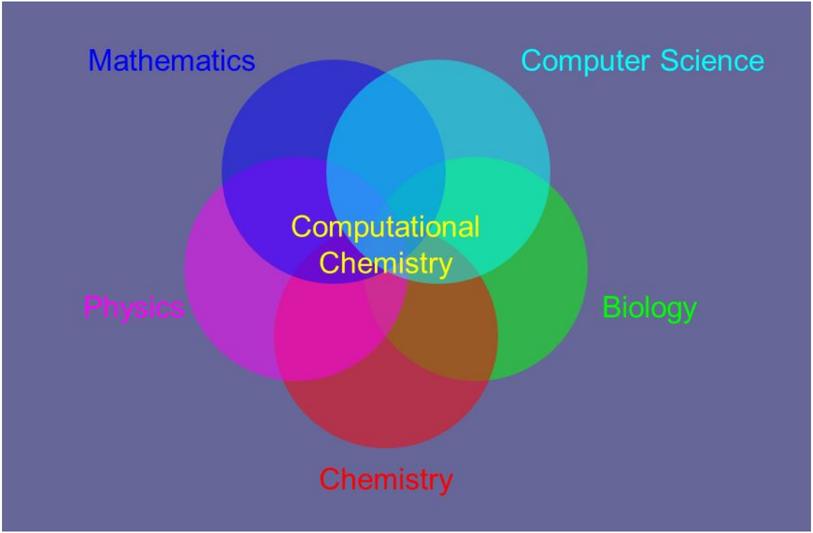
Finite range decomposition for multimarginal transport

Mircea Petrache, PUC Chile

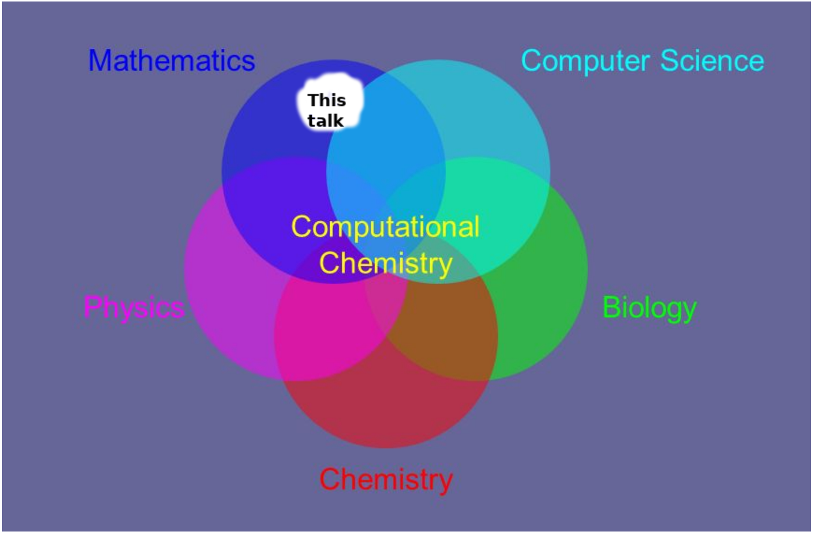
Optimal Transport Methods in Density Functional Theory, BIRS, Canada

February 2019 (15 min. talk)

INTRO: WHAT'S IN THE TALK



INTRO: WHAT'S IN THE TALK



Intro: Optimal transport with N marginals

- Fix $\mu \in \mathcal{P}(\mathbb{R}^d), \gamma_N \in \mathcal{P}((\mathbb{R}^d)^N)$ and $\mathsf{c} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$.
- $\gamma_N \mapsto \mu$ means that γ_N has N marginals equal to μ, i.e. $(\pi_j)_{\#}\gamma_N = \mu$ for $j = 1, \dots, N$.

