t_{i_1,\ldots,i_d}}, \mathcal{U}\in\Gamma_{cl}(\mathbf{p}_1,\ldots,\mathbf{p}_d)<\infty$

iff $\mathcal{U}\in\Gamma_{cl}(\mathbf{p}_1,\ldots,\mathbf{p}_d,\mathcal{T})$. Then $\min\{g(\mathcal{U}),\mathcal{U}\in\Gamma_{cl}(\mathbf{p}_1,\ldots,\mathbf{p}_d,\mathcal{T})\}$

achieved at unique $\mathcal{U}(\mathcal{T})$ diagonally scallable to $\mathcal{T}$

Sinkhorn scaling algorithm

Assume $\mathbf{0} < \mathbf{p}^A \in \Pi_m, \mathbf{0} < \mathbf{p}^B \in \Pi_n$

For $R \in \mathbb{R}_{++}^{m\times n}$: set $R_0 = R$

$R_{2k+1} = D_{2k}R_{2k}, \; D_{2k} = \mathrm{diag}(d_{1,2k}, \ldots, d_{m,2k})$

$d_{i,2k} = \frac{p_i}{\sum_{j=1}^n a_{i,j,2k}}, i \in [m] \Rightarrow R_{2k+1}\mathbf{1}_n = \mathbf{p} > \mathbf{0}$

$R_{2k} = R_{2k-1}D_{2k-1}, \; D_{2k-1} = \mathrm{diag}(d_{1,2k-1}, \ldots, d_{n,2k-1})$

$d_{j,2k-1} = \frac{q_j}{\sum_{i=1}^m a_{i,j,2k-1}}, j \in [n] \Rightarrow R_{2k+1}^\top \mathbf{1}_m = \mathbf{q} > \mathbf{0}$

Theorem II: $R_k, k \in \mathbb{N}$ converges geometrically to $U \in \Gamma_{cl}(\mathbf{p}, \mathbf{q})$

$\|R_k - U\| \le Kt^k, k \in \mathbb{N}$ for some $K > 0, t \in (0, 1)$

Sinkhorn proved for $m = n, \mathbf{p}^A = \mathbf{p}^B = (1/n)\mathbf{1}_n$

Minimization of convex functions

Lemma: $f \in \mathrm{C}^2(\mathbb{R}^n)$ with positive definite Hessian $H(f)$ everywhere

Then (1) $\lim_{\|\mathbf{x}\| \to \infty} f(\mathbf{x}) = \infty$ iff f has a unique critical point

Another convex function related to scaling I

$\mathbf{x}_i = (x_{1,i}, \ldots, x_{n_i,i})^\top \in \mathbb{R}^{n_i}, \mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_d) \in \mathbb{R}^N = \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_d}$

$h(\mathbf{x}) = \sum_{i_j \in [n_j], j \in [d]} t_{i_1,\ldots,i_d} \exp(\sum_{j=1}^d x_{i_j,j}), \quad \mathcal{T} \gneq 0$

h convex and constant on an affine space:

(1) $\sum_{j=1}^d x_{i_j,j} = a_{i_1,\cdots,i_d}$ if $t_{i_1,\ldots,i_d} > 0$ for all $(i_1, \ldots, i_d)$

Example: (2) $\mathbf{x}_i = t_i \mathbf{1}_{n_i}, i \in [d], \sum_{i=1}^d t_i = 0$ (or constant)

Claim: If $\mathcal{T} > 0$ ker $H(h)(\mathbf{x})$ is (2)

Another convex function related to scaling II

$\mathcal{T}(\mathbf{x}) = [\exp(\sum_{j=1}^{d} x_{i_j,j}) t_{i_1,\ldots,i_d}]$, $h(\mathbf{x}) = \langle \mathcal{T}(\mathbf{x}), \otimes_{j=1}^{d} \mathbf{1}_{n_j} \rangle$

For $\mathbf{p} \in \mathbb{R}^n$: $\mathrm{L}(\mathbf{p}) = \{\mathbf{x} \in \mathbb{R}^n, \sum_{j=1}^{n} p_j x_j = 0\}$

