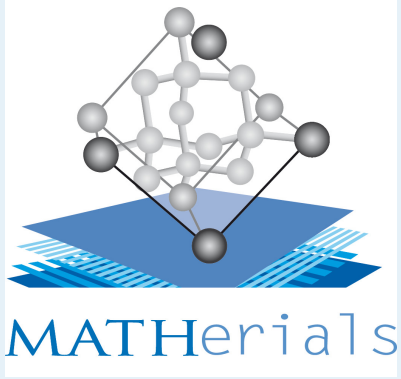
Surrogate modeling in electronic structure calculations using optimal transport

Virginie Ehrlacher

Ecole des Ponts ParisTech & INRIA

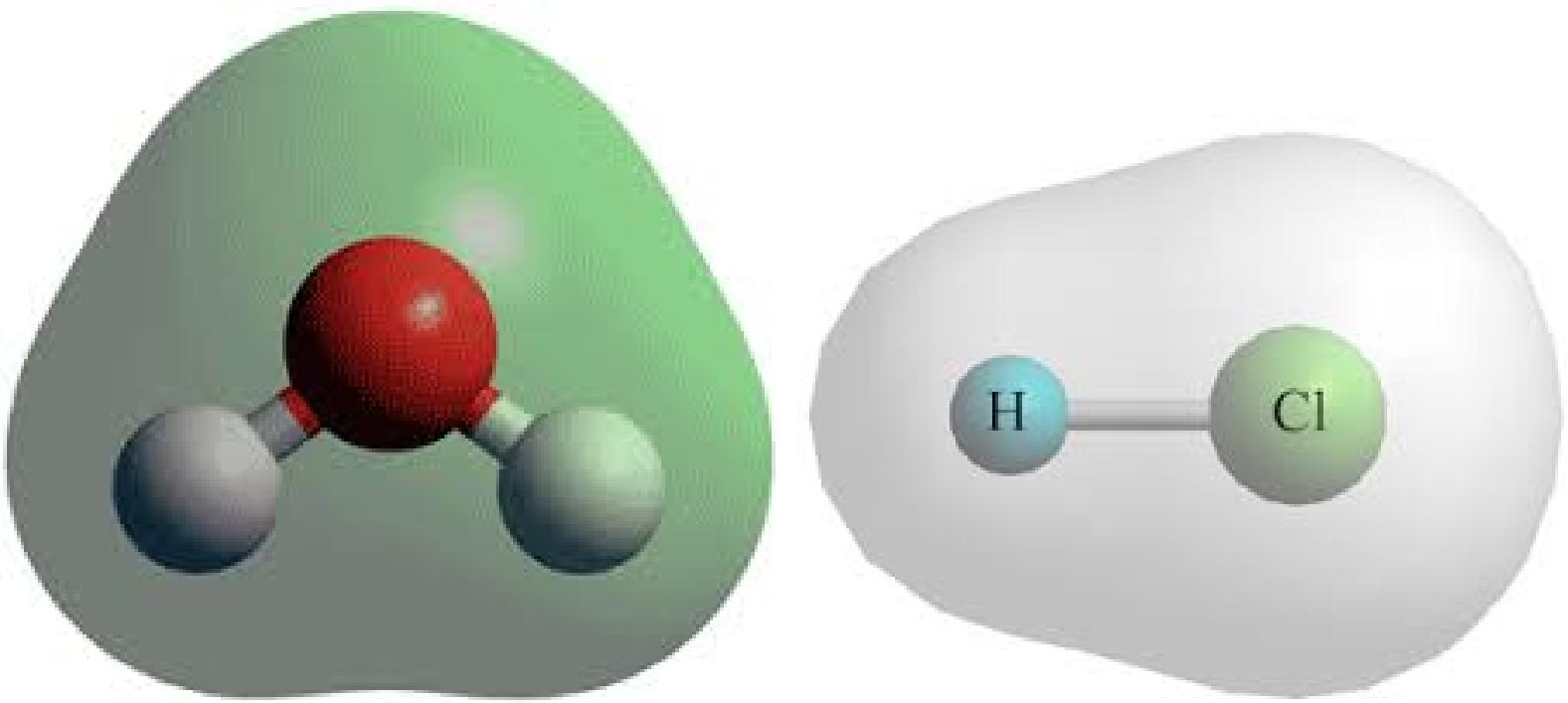
joint work with Maxime Dalery (UFC), Geneviève Dusson (CNRS & Université Franche-Comté), Alexei Lozinski (Université Franche-Comté), Etienne Polack (Ecole des Ponts)



