Strictly-correlated Electron Functional and Multimarginal Optimal Transport

Lexing Ying
Stanford University

Joint work with Yuehaw Khoo, Lin Lin, and Michael Lindsey
April 2021, IPAM, UCLA

Outline

- Multimarginal optimal transport
- Strictly-correlated Electrons
- Convex relaxation approach

Optimal transport

- Given probability distributions ρ_1, ρ_2 on X
- Cost function $c(x, y)$
- Optimal transport problem

$$\inf_{\mu \in \Pi(\rho_1, \rho_2)} \int_X \int_X c(x, y) d\mu(x, y)$$

  where $\Pi(\rho_1, \rho_2)$ is the set of joint distributions on $X \times X$ with marginals ρ_1, ρ_2.
- Applications
  - Operational research, ...
  - Generative adversarial network (GAN), ...

Multi-marginal optimal transport

- Given probability distributions $\rho_1, \ldots, \rho_N$ on X
- Cost function $C(x_1, \ldots, x_N)$
- Multimarginal OT problem

$$\inf_{\mu \in \Pi(\rho_1, \ldots, \rho_N)} \int_{X \times \cdots \times X} C(x_1, \ldots, x_N) d\mu(x_1, \ldots, x_N)$$

  with $\Pi(\rho_1, \ldots, \rho_N)$ the set of distributions on $X \times \cdots \times X$ with marginals $\rho_1, \ldots, \rho_N$
- We are concerned with an example from many-body Schrödinger equation

Outline

- Multimarginal optimal transport
- Strictly-correlated Electrons
- Convex relaxation approach

Schrödinger equation

- Many-body Schrödinger equation

$$H\Psi \equiv \left(\sum_{i=1}^{N} -\Delta_{x_i} + \sum_{i<j} \frac{1}{|x_i - x_j|} + \sum_i v_{\mathsf{ext}}(x_i) \right) \Psi \equiv (K+C+V_{\mathsf{ext}})\Psi$$

with $\Psi = \Psi(x_1, \ldots, x_N)$ antisymmetric and $\|\Psi\|_{\mathcal{L}_2} = 1$.

- Ground state energy

$$\inf_{\Psi} \langle \Psi | H | \Psi \rangle = \inf_{\Psi} \langle \Psi | K + C + V_{\mathsf{ext}} | \Psi \rangle$$

- High dimensional problem for N large, hard to solve.

Density functional theory (DFT)

- Define density (1-marginal)

$$\rho(x) \equiv \int |\Psi(x, x_2, \dots, x_N)|^2 dx_2, \dots dx_N$$

- DFT uses nested optimization using ρ

$$\inf_{\rho} \inf_{\Psi \to \rho} \langle \Psi | (K + C) + V_{\text{ext}} | \Psi \rangle \equiv \inf_{\rho} F[\rho] + V_{\text{ext}}[\rho]$$

$$V_{\text{ext}}[\rho] \equiv \inf_{\Psi \to \rho} \langle \Psi | V_{\text{ext}} | \Psi \rangle = \inf_{\Psi \to \rho} \langle \Psi | \sum_i v_{\text{ext}}(x_i) | \Psi \rangle = N \int v_{\text{ext}}(x) \rho(x) dx$$

$$F[\rho] \equiv \inf_{\Psi \to \rho} \langle \Psi | K + C | \Psi \rangle$$

(unknown universal functional, Hohenberg-Kohn 64).

Approximating $F[\rho]$

- When K dominant, approximate $F[\rho]$ with Kohn-Sham functional
- When C dominant, ignore K and approximate $F[\rho]$ with strictly-correlated electron (SCE) functional (Seidl 99)

$$F[\rho] \approx V_{\text{ee}}^{\text{SCE}}[\rho] = \inf_{\Psi \to \rho} \int \sum_{i<j} \frac{1}{|x_i - x_j|} |\Psi(x_1, \ldots, x_N)|^2 dx_1 \ldots dx_N$$

$$= \inf_{\mu \in \Pi(\rho, \cdots, \rho)} \int \sum_{i<j} \frac{1}{|x_i - x_j|} \mu(x_1, \ldots, x_N) dx_1 \ldots dx_N$$

μ symmetric, $\int d\mu = 1, \mu \geq 0$, $\mu(\ldots, x_i, \ldots, x_j, \ldots) = 0$ if $x_i = x_j$.