For $\mathcal{T} > 0, \mathbf{p}_i \in \Pi_{n_i}^{o}, i \in [d]$ h is strictly convex on $\mathrm{L} := \prod_{i=1}^{n} \mathrm{L}(\mathbf{p}_i) \equiv \mathbb{R}^{N'}$

$\lim_{\|\mathbf{x}\| \to \infty, \mathbf{x} \in \mathrm{L}} h(\mathbf{x}) = \infty$, equivalently, h has a critical point in L

Advantage over Franklin-Lorentz: Minimizing over much smaller space

Partial minimization algorithm

$\mathbb{R}^N = \mathbb{R}^{N_1} \times \cdots \times \mathbb{R}^{N_d}$, $\mathbf{x} = (\mathbf{x}_j, \mathbf{x}^j) \in \mathbb{R}^N$, $\nabla\phi = (\nabla_{\mathbf{x}_1}\phi, \ldots, \nabla_{\mathbf{x}_d}\phi)$

$\phi : \mathrm{C}^2(\mathbb{R}^N)$, $H(\phi)(\mathbf{x}) \succ 0$, $\lim_{\|\mathbf{x}\|\to\infty} \phi(\mathbf{x}) = \infty$

partial minimization: $\min\{\phi(\mathbf{x}_j, \mathbf{x}^j), \mathbf{x}_j \in \mathbb{R}^{N_j}\} = \phi(\mathbf{x}_j(\mathbf{x}^j), \mathbf{x}^j)$

$\|\mathbf{x}\|_s := \left(\sum_{i=1}^N |x_i|^s\right)^{1/s}$, $s \in [1, \infty]$, Fix $s \in [1, 2]$

PMThm: Algorithm below converges geomet. to min ϕ & min point:

Start with any $\mathbf{x}^{(0)} \in \mathbb{R}^{\mathbb{N}}$

If $\max\{\|\nabla_{\mathbf{x}_j}(\mathbf{x}^{(l)})\|_s, j \in [d]\} = 0$ $\mathbf{x}^{(l)}$ min point, exit

At $\mathbf{x}^{(l)}$ set $k(l) = \arg\max\{\|\nabla_{\mathbf{x}_j}(\mathbf{x}^{(l)})\|_s, j \in [d]\}$, $\mathbf{x}^{(l)} = (\mathbf{x}^{(l)}_{k(l)}, \mathbf{y}^{(l)})$

$\mathbf{x}^{(l+1)} = (\mathbf{x}^{(l+1)}_{k(l)}, \mathbf{y}^{(l)}))$, $\mathbf{x}^{(l+1)}_{k(l)} = \mathbf{x}_{k(l)}(\mathbf{y}_l) - \mathbf{x}^{(l)}_{k(l)})$

Note: $\nabla_{\mathbf{x}_{k(l-1)}}(\mathbf{x}^{(l)}) = \mathbf{0}$ for $l \geq 1$

Sinkhorn algorithm is partial minimization algorithm

$\mathcal{T} > 0$, $\mathbb{R}^{N'} = \mathrm{L} = \mathrm{L}(\mathbf{p}_1) \times \cdots \times \mathrm{L}(\mathbf{p}_d)$

Claim: For $\mathbf{x} = (\mathbf{x}_k, \mathbf{x}^k)$: $\mathbf{x}_k(\mathbf{x}^k) = \mathbf{z}_k = (z_1, \ldots, z_{n_k}) \in \mathrm{L}(\mathbf{p}_j)$

is essentially given by Sinkhorn scaling:

Part II: Matrix Quantum Optimal Transport

$\mathrm{H}_n \supset \mathrm{H}_{n,+} \supset \Omega_n$ hermitian, positive semidefinite, density matrices

For $\rho^A \in \Omega_m, \rho^B \in \Omega_n$: $\Gamma(\rho^A, \rho^B) = \{R \in \Omega_{mn}, \mathrm{Tr}_B R = \rho^A, \mathrm{Tr}_A R = \rho^B\}$