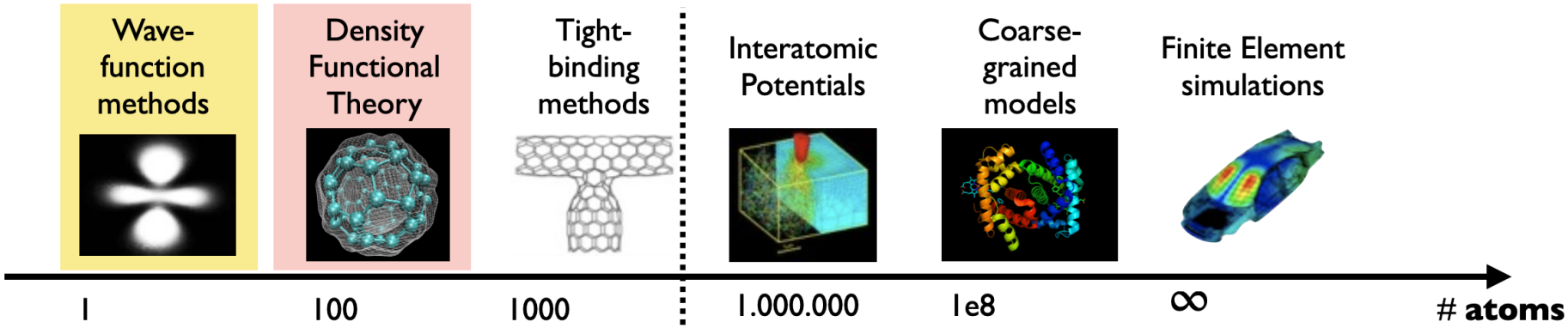
Context : Electronic structure calculation

Molecules are composed of :

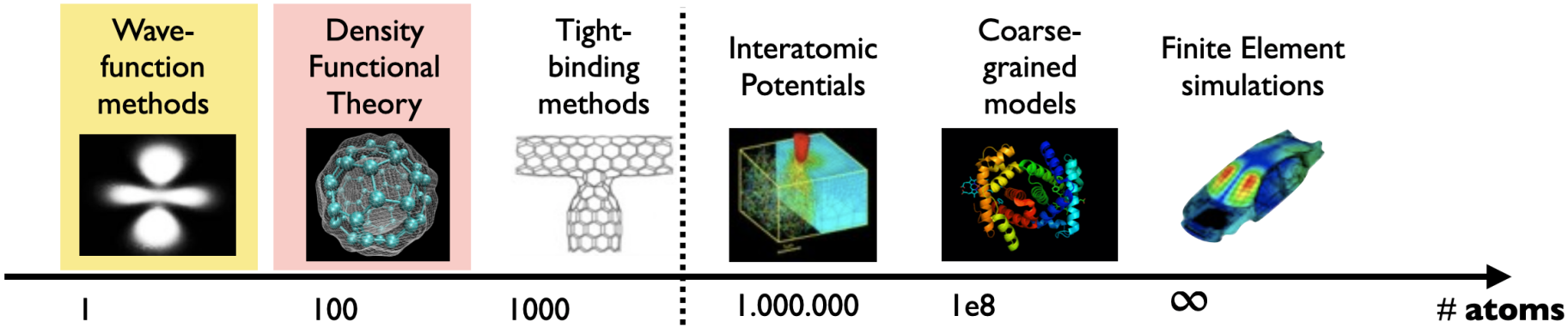
- nuclei, composed of protons and neutrons, that are often modeled as classical point charges ; (Born-Oppenheimer approximation) ;
- electrons, which are much lighter particles than nuclei and have to be modeled as quantum particles.