- This is a special multimarginal OT problem
  - Same 1-marginal ρ for each dimension
  - Cost

$$C(x_1, \ldots, x_N) = \sum_{i<j} \frac{1}{|x_i - x_j|}$$

  - But the support of μ is singular

Breaking the complexity barrier

- Numerics-related previous work
  - (Mendl-Lin 12): Solve the dual problem of $V_{ee}^{\text{SCE}}[\rho]$ (exponential number of constraints)
  - (Benamou-Carlier-Nenna 16): Sinkhorn scaling (exponential number of variables)
  - (Friesecke-Vogler 18): Existence of sparse solution for multimarginal OT
- Q: Can we solve this multimarginal OT with polynomial complexity?
- Approach: use convex relaxation techniques to obtain useful lower and upper bounds for the SCE optimization problem
  - Yuehaw Khoo and Lexing Ying. Convex relaxation approaches for strictly correlated density functional theory. SIAM Journal on Scientific Computing 41-4, (2019).

Outline

- Multimarginal optimal transport
- Strictly-correlated Electrons
- Convex relaxation approach

Discretization

- In SCE $C(x_1, \ldots, x_N) = \sum_{i<j} \frac{1}{|x_i - x_j|}$, multimarginal OT with pairwise cost

$$\inf_{\mu \in \Pi(\rho, \cdots, \rho)} \int_{X \times \ldots \times X} \sum_{k<l} \frac{1}{|x_k - x_l|} d\mu(x_1, \ldots, x_N)$$

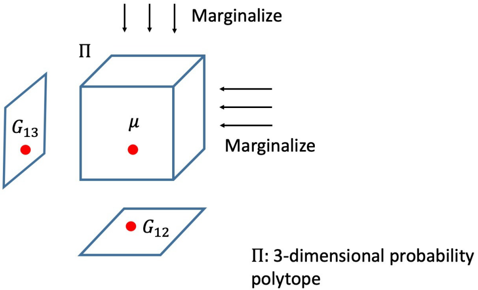
- Discretize X with a grid $\mathcal{L}$ of L pts and introduce $c(x, y) \equiv \frac{1}{|x-y|}$.
- From now on, focus on the discrete problem

$$\min_{\mu \in \Pi(\rho, \cdots, \rho)} \sum_{x_1, \ldots x_N \in \mathcal{L}} \left(\sum_{k<l} c(x_k, x_l) \right) \mu(x_1, \ldots, x_N)$$

Reducing dimensionality

$$\min_{\mu \in \Pi(\rho, \cdots, \rho)} \sum_{x_1, \ldots x_N \in \mathcal{L}} \left(\sum_{k<l} c(x_k, x_l) \right) \mu(x_1, \ldots, x_N)$$

- Rewrite the problem in terms of 2-marginals



In terms of 2-marginals

- Write μ as convex combination of extreme points:

$$\mu = \sum_{x_1,\dots x_N} \mu_{x_1,\dots,x_N} e_{x_1} \otimes \cdots \otimes e_{x_N}$$

- Let $\{e_x : x \in \mathcal{L}\}$ be the set of L canonical basis vectors.
- $\sum_{x_1,\dots x_N} \mu_{x_1,\dots,x_N} = 1$ and $\mu_{x_1,\dots,x_N} \geq 0$.
- The generalized 2-marginal of the (k,l) slice

$$G^{kl} = \sum_{x_1,\dots x_N} \mu_{x_1,\dots,x_N} e_{x_k} e_{x_l}^{\mathsf{T}} \in \mathbb{R}^{L\times L}$$

- Due to symmetries (identical electrons), all G^{kl} for $k \neq l$ are the same and all G^{kk} are the same. Thus define

$$\gamma \equiv G^{kl}, \quad \epsilon \equiv G^{kk}$$

Equivalent multimarginal OT form

- $\min_{\mu\in\Pi(\rho,\cdots,\rho)} \sum_{x_1,\dots x_N\in\mathcal{L}} \left(\sum_{k<l} c(x_k,x_l)\right)\mu(x_1,\dots,x_N)$

$$\sum_{k<l}\sum_{x_k,x_l} c(x_k,x_l) \sum_{\text{rest } x_i} \mu(x_1,\dots,x_N) \equiv \sum_{k<l}\sum_{x_k,x_l} c(x_k,x_l) G^{kl}(x_k,x_l)$$

