

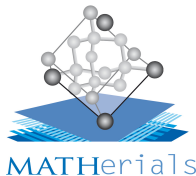
Wasserstein-type interpolation for generic mixture models: application to model-order reduction in quantum chemistry

G. Dusson^{1,3}, V. Ehrlacher², N. Nouaime³

¹CNRS

²Ecole des Ponts ParisTech & INRIA

³Université de Besançon



Outline of the talk

Motivation: model-order reduction for electronic structure calculations

Introduction to optimal transport, Wasserstein spaces and barycenters

Modified Wasserstein metric for mixtures

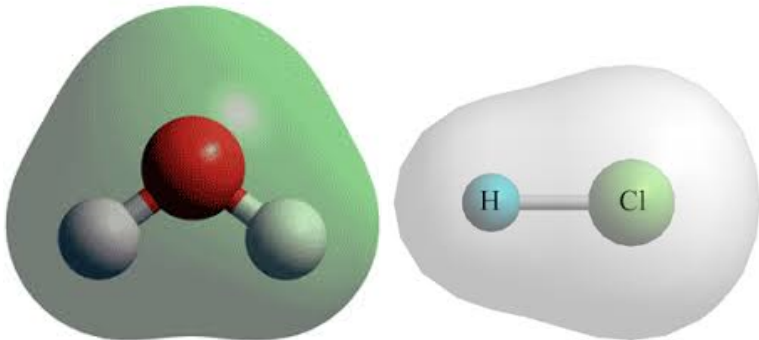
Outline of the talk

Motivation: model-order reduction for electronic structure calculations

Introduction to optimal transport, Wasserstein spaces and barycenters

Modified Wasserstein metric for mixtures

Motivation: electronic structure calculation for molecules



Computation of the **ground state of electrons in a molecule**: electric, optical, magnetic properties, prediction of chemical reactions, computation of inter-atomic or inter-molecular potentials for MD simulations...

Electronic Schrödinger model: Born-Oppenheimer approximation

Let us consider a physical system composed of:

- **M nuclei**, composed of protons and neutrons, that are assumed to be (fixed) classical point charges whose:
 - positions in \mathbb{R}^3 are denoted by $R_1, \dots, R_M \in \mathbb{R}^3$;
 - electric charges are denoted by $Z_1, \dots, Z_M > 0$.
- **N electrons**, considered as quantum particles, and represented by a complex-valued wavefunction $\psi(x_1, \dots, x_N)$, where for all $1 \leq i \leq N$, $x_i \in \mathbb{R}^3$. Since electrons are fermionic particles, ψ is antisymmetric with respect to permutation of the order of the variables.

Physical interpretation: $|\psi(x_1, \dots, x_N)|^2$ represents the probability density of finding the N electrons at positions $(x_1, \dots, x_N) \in \mathbb{R}^{3N}$

$$\|\psi\|_{L^2(\mathbb{R}^{3N})}^2 = \int_{\mathbb{R}^{3N}} |\psi|^2 = 1.$$

Electronic (one or two-body) density:

$$\rho(x) = \int_{\mathbb{R}^{3(N-1)}} |\psi(x, x_2, \dots, x_N)|^2, \quad \tau(x, y) = \int_{\mathbb{R}^{3(N-2)}} |\psi(x, y, x_3, \dots, x_N)|^2,$$

Electronic Schrödinger model: Born-Oppenheimer approximation

Let $\mu = (R_1, \dots, R_M) \in \mathbb{R}^{3M}$ the set of positions of the nuclei.

$$V_\mu(x) = - \sum_{k=1}^M \frac{Z_k}{|x - R_k|}, \quad x \in \mathbb{R}^3.$$

For a given value of μ , the corresponding ground state electronic wavefunction ψ_μ is solution to the eigenvalue problem (electronic Schrödinger problem)

$$H_\mu \psi_\mu = E_\mu \psi_\mu,$$

with

$$H_\mu = -\Delta_{x_1, \dots, x_N} + \sum_{i=1}^N V_\mu(x_i) + \sum_{1 \leq i < j \leq N} \frac{1}{|x_i - x_j|}$$

Problem: Linear eigenvalue problem for functions defined on the high-dimensional space \mathbb{R}^{3N} .

Density Functional Theory and Kohn-Sham models

Instead, quantum chemists prefer to solve **approximate problems**, which enable to obtain approximations of the electronic one-body density $\rho_\mu(x)$ (or electronic two-body densities $\tau_\mu(x, y)$).