Context : Electronic structure calculations

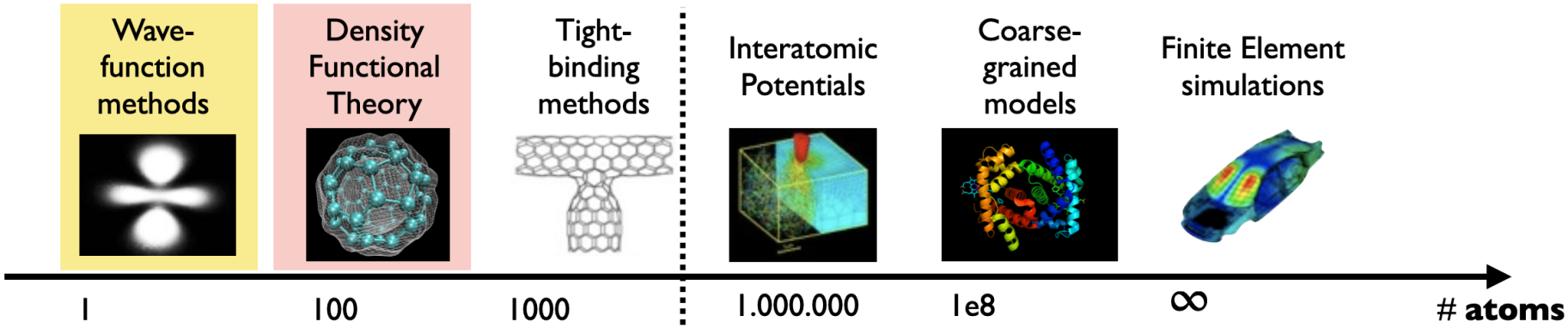


Context : Electronic structure calculations



- Costly but high accuracy $O(N^3)$
- A large number of models : Schrödinger, Hartree-Fock, Density Functional Theory

Context : Electronic structure calculations

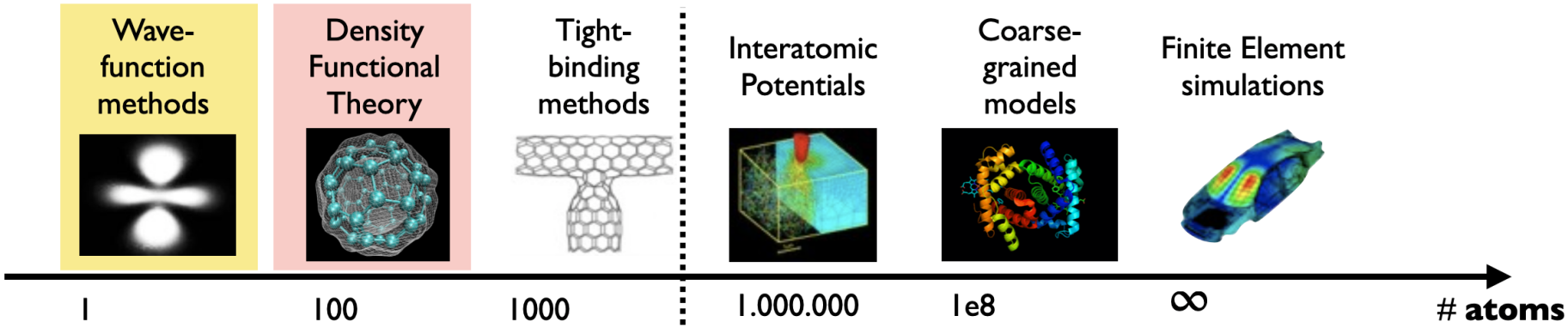


- Costly but high accuracy $O(N^3)$
- A large number of models : Schrödinger, Hartree-Fock, Density Functional Theory

Used for different purposes :

- Electric, optical, magnetic properties, prediction of chemical reactions...
- Ab initio molecular dynamics
- Building databases to construct interatomic potentials

Context : Electronic structure calculations



- Costly but high accuracy $O(N^3)$
- A large number of models : Schrödinger, Hartree-Fock, Density Functional Theory

Used for different purposes :

- Electric, optical, magnetic properties, prediction of chemical reactions...
- Ab initio molecular dynamics
- Building databases to construct interatomic potentials

Focus on the **ground state of electrons**, i.e. the state of electrons with the lowest possible energy.

Quantum mechanics

Let us consider a physical system composed of :

Quantum mechanics

Let us consider a physical system composed of :

- M nuclei with
  - positions in $\mathbb{R}^3$ are denoted by $R_1, \ldots, R_M \in \mathbb{R}^3$;
  - electric charges are denoted by $Z_1, \ldots, Z_M > 0$.