$R = [r_{(i,p)(j,q)}], i, j \in [m], p, q \in [n]$, $\mathrm{Tr}_B R = [\sum_{p=1}^n r_{(i,p)(j,p)}]$

$\rho^A \otimes \rho^B \in \Gamma(\rho^A, \rho^B)$

MQOT: $\mathrm{T}_C^Q(\rho^A, \rho^B) = \min\{\mathrm{Tr}\, CR, R \in \Gamma(\rho^A, \rho^B)\}$ for $C \in \mathrm{H}_{mn}$

If $0 \preceq C \preceq \mathbb{I}_{mn}$ then $\mathrm{T}_C^Q(\rho^A, \rho^B)$

is the minimum expected value of the observable C given ρ^A, ρ^B

$\mathrm{T}_C^Q(\rho^A, \rho^B)$ is semidefinite problem that can be computed

with precision $\varepsilon > 0$ in polynomial time in data and $\log(1/\varepsilon)$

Connection between OT and QOT

For $\mathbf{x} = (x_1, \ldots, x_n)^\top, X \in \mathbb{R}^{n\times n}$, $\mathrm{diag}(\mathbf{x}) = \mathrm{diag}(x_1, \ldots, x_n) \in \mathbb{R}^{n\times n}$, and

$\mathrm{diag}(X) \in \mathbb{R}^{n\times n}$ is a diagonal matrix with diagonal entries of X

For $\rho \in \Omega_n$: $\mathrm{diag}(\rho) = \mathrm{diag}(\mathbf{p}) \in \Omega_n, \mathbf{p} \in \Pi_n$ is decoherence of ρ

$\Gamma_{cl}(\mathbf{p}^A, \mathbf{p}^B)$ isomorphic to $\mathrm{diag}(\Gamma(\mathrm{diag}(\mathbf{p}^A), \mathrm{diag}(\mathbf{p}^B))$:

$[r_{ij}]$ corresponds to $\mathrm{diag}(r_{11}, \ldots, r_{mn})$

$\mathrm{T}_C(\mathbf{p}^A, \mathbf{p}^B) \ge \mathrm{T}^Q_{\tilde{C}}(\mathrm{diag}(\mathbf{p}^A), \mathrm{diag}(\mathbf{p}^B))$ if

$\mathrm{diag}(\tilde{C}) = \mathrm{diag}(c_{11}, \ldots, c_{mn})$

Classical OT greater than QOT

Semi-distances and weak distances

sd: $\mathbf{X} \times \mathbf{X} \to [0, \infty)$ is semi-distance if

(a) $\mathrm{sd}(x, y) = \mathrm{sd}(y, x)$ (symmetric)

(b) $\mathrm{sd}(x, y) \geq 0$ and equality holds iff $x = y$ (positivity)

sd is a metric if it satisfies the triangle inequality:

$\mathrm{sd}(x, y) \leq \mathrm{sd}(x, z) + \mathrm{sd}(y, z)$

semi-distance is a weak distance if $\exists$ metric d on $\mathbf{X}$ satisfying

(1) $\mathrm{sd}(x, y) \geq \mathrm{d}(x, y)$

THM: For a given weak sd on $\mathbf{X}$ $\exists!$ maximizing metric D satisfying (1):

$\mathrm{D}(x, y) = \lim_{N \to \infty} \inf_{x_0 = x, x_N = y} \sum_{i=1}^{N} \mathrm{sd}(x_{i-1}, x_i)$

Special metrics on Hermitian matrices

Denote by $\mathrm{U}(n) \subset \mathbb{C}^{n\times n}$ the group of unitary matrices

$d(\rho^A, \rho^B) = \max\{\big(|U(\rho^A - \rho^B)U^*)_{11}|U \in \mathrm{U}(n)\}$ metric on H_n

Can be generalized

2-Wasserstein metric induced by QOT

THM 1: $\mathrm{T}_C^Q(\rho^A, \rho^B)$ induces semi-distance on Ω_n if and only if C is positive semidefinite and vanishes on symmetric matrices $\mathcal{H}_S$