These approximate models boil down to solving **nonlinear eigenvalue problems** for functions defined on \mathbb{R}^3 (or \mathbb{R}^6).

Find $\Phi_\mu(x) = (\phi_{1,\mu}(x), \dots, \phi_{N,\mu}(x))$ solution to

$$\begin{cases} (-\Delta_x + V_\mu(x) + W[\rho_\mu](x)) \phi_{i,\mu}(x) = \epsilon_{i,\mu} \phi_{i,\mu}(x) \\ \rho_\mu(x) = \frac{1}{N} \sum_{i=1}^N |\phi_{i,\mu}(x)|^2 \end{cases}$$

Example: $W[\rho](x) = \int_{y \in \mathbb{R}^3} \frac{1}{|x-y|} \rho(y) dy$

Roothan algorithms

In practice, these problems are solved by means of **iterative algorithms**: starting from an **initial guess** ρ_μ^0 for the electronic density,

Iteration k :

1. From ρ_μ^k , compute $W_\mu^k := W[\rho_\mu^k]$
2. Compute $\Phi_\mu^k = (\phi_{1,\mu}^k, \dots, \phi_{N,\mu}^k)$ solution to

$$(-\Delta + V_\mu + W_\mu^k)\phi_{i,\mu}^k = \epsilon_{i,\mu}^k \phi_{i,\mu}^k$$

3. Compute $\rho_\mu^{k+1}(x) = \frac{1}{N} \sum_{i=1}^N |\phi_{i,\mu}^k(x)|^2$.

Objective: We would like to have numerical methods which enable to **quickly compute** ρ_μ **for many values of** μ : ROM!

Challenge: The computational cost of these iterative algorithms **very strongly depend on the value of the initial guess** for the electronic density.

Model-order reduction: interpolation

Approach: Compute an approximation of ρ_μ as an interpolation in a "good sense" from a few well-selected snapshots $\rho_{\mu_1}, \dots, \rho_{\mu_n}$.

Important property: If the set of positions of the nuclei $\mu = (R_1, \dots, R_k)$ is shifted by a translation vector b , ($\mu' = (R_1 + b, \dots, R_M + b)$), then the electronic density is also shifted by the same translation vector:

$$\rho_{\mu'}(x) = \rho_\mu(x + b)$$

Similar issue as in ROMs for transport-dominated problems...

Use of **optimal transport theory** to build interpolations which respect this property.

Optimal transport for model order reduction

Non-exhaustive list...

- [Iollo, Lombardi, 2014]: transport maps computed as linear combinations of optimal transport maps
- [VE, Lombardi, Mula, Vialard, 2020]: use of Wasserstein barycenters (1d) with greedy algorithms
- [Iollo, Taddei, 2022]: use of Gaussian models
- [Do, Feydy, Mula, 2023]: extension of the Wasserstein barycenter methodology to higher-dimensional settings
- [Rim, Peherstorfer, Mandli, 2023]: towards a Galerkin approach combined with optimal transport

Outline of the talk

Motivation: model-order reduction for electronic structure calculations

Introduction to optimal transport, Wasserstein spaces and barycenters

Modified Wasserstein metric for mixtures

Probability measures with finite second-order moments

Let $d \in \mathbb{N}^*$ and $\Omega \subset \mathbb{R}^d$ an open domain.

Let $\mathcal{P}_2(\Omega)$ denote the set of **probability measures** u on Ω with finite second-order moments, i.e.

$$\int_{\Omega} u(dx) = 1, \quad \int_{\Omega} (1 + |x|)^2 u(dx) < +\infty.$$

Example: Let $\rho \in L^1(\Omega)$ such that

$$\rho \geq 0, \quad \int_{\Omega} \rho(x) dx = 1, \quad \int_{\Omega} (1 + |x|)^2 \rho(x) dx < +\infty. \quad (1)$$

Then, the probability measure $u(dx) := \rho(x) dx$ belongs to $\mathcal{P}_2(\Omega)$.

Wasserstein space

The **2-Wasserstein (or Kantorovich-Rubinstein) metric** is a distance function defined between two probability measures $\mu, \nu \in \mathcal{P}_2(\Omega)$ and is denoted by

$$W_2(\mu, \nu).$$

The set $(\mathcal{P}_2(\Omega), W_2)$ then defines a metric space, called the **Wasserstein space**.

Its precise definition will come later... **Patience!**

I first would like to explain to you the interest of this distance with respect to **interpolation properties**.