Our problem:

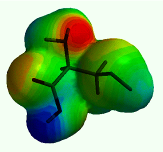
- γ_N assumed symmetric,
- power-law potential $\mathsf{c}(x, y) := \frac{1}{|x-y|^s}$ (or $:= \log \frac{1}{|x-y|}$ for $s = 0$)

$$\mathsf{OT}_{N,s}(\mu) := \min \left\{ \int_{(\mathbb{R}^d)^N} \sum_{i \neq j}^{N} \frac{1}{|x_i - x_j|^s} d\gamma_N(x_1, \dots, x_N) \;\middle|\; \begin{array}{l} \gamma_N \in \mathcal{P}_{sym}((\mathbb{R}^d)^N), \\ \gamma_N \mapsto \mu \end{array} \right\}.$$

INTRO: DENSITY FUNCTIONAL THEORY

- **Curse of dimensionality**:
  - Schrödinger equation $H\Psi = E_0\Psi$

    Ψ = state of N-particle system,
    H = operator on $\mathbb{R}^{3N}$,
    E_0 = ground state energy.
  - Chemical behavior $\sim$ energy differences $\ll$ total energy



Cystein molecule simulation,
(from Walter Kohn's Nobel prize laudation page)

Intro: Hohenberg-Kohn-Sham model

- Hohenberg-Kohn-Sham (HK) model
- (most of you know it better than me).
- Formulated in terms of the normalized one-particle density ρ.
- **Computational bottleneck:** Given ρ, compute the N-electron minimum energy at fixed one-particle density ρ.
- **Second step:** Optimize ρ including the interaction with the nuclei.

INTRO: DFT AND MULTIMARGINAL OT

- Hohenberg-Kohn functional: energy of N electrons of density ρ

$$\mathsf{HK}_N[\rho] := \text{Minimize:} \quad \langle \Psi_N, (\hbar^2 \Delta_{\mathbb{R}^{Nd}} + E_N)\Psi_N \rangle \quad \text{where:}$$

  - “$|\Psi_N|^2$” $\in \mathcal{P}((\mathbb{R}^d)^N)$ + other properties,
  - The measure $|\Psi_N|^2$ has marginals all equal to ρ,
  - $E_N(x_1, .., x_N) := \sum_{i \neq j} \frac{1}{|x_i - x_j|^s}$ (or take another $c(x, y)$ instead?)
- $\lim_{\hbar \to 0} \mathsf{HK}_N[\rho] = \mathsf{OT}_N(\rho)$

To know about this, ask Codina/Gero/Luigi/Mathieu/Ugo (in alphabetical order).

Intro: Leading order term = mean field

Theorem (Cotar-Friesecke-Pass '15, Petrache '15)

$$\begin{aligned} \mathsf{OT}_N(\mu) &= N^2\,\mathsf{MF}(\mu) + o(N^2), \\ \mathsf{MF}(\mu) &:= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mathsf{c}(x-y)\,d\mu(x)\,d\mu(y) \end{aligned}$$

if and only if $\mathsf{c}(x-y)$ *is balanced positive definite, i.e.*

$$\int\int \mathsf{c}(x-y)f(x)f(y)\,dx\,dy \geq 0 \quad \textit{whenever} \quad \int f = 0\,.$$

- Define $\mathsf{Exc}_N(\mu) := \mathsf{OT}_N(\mu) - N^2\mathsf{MF}(\mu)$.
- Theorem says:

$$\mathsf{Exc}_N(\mu) = o(N^2) \quad \Leftrightarrow \quad \mathsf{c}\ \textbf{balanced positive definite}.$$

NEXT-ORDER TERM FOR INVERSE POWER LAWS, $0 < s < d$

- $d = 1$, general kernels: unpublished note by Di Marino
- $s = 1, d = 3$: Lewin-Lieb-Seiringer '17, using Graf-Schenker '95
- Improving upon the different strategy Fefferman '85, we get:

Theorem (Cotar-Petrache, Adv. Math. 2019)

Let $d \geq 1$, $\mathsf{c}(x, y) = |x - y|^{-s}$ with $0 < s < d$. Under suitable hypotheses on ρ, as $N \to \infty$ we have

$$\mathsf{Exc}_N(\rho) = N^{1+\frac{s}{d}} \left(C_{UG}(d, s) \int_{\mathbb{R}^d} \rho^{1+\frac{s}{d}}(x) dx + o(1) \right),$$

where $C_{UG}(d, s) = \min$ energy of an "Uniform Riesz Gas" (special case: "Uniform Electron Gas" from DFT, for $s = d - 2$).

- In the above Cotar-Petrache '19 we show **a bit more**, bounding the **"third-order term"** asymptotic contribution as $N \to \infty$.

The problem of precise localization

- Idea of proof:
  - split $\mathrm{supp}(\rho)$ into small cubes,
  - use scaling (get power $1 + s/d$),
  - approximate $\int \rho^{1+s/d}(x)dx$ by a Riemann sum.
- **Main topic of the talk**:
  get "independence of contributions" coming from disjoint cubes.
- **Two linked topics**:
  1. Kernel decompositions for c
     (positive definite + finite range pieces: allows superadditivity)
  2. Space cut-off of ρ
     (Ruelle approach to subadditivity, classical tool in Stat. Phys.)