If $0 \neq C \succeq 0, \ker C = \mathcal{H}_S$ then $\sqrt{\mathrm{T}_C^Q(\rho^A, \rho^B)}$ is a weak distance inducing 2-Wasserstein metric $W_C^Q(\rho^A, \rho^B)$

THM 2: Assume that C^Q is projection on skew symmeric matrices $\mathcal{H}_A$:

(a) $\sqrt{\mathrm{T}_{C^Q}^Q(\rho^A, \rho^B)}$ is a metric on pure states $(\mathrm{rank}\ \rho = 1)$

(b) $\sqrt{\mathrm{T}_{C^Q}^Q(\rho^A, \rho^B)}$ is a metric on Ω_2 (qubits)

(c) $\left(\mathrm{T}_{C^Q}^Q(\rho^A, \rho^B)\right)^{1/t}$ is not a metric for $t \in [1, 2)$

Open Problem

Let $S:\mathcal{H}_n\otimes\mathcal{H}_n\to\mathcal{H}_n\otimes\mathcal{H}_n$ SWAP operator $\mathbf{x}\otimes\mathbf{y}\to\mathbf{y}\otimes\mathbf{x}$ $(X\to X^\top)$

$C^Q=\frac{1}{2}(\mathbb{I}-S)$ projection on $\mathcal{H}_A$

Open Problem: Is $\sqrt{\mathrm{T}^Q_{C^Q}(\rho^A,\rho^B)}$ is a metric on Ω_n?

Numerical simulations for $n=3,4$ yes!

The dual problem

$\mathrm{T}_C^Q(\rho^A,\rho^B) = \sup\{\mathrm{Tr}\,(\sigma^A\rho^A + \sigma^B\rho^B), F = C - \sigma^A\otimes\mathbb{I}_n - \mathbb{I}_m\otimes\sigma^B \succeq 0\}$

$\mathrm{Tr}\, FR = 0$ for $R\in\Gamma(\rho^A,\rho^B) \iff R$ minimizing and F maximizing resp.

If ρ^A,ρ^B positive definite, maximizing F exists

THM: $\mathrm{T}_{C^Q}(\rho^A,\rho^B) \ge \frac{1}{2}d(\rho^A,\rho^B)^2$

For qubits ($n=2$) equality holds

Outline of the proof:

(a) $\mathrm{T}_{C^Q}(\rho^A,\rho^B) = \mathrm{T}_{C^Q}(U\rho^A U^*, U\rho^B U^*)$

(b) $\mathrm{T}_{C^Q}(\rho^A,\rho^B) \ge \mathrm{T}_{C^Q}(\mathrm{diag}(\rho^A),\mathrm{diag}(\rho^B))$

(c) lower bounds using the dual version for diagonal density matrices

QOT for tensors (multipartite states)

$\mathcal{H}_{n_k}$ - n_k-dimensional Hilbert space

$\mathrm{S}_+(\otimes_{j=1}^d \mathcal{H}_{n_j})$- self adjoint positive semidefinite operators

For $\rho^{A_j} \in \Omega_{n_j}$, $j \in [d] = \{1, \dots, d\}$, $d \geq 3$

$R \in \mathrm{S}_+(\otimes_{j=1}^d \mathcal{H}_{n_j})$, $\mathrm{Tr}_k\, R \in \mathrm{S}_+(\mathcal{H}_{n_k})$

$\Gamma(\rho^{A_1}, \dots, \rho^{A_d}) = \{R \in \mathrm{B}_+(\otimes_{j=1}^d \mathcal{H}_{n_j}), \mathrm{Tr}_k\, R = \rho^{A_k}, k \in [d]\}$

$\mathrm{T}_C^Q(\rho^{A_1}, \dots, \rho^{A_d}) = \min\{\mathrm{Tr}\, CR, R \in \Gamma(\rho^{A_1}, \dots, \rho^{A_d})\}$

is a continuous convex function on $\Omega_{n_1} \times \cdots \times \Omega_{n_d}$

Computation of QOT is an Semidefinite Problem (SDP)

Equidimensional case

$n_1 = \cdots = n_d = n$, $\mathrm{S}^d\mathcal{H}_n$-symmetric tensors (bosons)