Interpolation in the Wasserstein or $L^2(\Omega)$ space

Let $\rho_1, \rho_2 \in L^2(\Omega) \cap L^1(\Omega)$ which satisfy (1).

Let $t \in [0, 1]$ and consider the two following problems:

- **Interpolation in the $L^2(\Omega)$ space:** Find $\rho_t^{L^2} \in L^2(\Omega)$ such that

$$\rho_t^{L^2} = \operatorname{argmin}_{\rho \in L^2(\Omega)} (1-t) \|\rho - \rho_1\|_{L^2(\Omega)}^2 + t \|\rho - \rho_2\|_{L^2(\Omega)}^2.$$

Then, we all know that the solution is $\rho_t^{L^2}$ is the barycentric combination of ρ_1 and ρ_2 , i.e.

$$\rho_t^{L^2} := (1-t)\rho_1 + t\rho_2.$$

- **Interpolation in the Wasserstein space:** Find $\rho_t^{W_2} \in \mathcal{P}_2(\Omega)$ such that

$$\rho_t^{W_2} = \operatorname{argmin}_{\rho \in \mathcal{P}_2(\Omega)} (1-t) W_2(\rho, \rho_1)^2 + t W_2(\rho, \rho_2)^2.$$

The measure $\rho_t^{W_2}$ is unique and is called the **McCann's interpolant** between ρ_1 and ρ_2 .

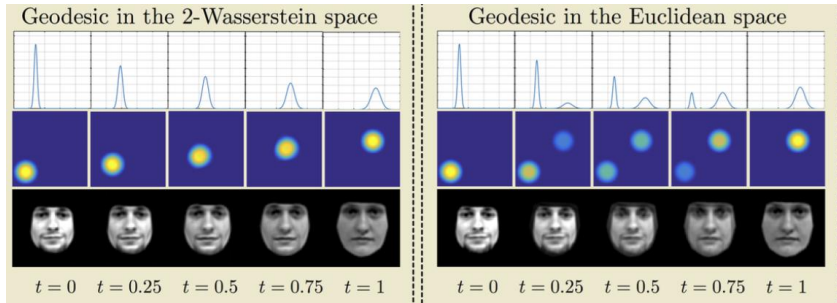
What does $\rho_t^{L^2}$ and $\rho_t^{W_2}$ look like?

Comparison between the Wasserstein and $L^2(\Omega)$ interpolation

[Kolouri et al. 2016]

$$\rho_t^{W_2}$$

$$\rho_t^{L^2}$$



Interesting property of the Wasserstein metric:

If $\rho_2 = \rho_1(\cdot - b)$ for some $b \in \mathbb{R}^d$, then it holds that

$$\rho_t^{W_2} = \rho_1(\cdot - tb).$$

Barycenters in the Wasserstein or $L^2(\Omega)$ space

Let $n \in \mathbb{N}^*$ and $(\rho_1, \rho_2, \dots, \rho_n) \in \mathcal{P}_2(\Omega)^n$. Let $\Lambda := (\lambda_1, \dots, \lambda_n) \in [0, 1]^n$ such that $\sum_{i=1}^n \lambda_i = 1$, and consider the minimization problem:

Find $\rho_\Lambda^{W_2} \in \mathcal{P}_2(\Omega)$ such that

$$\rho_\Lambda^{W_2} = \operatorname{argmin}_{\rho \in \mathcal{P}_2(\Omega)} \sum_{i=1}^n \lambda_i W_2(\rho, \rho_i)^2.$$

The measure $\rho_\Lambda^{W_2}$ is unique and is called the **Wasserstein barycenter** of (ρ_1, \dots, ρ_n) with weights Λ .

This object is the Wasserstein counterpart of the $L^2(\Omega)$ barycenter of a set of functions $(\rho_1, \dots, \rho_n) \in L^2(\Omega)^n$ with barycentric weight Λ . Indeed,

$$\rho_\Lambda^{L^2} := \sum_{i=1}^n \lambda_i \rho_i,$$

is equivalently the unique minimizer of

$$\rho_\Lambda^{L^2} = \operatorname{argmin}_{\rho \in L^2(\Omega)} \sum_{i=1}^n \lambda_i \|\rho - \rho_i\|_{L^2(\Omega)}^2.$$