- Introduce $G, C \in \mathbb{R}^{NL\times NL}$:

$$G := \begin{bmatrix} G^{11} & \cdots & G^{1N} \\ \vdots & \ddots & \vdots \\ G^{N1} & \cdots & G^{NN} \end{bmatrix} = \begin{bmatrix} \epsilon & \cdots & \gamma \\ \vdots & \ddots & \vdots \\ \gamma & \cdots & \epsilon \end{bmatrix}, \quad C := \begin{bmatrix} 0 & \cdots & c \\ \vdots & \ddots & \vdots \\ c & \cdots & 0 \end{bmatrix}$$

- We can the optimization problem in terms of 2-marginals:

$$\min_{G\sim\Pi(\rho,\dots,\rho)} \mathrm{Tr}(CG)$$

- But non-trivial constraints on G

Convex relaxation

- $\min_{G\sim\Pi(\rho,\ldots,\rho)} \mathrm{Tr}(CG)$ with $G = \begin{bmatrix} G^{11} & \cdots & G^{1N} \\ \vdots & \ddots & \vdots \\ G^{N1} & \cdots & G^{NN} \end{bmatrix}$
- What are some necessary conditions on G?
  - $G^{ij}\mathbf{1} \equiv \gamma\mathbf{1} = \rho$
  - $G^{ii} \equiv \epsilon = \mathrm{diag}(\rho)$
  - $G \geq 0$, $G \succeq 0$
- Relax: drop all other constraints and obtain the convex problem

$$\min_{G=[G^{ij}]} \mathrm{Tr}(CG)$$

s.t. $G^{ij}\mathbf{1} \equiv \gamma\mathbf{1} = \rho,\ G^{ii} \equiv \epsilon = \mathrm{diag}(\rho),\ G \geq 0,\ G \succeq 0.$

- Polytope w. exp. num. of constraints. $\Rightarrow$ Polytope w. poly. num. of constraints.

Final SDP form

- Rewrite the cost $\mathrm{Tr}(CG) = \frac{N(N-1)}{2}\,\mathrm{Tr}(c\gamma)$
- Introduce $\delta \in \mathbb{R}^{L\times L}$

$$\delta = \frac{1}{N^2}\left[I \cdots I\right] G \begin{bmatrix} I \\ \vdots \\ I \end{bmatrix} = \frac{1}{N^2}\left[I \cdots I\right] \begin{bmatrix} \epsilon & \cdots & \gamma \\ \vdots & \ddots & \vdots \\ \gamma & \cdots & \epsilon \end{bmatrix} \begin{bmatrix} I \\ \vdots \\ I \end{bmatrix}$$

- Then

$$\delta = \frac{1}{N}\epsilon + \frac{N-1}{N}\gamma, \quad \delta\mathbf{1} = \rho, \quad \gamma = \frac{N\delta - \mathrm{diag}(\delta\mathbf{1})}{N-1}$$

- A convex-relaxed SDP lower bound for $V_{ee}^{\mathrm{SCE}}[\rho]$:

$$\min_{\delta\in\mathbb{R}^{L\times L}} \frac{N(N-1)}{2}\,\mathrm{Tr}\left(c\left(\frac{N}{N-1}\delta - \frac{1}{N-1}\mathrm{diag}(\rho)\right)\right)$$

s.t. $\delta\mathbf{1} = \rho$, $\delta \succeq 0$, $\delta \geq 0$, $\mathrm{diag}(\delta) = \frac{\rho}{N}$.

Why this relaxation is reasonable?

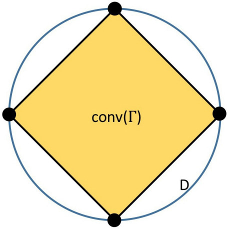
- Theorem (Friesecke-Vogler 18): The set of extreme points of N-representable symmetric 2-marginals (with Coulombic cost) is

$$\Gamma = \left\{ \frac{N}{N-1}\lambda\lambda^{\mathsf{T}} - \frac{1}{N-1}\mathrm{diag}(\lambda) \;\middle|\; \lambda \in \left\{0, \frac{1}{N}\right\}^{L},\ \lambda^{\mathsf{T}}\mathbf{1} = 1 \right\}$$

- Theorem (Khoo-Y. 19): Γ is a subset of the extreme points of

$$D \equiv \left\{ \frac{N}{N-1}\delta - \frac{1}{N-1}\mathrm{diag}(\delta\mathbf{1}) \;\middle|\; \delta \succeq 0,\ \delta \geq 0,\ \mathrm{diag}(\delta) = \frac{\delta\mathbf{1}}{N} \right\}$$

(Note that D is the set of feasible 2-marginals γ for the SDP).