Quantum mechanics

Let us consider a physical system composed of :

- M nuclei with
  - positions in $\mathbb{R}^3$ are denoted by $R_1, \ldots, R_M \in \mathbb{R}^3$;
  - electric charges are denoted by $Z_1, \ldots, Z_M > 0$.
- N electrons represented by a complex-valued wavefunction $\psi(x_1, \ldots, x_N)$, 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.

Quantum mechanics

Let us consider a physical system composed of :

- M nuclei with
  - positions in $\mathbb{R}^3$ are denoted by $R_1, \ldots, R_M \in \mathbb{R}^3$;
  - electric charges are denoted by $Z_1, \ldots, Z_M > 0$.
- N electrons represented by a complex-valued wavefunction $\psi(x_1, \ldots, x_N)$, 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 : $|\psi(x_1, \cdots, x_N)|^2$ represents the probability density of finding the N electrons at positions $(x_1, \cdots, x_N) \in \mathbb{R}^{3N}$

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

Quantum mechanics

Let us consider a physical system composed of :

- M nuclei with
  - positions in $\mathbb{R}^3$ are denoted by $R_1, \ldots, R_M \in \mathbb{R}^3$;
  - electric charges are denoted by $Z_1, \ldots, Z_M > 0$.
- N electrons represented by a complex-valued wavefunction $\psi(x_1, \ldots, x_N)$, 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 : $|\psi(x_1, \cdots, x_N)|^2$ represents the probability density of finding the N electrons at positions $(x_1, \cdots, x_N) \in \mathbb{R}^{3N}$

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

Electronic (one or two-body) density :

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

Electronic Schrödinger model

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

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

Electronic Schrödinger model

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

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

For a given value of $\mathbf{R}$, the corresponding ground state electronic wavefunction $\psi_{\mathbf{R}}$ is solution to the eigenvalue problem (electronic Schrödinger problem)

$$H_{\mathbf{R}}\psi_{\mathbf{R}} = E_{\mathbf{R}}\psi_{\mathbf{R}}$$

with

$$H_{\mathbf{R}} = -\Delta_{x_1,\ldots,x_N} + \sum_{i=1}^{N} V_{\mathbf{R}}(x_i) + \sum_{1 \le i < j \le N} \frac{1}{|x_i - x_j|}$$

and $E_{\mathbf{R}}$ the lowest eigenvalue of $H_{\mathbf{R}}$.

Electronic Schrödinger model

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

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

For a given value of $\mathbf{R}$, the corresponding ground state electronic wavefunction $\psi_{\mathbf{R}}$ is solution to the eigenvalue problem (electronic Schrödinger problem)

$$H_{\mathbf{R}}\psi_{\mathbf{R}} = E_{\mathbf{R}}\psi_{\mathbf{R}}$$

with

$$H_{\mathbf{R}} = -\Delta_{x_1,\ldots,x_N} + \sum_{i=1}^{N} V_{\mathbf{R}}(x_i) + \sum_{1 \le i < j \le N} \frac{1}{|x_i - x_j|}$$

and $E_{\mathbf{R}}$ the lowest eigenvalue of $H_{\mathbf{R}}$.

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

Model-order reduction : interpolation

Approach : Compute an approximation of $\rho_{\mathbf{R}}$ as an interpolation in a "good sense" from a few well-selected snapshots $\rho_{\mathbf{R}_1}, \ldots, \rho_{\mathbf{R}_n}$.

Model-order reduction : interpolation

Approach : Compute an approximation of $\rho_{\mathbf{R}}$ as an interpolation in a "good sense" from a few well-selected snapshots $\rho_{\mathbf{R}_1}, \ldots, \rho_{\mathbf{R}_n}$.

Important property : If the set of positions of the nuclei $\mathbf{R} = (R_1, \ldots, R_M)$ is shifted by a translation vector b, $(\mathbf{R}' = (R_1 + b, \ldots, R_M + b))$, then the electronic density is also shifted by the same translation vector :

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

Model-order reduction : interpolation

Approach : Compute an approximation of $\rho_{\mathbf{R}}$ as an interpolation in a "good sense" from a few well-selected snapshots $\rho_{\mathbf{R}_1}, \ldots, \rho_{\mathbf{R}_n}$.