Wasserstein barycenters

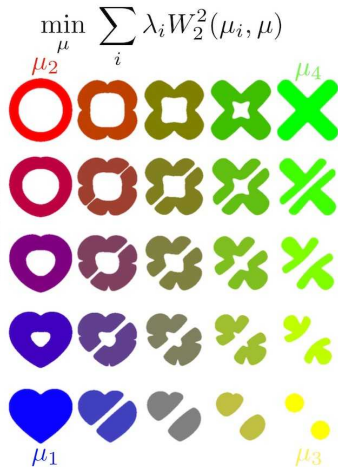
Barycenters in the Wasserstein space

Martial Agueh ^{*}, Guillaume Carlier [†]

August 17, 2010

Abstract

In this paper, we introduce a notion of barycenter in the Wasserstein space which generalizes McCann's interpolation to the case of more than two measures. We provide existence, uniqueness, characterizations and regularity of the barycenter, and relate it to the multi-marginal optimal transport problem considered by Gangbo and Świąch in [8]. We also consider some examples and in particular rigorously solve the gaussian case. We finally discuss convexity of functionals in the Wasserstein space.



Wasserstein metric: optimal transport

Let $\rho_1, \rho_2 \in \mathcal{P}_2(\Omega)$. Then,

$$W_2(\rho_1, \rho_2)^2 := \inf_{\pi \in \mathcal{P}(\Omega \times \Omega)} \int_{\Omega \times \Omega} |x - y|^2 \pi(x, y) dx dy.$$

$$\int_{y \in \Omega} \pi(x, y) dy = \rho_1(x)$$

$$\int_{x \in \Omega} \pi(x, y) dx = \rho_2(y)$$

where $\mathcal{P}(\Omega \times \Omega)$ is the set of probability measures on $\Omega \times \Omega$.

Kantorovich formulation of optimal transport problem

Numerical methods to compute Wasserstein barycenters

- Simple explicit formulas for one-dimensional densities
- Several numerical methods exist for solving such problems and computing Wasserstein barycenters in higher-dimensional settings: linear programming, auction algorithm, entropic regularization (Sinkhorn algorithm)...

Nevertheless, all these methods **remain quite expensive from a computational point of view.**

- There are some families of densities for which analytical formulas are explicitly known.

Example: Gaussian densities [lollo, Taddei, 2022]

$$\rho_1 \sim \mathcal{N}(m_1, \Sigma_1) \quad \rho_2 \sim \mathcal{N}(m_2, \Sigma_2)$$

$$W_2^2(\rho_1, \rho_2) = \|m_1 - m_2\|^2 + \text{Tr} \left(\Sigma_1 + \Sigma_2 - 2 \left(\Sigma_1^{1/2} \Sigma_2 \Sigma_1^{1/2} \right)^{1/2} \right)$$

$$\rho_t^{W_2} \sim \mathcal{N}(m_t, \Sigma_t) \text{ with explicit formulas for } m_t \text{ and } \Sigma_t.$$

Outline of the talk

Motivation: model-order reduction for electronic structure calculations

Introduction to optimal transport, Wasserstein spaces and barycenters

Modified Wasserstein metric for mixtures



Gaussian mixtures

In [Delon, Desolgneux, 2020]: mathematical analysis of a **modified Wasserstein metric** defined on the space of Gaussian mixtures, i.e. probability densities that are convex combinations of Gaussians.

$$\rho = \sum_{k=1}^K \pi_k \rho_k, \quad \pi_k \geq 0 \text{ with } \sum_{k=1}^K \pi_k = 1 \text{ and } \rho_k \sim \mathcal{N}(m_k, \Sigma_k)$$

Let $\rho^1 = \sum_{k_1=1}^{K_1} \pi_{k_1}^1 \rho_{k_1}^1$ and $\rho^2 = \sum_{k_2=1}^{K_2} \pi_{k_2}^2 \rho_{k_2}^2$ be two such mixtures

$$MW_2^2(\rho^1, \rho^2) := \min_{(w_{k_1 k_2})} \sum_{k_1, k_2} w_{k_1 k_2} W_2^2(\rho_{k_1}^1, \rho_{k_2}^2), \quad (2)$$

over the set of $w_{k_1 k_2} \geq 0$ such that

$$\sum_{k_1} w_{k_1 k_2} = \pi_{k_2}^2 \quad \sum_{k_2} w_{k_1 k_2} = \pi_{k_1}^1$$