1. Finite-range decomp. and superadditivity

- **Input:** $\mathsf{c} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ which is positive definite (e.g. $1/|x-y|^s$ or $e^{-c|x-y|^2}$ or $|x-y|^2$ or products of these..)
- **Output:** splitting $\mathsf{c}_r,\ r \in \mathcal{R}$ such that
  - c_r is positive definite
  - c_r has finite range ($\mathsf{c}_r(x,y) = 0$ if $|x-y| > 2r$),
  - c is completely split $\mathsf{c} = \sum_r \mathsf{c}_r$.
- **Use of this in OT:**
  - $\mathsf{MF}[\mathsf{c}](\rho) = \sum_r \mathsf{MF}[\mathsf{c}_r](\rho)$ (by linearity)
  - $\mathsf{OT}_N[\mathsf{c}](\rho) \geq \sum_r \mathsf{OT}_N[\mathsf{c}_r](\rho)$ (by linearity $+\min \sum_r P_r \geq \sum_r \min P_r$)

$$\Rightarrow \qquad \mathsf{Exc}_N[\mathsf{c}](\rho) \geq \sum_r \mathsf{Exc}_N[\mathsf{c}_r](\rho).$$

..the N^2-contribution cancels as $N \to \infty$, only next-order remains!

(by leading order theo. + **positive definiteness** of c_r).

2. Convex envelope and subadditivity

Rewrite $\mathsf{Exc}_N(\mu) = \mathsf{OT}_N(\mu) - N^2\mathsf{MF}(\mu)$ for $N \in \mathbb{N}$, by new formula:

$$\mathsf{Exc}(\nu) := \mathsf{OT}_{|\nu|}\left(\frac{\nu}{|\nu|}\right) - \mathsf{MF}(\nu) \quad \begin{cases} \mathsf{Exc}(\nu) = \mathsf{Exc}_N(\mu) \\ \text{if } \nu = N\mu \text{ and } |\nu| = N. \end{cases}$$

- This **agrees with** Exc_N across different $N \in \mathbb{N}$, **and it's subadditive**:

$$\mathsf{Exc}\left(\sum_i \nu_i\right) \leq \sum_i \mathsf{Exc}(\nu_i)$$

(if all above measures have integer mass)

- $\overline{\mathsf{Exc}}$:= **(lower) convex envelope of** Exc.
  - We get a "**fractional number of marginals**" OT-problem
  - Physically, it's the **grand-canonical version** of Exc.
  - The approach is ubiquitous in classical Statistical Mechanics.

3. Random packings for mixing the ingredients

- Localization: split c or ν into local parts:

$$\overline{\mathsf{Exc}}[\mathsf{c}](\nu) \geq \sum_r \overline{\mathsf{Exc}}[\mathsf{c}_r](\nu), \tag{1}$$

$$\overline{\mathsf{Exc}}[\mathsf{c}]\Big(\sum_i \nu_i\Big) \leq \sum_i \overline{\mathsf{Exc}}[\mathsf{c}](\nu_i). \tag{2}$$

- **Can we use both contemporarily?**
- **Use construction of c_r in order to match the two setups**

$$\begin{aligned} \mathsf{h}_r(x-y) &:= 1_{B_r} * 1_{B_r}(x-y) \quad \text{positive definite,} \\ \widetilde{\mathsf{c}}_r(x,y) &:= \int \left[1_{B_r(p)}(x) 1_{B_r(p)}(y) \mathsf{c}(x-y)\right] dp \\ &= \mathsf{h}_r(x-y)\mathsf{c}(x-y) \quad \text{positive definite.} \end{aligned}$$

  - $\widetilde{\mathsf{c}}_r$ **fits in** (1)
  - The integrand gives a **cut-off like in** (2) on the ball $B_r(p)$
- **Strategy that worked: cut-off along "random" packings!**

Where this seems to be going (personal view)

- We have a simple "averaging" amongst packings:

  Via stochastic geometry we can extend this further

- So far we tried "simple/basic" cut-off functions:

  Finite-range decomposition theory connects it to PDE-ideas

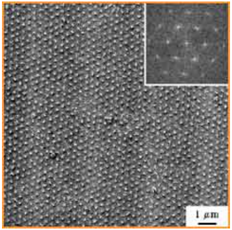
- We did sharp asymptotics for $N \to \infty$, oscillation bounds:

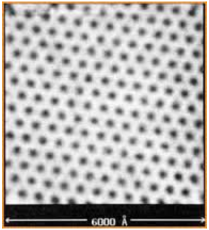
  What about sharper (randomized) algorithm analysis for "large N optimal transport"?