C^Q projection on the othogonal complement of symmetric tensors

Claim: $\mathrm{T}^Q_{C^Q}(\rho^{A_1}, \ldots, \rho^{A_k}) \geq 0$ equality hods iff $\rho^{A_1} = \cdots = \rho^{A_k}$

Relation to permanents

$\rho^{A_j} = \sum_{i=1}^{n} \lambda_{i,j} \mathbf{x}_{i,j} \mathbf{x}_{i,j}^*$ for $j \in [d]$

$G(\mathbf{y}_1, \ldots, \mathbf{y}_d) = [\mathbf{y}_i^* \mathbf{y}_j] \in \mathrm{H}_d$-Grammian of $\mathbf{y}_1, \ldots, \mathbf{y}_d \in \mathbb{C}^n$

$\mathrm{Tr}\, C^Q(\otimes_{j=1}^{d} \rho^{A_j}) = 1 - \frac{1}{d!} \sum_{i_1,\ldots,i_d \in [n]} \left(\prod_{j \in [d]} \lambda_{i_j,j} \right) \mathrm{perm}\, G(\mathbf{x}_{i_1,1}, \ldots, \mathbf{x}_{i_d,d})$

$\mathrm{T}_C^Q(\rho^{A_1}, \ldots, \rho^{A_d}) \le$

$1 - \frac{1}{d!} \sum_{i_1,\ldots,i_d \in [n]} \left(\prod_{j \in [d]} \lambda_{i_j,j} \right) \mathrm{perm}\, G(\mathbf{x}_{i_1,1}, \ldots, \mathbf{x}_{i_d,d})$

References 1

J. Altschuler, J. Weed and P. Rigollet, Near-linear time approximation algorithms for optimal transport via Sinkhorn iteration, NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems, December 2017, Pages 1961-1971.

R.B. Bapat D_1AD_2 theorems for multidimensional matrices, *Linear Algebra Appl.* 48 (1982), 437–442.

R.B. Bapat and T.E.S. Raghavan, An extension of a theorem of Darroch and Ratcliff in loglinear models and its application to scaling multidimensional matrices, *Linear Algebra Appl.* 114/115 (1989), 705-715.

References 2

- M. Cuturi, Sinkhorn distances: Lightspeed computation of optimal transport, In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, Advances in Neural Information Processing Systems 26, 2292-2300. Curran Associates, Inc., 2013.
- J. Franklin and J. Lorenz, On the scaling of multidimensional matrices, *Linear Algebra Appl.* 114/115 (1989), 717-735.
- S. Friedland, Positive diagonal scaling of a nonnegative tensor to one with prescribed slice sums, *Linear Algebra Appl.*, 434 (2011), 1615-1619.

References 3

- Sam Cole, Michał Eckstein, Shmuel Friedland, Karol Życzkowski, Quantum Optimal Transport, arXiv:2105.06922, May, 2021.
- Shmuel Friedland, Michał Eckstein, Sam Cole, Karol Życzkowski, Quantum Monge-Kantorovich problem and transport distance between density matrices, arXiv:2102.07787, 2021.
- Shmuel Friedland, Tensor optimal transport, distance between sets of measures and tensor scaling, arXiv:2005.00945, 2020.

References 4

- L.V Kantorovich, On the translocation of masses. Dokl. Akad. Nauk. USSR 37 (1942), 199-201. English translation in J. Math. Sci. 133, 4 (2006), 1381-1382.
- M.V. Menon, Matrix links, an extremisation problem and the reduction of a nonnegative matrix to one with with prescribed row and column sums, *Canad. J. Math* 20 (1968), 225-232.
- T.E.S. Raghavan, On pairs of multidimensional matrices, *Linear Algebra Appl.* 62 (1984), 263-268.
- R.A. Sinkhorn, A relationship between arbitrary positive matrices and doubly stochastic matrices, *Ann. Math. Statist.* 35 (1964), 876–879.
- C. Villani, Optimal transport, Old and new, Grundlehren der Mathematischen Wissenschaften, 338, Springer-Verlag, Berlin, 2009.