It holds that

$$MW_2^2(\rho^1, \rho^2) \geq W_2^2(\rho^1, \rho^2)$$



Modified Wasserstein barycenter

Denoting by $(w_{k_1 k_2}^*)$ a minimizer of problem (2), $t \in [0, 1]$,

$$\rho_t^{MW_2} = \sum_{k_1 k_2} w_{k_1 k_2}^* \rho_{t, k_1 k_2}^{W_2},$$

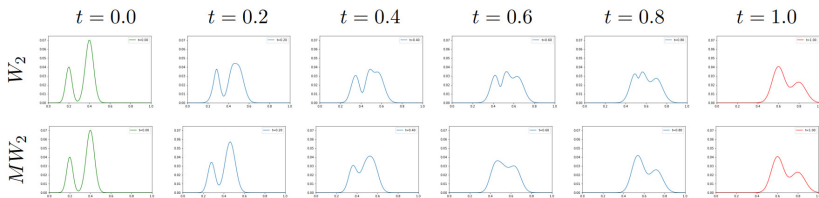
where $\rho_{t, k_1 k_2}^{W_2}$ is the Wasserstein barycenter with weight t between $\rho_{k_1}^1$ and $\rho_{k_2}^2$.

Similar formulas for barycenters of more than two gaussian mixtures.



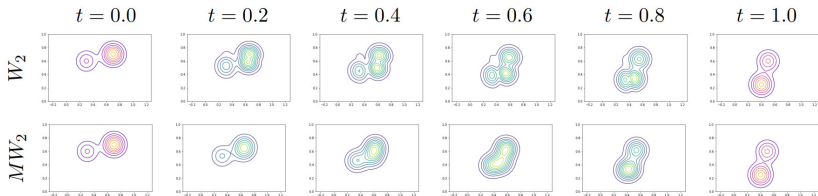
Advantages of the modified Wasserstein metric

- **Very quick to compute in any dimension** using analytic formulas for Gaussians: resolution of a small linear programming problem (provided that the number of gaussians in the mixture remains reasonably small).
- Better interpolation properties than the original Wasserstein metric: in particular, **the MW_2 barycenter of mixtures of gaussians is still a mixture of gaussians** (this is not the case for the W_2 barycenter)



Advantages of the modified Wasserstein metric

- **Very quick to compute in any dimension** using analytic formulas for Gaussians: resolution of a small linear programming problem (provided that the number of gaussians in the mixture remains reasonably small).
- Better interpolation properties than the original Wasserstein metric: in particular, **the MW_2 barycenter of mixtures of gaussians is still a mixture of gaussians** (this is not the case for the W_2 barycenter)



Extension to other types of mixtures

In [Dusson, VE, Nouaime, 2023], we extend the notion of this modified Wasserstein metric to other types of mixtures with a view to use it in order to interpolate electronic densities in quantum chemistry applications.

In general, let $\mathcal{A} \subset \mathcal{P}_2(\Omega)$ be a subset of probability measures, called the atomic set. We consider mixtures of elements of \mathcal{A} defined as convex combination of elements of \mathcal{A} ,

$$\rho = \sum_{k=1}^K \pi_k \mathbf{a}_k, \quad \mathbf{a}_k \in \mathcal{A}, \quad \pi_k \geq 0, \quad \sum_{k=1}^K \pi_k = 1.$$

Let $\rho^1 = \sum_{k_1=1}^{K_1} \pi_{k_1}^1 \mathbf{a}_{k_1}^1$ and $\rho^2 = \sum_{k_2=1}^{K_2} \pi_{k_2}^2 \mathbf{a}_{k_2}^2$ be two such mixtures

$$MW_2^2(\rho^1, \rho^2) := \min_{(w_{k_1 k_2})} \sum_{k_1, k_2} w_{k_1 k_2} W_2^2(\mathbf{a}_{k_1}^1, \mathbf{a}_{k_2}^2), \quad (3)$$

over the set of $w_{k_1 k_2} \geq 0$ such that

$$\sum_{k_1} w_{k_1 k_2} = \pi_{k_2}^2 \quad \sum_{k_2} w_{k_1 k_2} = \pi_{k_1}^1$$

Similar formulas as in the Gaussian case for modified Wasserstein barycenters.

Location-scatter measure case

Example of interesting atomic set: location-scatter measures

Let $a \in \mathcal{P}_2(\Omega)$. The set of location-scatter atoms generated from a is

$$\mathcal{A} := \left\{ T \# a, T : x \in \mathbb{R}^d \mapsto Ax + b, b \in \mathbb{R}^d, A \in \mathcal{S}_d \right\}$$

where \mathcal{S}_d is the set of symmetric positive definite matrices of $\mathbb{R}^{d \times d}$.