Important property : If the set of positions of the nuclei $\mathbf{R} = (R_1, \ldots, R_M)$ is shifted by a translation vector b, $(\mathbf{R}' = (R_1 + b, \ldots, R_M + b))$, then the electronic density is also shifted by the same translation vector :

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

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

Model-order reduction : interpolation

Approach : Compute an approximation of $\rho_{\mathbf{R}}$ as an interpolation in a "good sense" from a few well-selected snapshots $\rho_{\mathbf{R}_1}, \ldots, \rho_{\mathbf{R}_n}$.

Important property : If the set of positions of the nuclei $\mathbf{R} = (R_1, \ldots, R_M)$ is shifted by a translation vector b, $(\mathbf{R}' = (R_1 + b, \ldots, R_M + b))$, then the electronic density is also shifted by the same translation vector :

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

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

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

Model-order reduction : interpolation

Approach : Compute an approximation of $\rho_{\mathbf{R}}$ as an interpolation in a "good sense" from a few well-selected snapshots $\rho_{\mathbf{R}_1}, \ldots, \rho_{\mathbf{R}_n}$.

Important property : If the set of positions of the nuclei $\mathbf{R} = (R_1, \ldots, R_M)$ is shifted by a translation vector b, $(\mathbf{R}' = (R_1 + b, \ldots, R_M + b))$, then the electronic density is also shifted by the same translation vector :

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

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

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

A problem parametrized by the nuclei positions.
Goal : approximate

$$\mathcal{M} := \{\Psi_{\mathbf{R}} \text{ for } \mathbf{R} \in \mathcal{R}\}, \quad \mathcal{R} \text{ being the set of configurations}$$

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

Example on a toy problem in 1D

Eigenvalue problem :

$$\begin{cases} -\frac{1}{2}\Psi_{\mathbf{R}}'' + \left(-\sum_{m=1}^{M} z_m \delta_{R_m}\right)\Psi_{\mathbf{R}} & = E_{\mathbf{R}}\Psi_{\mathbf{R}} \\ \|\Psi_{\mathbf{R}}\|_{L^2(\mathbb{R})} & = 1. \end{cases}$$

Example on a toy problem in 1D

Eigenvalue problem :

$$\begin{cases} -\frac{1}{2}\Psi_{\mathbf{R}}'' + \left(-\sum_{m=1}^{M} z_m \delta_{R_m}\right)\Psi_{\mathbf{R}} & = E_{\mathbf{R}}\Psi_{\mathbf{R}} \\ \|\Psi_{\mathbf{R}}\|_{L^2(\mathbb{R})} & = 1. \end{cases}$$

- Dirac potential

Example on a toy problem in 1D

Eigenvalue problem :

$$\begin{cases} -\frac{1}{2}\Psi_{\mathbf{R}}'' + \left(-\sum_{m=1}^{M} z_m \delta_{R_m}\right)\Psi_{\mathbf{R}} & = E_{\mathbf{R}}\Psi_{\mathbf{R}} \\ \|\Psi_{\mathbf{R}}\|_{L^2(\mathbb{R})} & = 1. \end{cases}$$

- Dirac potential
- Similar regularity as in the 3D case with Coulomb

Example on a toy problem in 1D

Eigenvalue problem :