Extensions

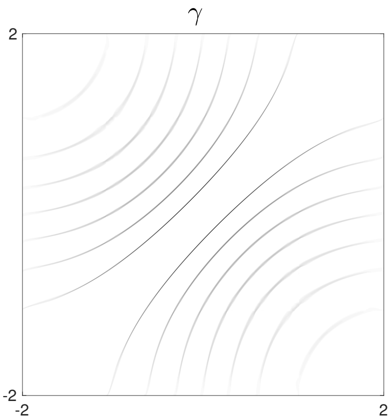
- An upper bound to $V_{ee}^{\mathsf{SCE}}[\rho]$ using three marginals

- The second quantization case
  - Yuehaw Khoo, Lin Lin, Michael Lindsey, Lexing Ying, Semidefinite relaxation of multi-marginal optimal transport for strictly correlated electrons in second quantization. SIAM J. Sci. Comput., 42(6), B1462B1489.

Numerical examples

- 1D electron: $N = 8$, $L = 1600$

$$\rho \propto \exp\left(-x^2/\sqrt{\pi}\right)$$

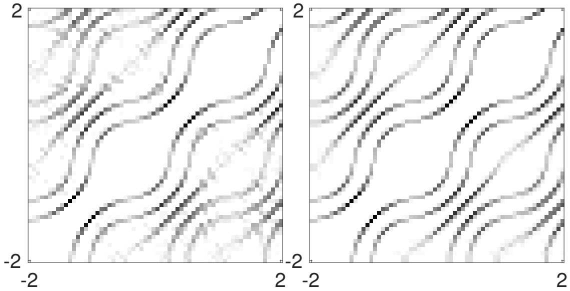


- 10^{25} entries if LP was used.

Numerical examples

- 1D electrons, $N = 8$

$$\rho(x) \propto \sin(4x) + 1.5$$

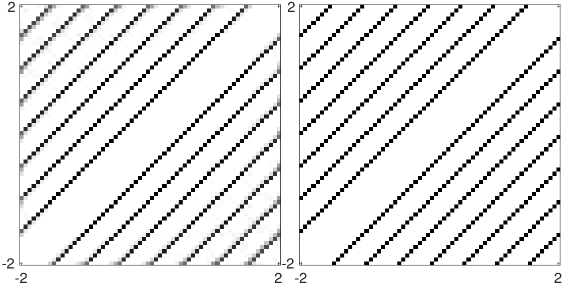


- Left: 2-marginal $\delta^* \to \gamma$. Relative gap = 4.2e-02
- Right: 3-marginal $\theta^* \to \gamma$. Relative gap = 3.9e-02

Numerical examples

- 1D electrons, $N = 8$

$$\rho(x) \propto 1$$

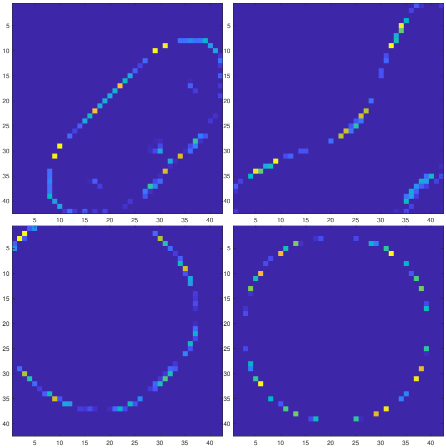


- Left: 2-marginal $\delta^* \to \gamma$. Relative gap = 4.9e-04
- Right: 3-marginal $\theta^* \to \gamma$. Relative gap = 1.0e-06

Numerical examples

- 2D electrons, $N = 5$

$$\rho(x,y) \propto 1$$



- Plots are slice of 2-marginal with one component fixed.
- 2-marginal $\delta^* \to \gamma$. Relative gap = 3.8e-02
- 3-marginal $\theta^* \to \gamma$. Relative gap = 3.5e-02

Thank you

- Email: lexing@stanford.edu
- URL: http://web.stanford.edu/~lexing