  (I.e. get better complexity bounds with high probability)

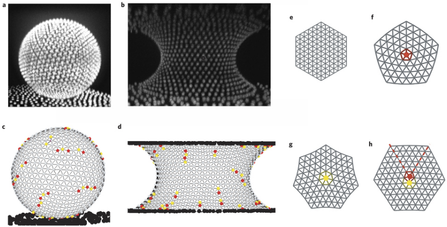
- Relate OT complexity-reduction problem to "pure" CS topics:

  cut decompositions / regularity lemmas / dimensionality reduction

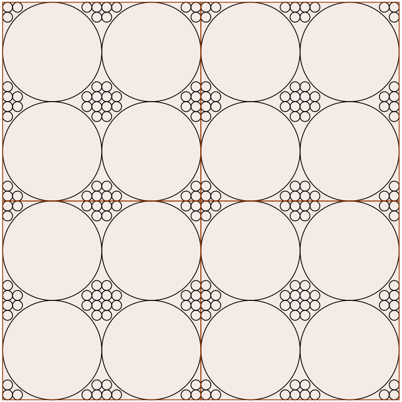




THANK YOU!



Our Packing, $M = 2$



Packing strategy

- "Swiss cheese" lemma Lebowitz-Lieb '72: Cover $[0,1]^d$ by balls $\mathcal{F} = \{B\}_B$ of radii $0 < R_1 < \cdots < R_M$ with
  - geometric growth: $R_{i+1} > C_d R_i$,
  - c_i :=(volume fraction covered by R_i-balls) $= 1/M + O(M^{-2})$.

  Extend by $\mathbb{Z}^d$-periodicity.
- For $f \in L^1$ with compact support, $\langle f\rangle(x,y) := \int_{\mathbb{R}^d} f(x+p, y+p)dp$. Then

$$\sum_{B\in\mathcal{F}} \langle 1_B(x)1_B(y)c(x-y)\rangle = c(x-y)\sum_{i=1}^{M} c_i \frac{1_{B_{R_i}} * 1_{B_{R_i}}(x-y)}{|B_{R_i}|}.$$

Positive definiteness criterion

Lemma (perturbative positive-definiteness criterion)

$$|\partial_x^\beta g(x)| \lesssim |x|^{-s-|\beta|} \text{ for all multiindices } |\beta| \le d.$$
$$\Rightarrow$$
$$|\hat{g}(\xi)| \lesssim |\xi|^{s-d}.$$

To use it we further mollify

$$Q_i(x) = \frac{1_{B_R} * 1_{B_R}(x)}{|B_R|} \quad \mapsto \quad Q_{i,\eta}(x) = \int_{1-\eta}^{1+\eta} \frac{1_{B_{tR}} * 1_{B_{tR}}(x)}{|B_{tR}|} \rho_\eta(t) dt.$$

(can still re-express as averaging over dilated packings)

Positive definite error term

Lemma (perturbative positive-definiteness criterion)

$$|\partial_x^\beta g(x)| \lesssim |x|^{-s-|\beta|} \textit{ for all multiindices } |\beta| \leq d.$$
$$\Rightarrow \quad |\widehat{g}(\xi)| \lesssim |\xi|^{s-d}.$$

By adding $\epsilon/|x-y|^s$, we ensure $\widehat{w}(\xi) = \widehat{err}(\xi) + C\epsilon|\xi|^{s-d} > 0$.
(Recall that $\widehat{w} > 0$ implies that w is positive definite.)

Proposition (kernel localization + small error)

$$\frac{1}{|x_1-x_2|^s} = \frac{1}{1-\epsilon}\left(\int_\Omega \left[\sum_{A\in F_\omega} \frac{1_A(x_1)1_A(x_2)}{|x_1-x_2|^s}\right] d\mathbb{P}(\omega) + w(x_1-x_2)\right),$$

where

1. w *is positive definite.*
2. *OT next-order term with kernel* w *exists and has good bounds.*