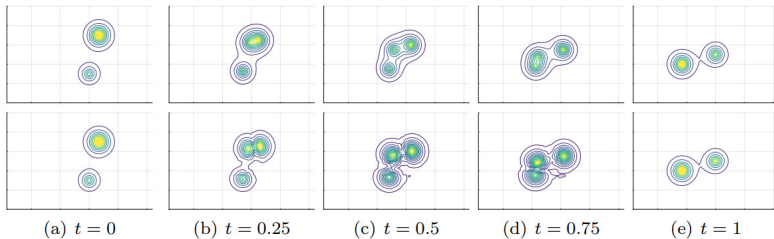
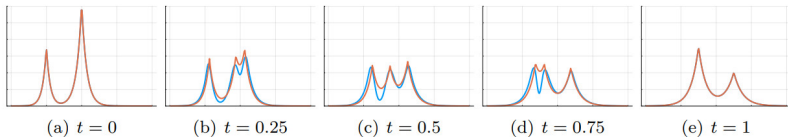
Why interesting? Because there are analytic formulas for the Wasserstein distance and barycenter between two elements of \mathcal{A} !

The computation of modified Wasserstein metric and barycenter between mixture of location-scatter atoms is then very easy provided that the number of terms in the mixture remains not too large.



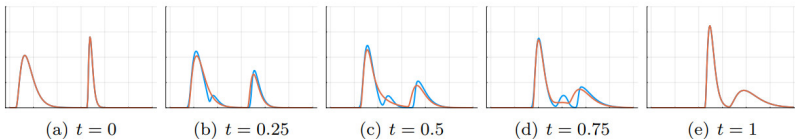
Examples

Slater distribution: $a(x) = e^{-|x|}$

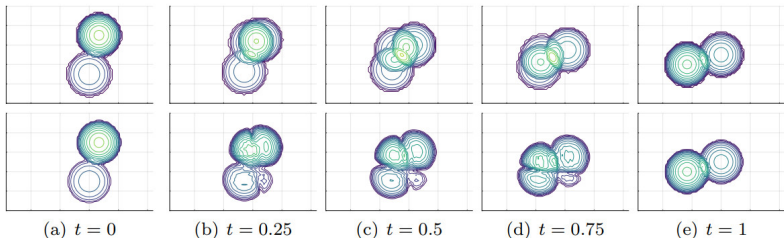


Examples

Gamma distribution: $x \geq 0$, $a_{\alpha,\beta}(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$



Wigner distribution: $a(x) = \sqrt{1 - |x|^2}$

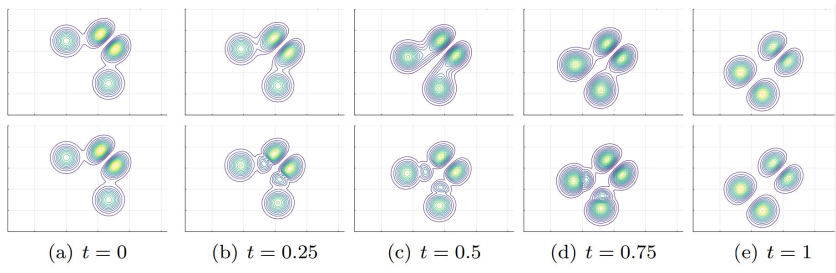


Mixture of squared Slater determinants

In quantum chemistry computation, an interesting set of atoms for two-body electronic densities is the set of squared Slater determinants of Gaussian or Slater functions: $x, y \in \mathbb{R}^d$

$$a(x, y) = \frac{1}{Z} (g_{m_1, \Sigma_1}(x) g_{m_2, \Sigma_2}(y) - g_{m_1, \Sigma_1}(y) g_{m_2, \Sigma_2}(x))^2$$

in order to interpolate between two-body densities



Conclusion and perspectives

Summary:

- The modified Wasserstein metric introduced by Delon and Desolneux for Gaussian mixtures has nice properties: easy to compute in arbitrary dimension provided that the number of terms in the mixture remains not too large, better interpolation properties than the exact Wasserstein metric
- With G. Dusson and N. Nouaime, we extend the notion of this modified Wasserstein metric to other types of mixtures, with a view to use it to build reduced-order model for electronic structure calculations in quantum chemistry

On-going work and perspectives:

- Implementation of this interpolation scheme to accelerate electronic structure calculations on real molecules (on-going work with Etienne Polak and Geneviève Dusson)
- Theoretical estimates of the error made by this type of modified Wasserstein interpolations (in simple cases) (joint work with Maxime Dalery, Geneviève Dusson and Alexeï Lozinski)
- useful for other types of applications where optimal transport can help?

Thank you for your attention!