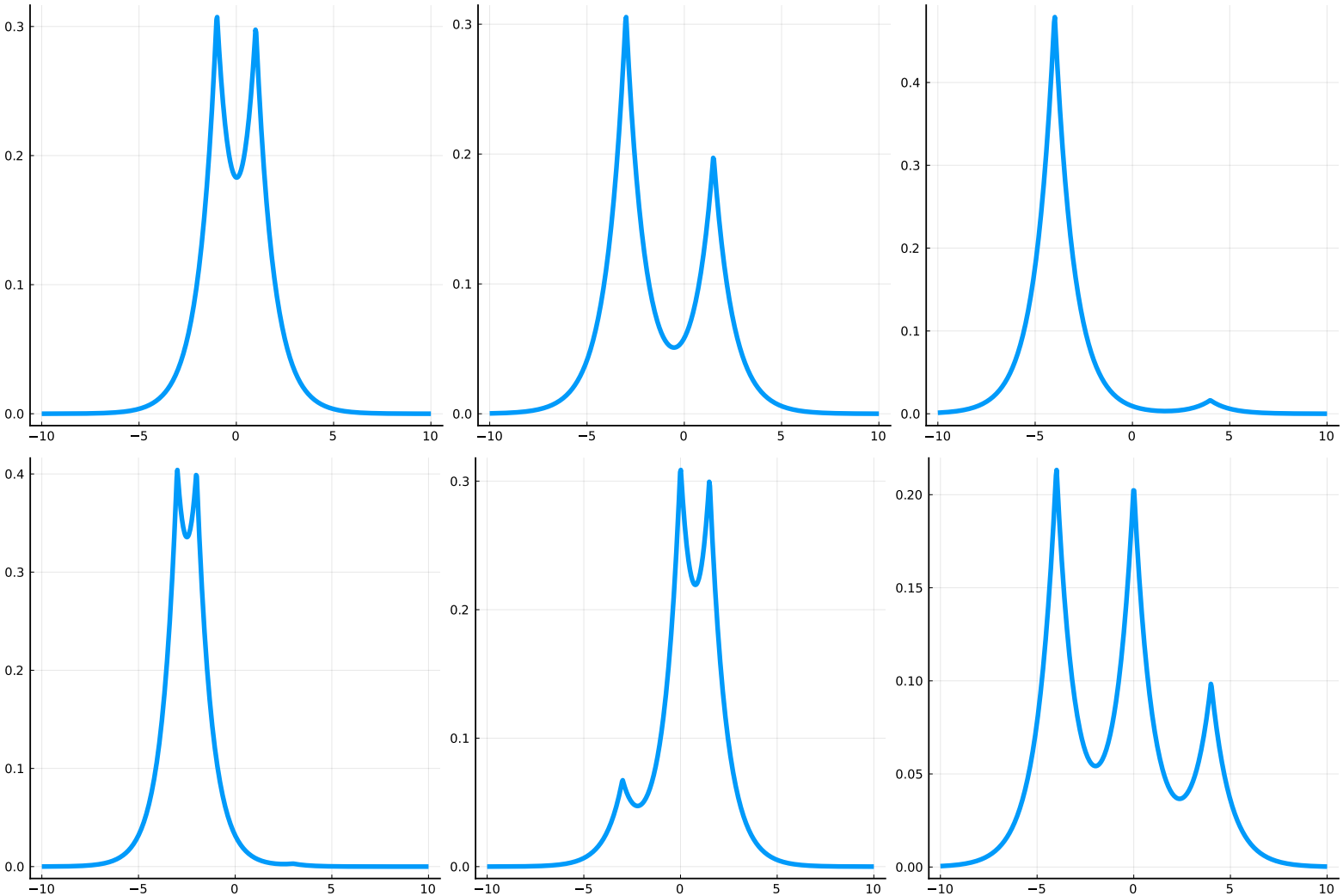
$$\begin{cases} -\frac{1}{2}\Psi_{\mathbf{R}}'' + \left(-\sum_{m=1}^{M} z_m \delta_{R_m}\right)\Psi_{\mathbf{R}} & = E_{\mathbf{R}}\Psi_{\mathbf{R}} \\ \|\Psi_{\mathbf{R}}\|_{L^2(\mathbb{R})} & = 1. \end{cases}$$

- Dirac potential
- Similar regularity as in the 3D case with Coulomb
- Analytic solutions

$$\Psi_{\mathbf{R}} = \sum_{m=1}^{M} \pi_m^{\mathbf{R}} e^{-\zeta_{\mathbf{R}}|x-R_m|},$$

for some positive weights $\boldsymbol{\pi}^{\mathbf{R}} = \left(\pi_m^{\mathbf{R}}\right)_{m=1}^{M} \in (\mathbf{R}_+)^M$ and $\zeta_{\mathbf{R}} > 0$.

Plots of a few solutions



Aim : efficiently approximate all solutions (for varying positions) from the computation of only **a few** solutions

Outline

Linear reduced-order modelling

Optimal transport : a few properties

Practical strategy and numerical results for a 1D toy problem

Linear reduced basis method

Problem : Parametrized PDE with parameters $\mathbf{R} \in \mathcal{R}$.

Needs to be solved for **many** parameters $\mathbf{R}$.

Linear reduced basis method

Problem : Parametrized PDE with parameters $\mathbf{R} \in \mathcal{R}$.

Needs to be solved for **many** parameters $\mathbf{R}$.

Offline part : Select accurate solutions for a few **wisely chosen** parameters $\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_K \in \mathcal{R}$

Linear reduced basis method

Problem : Parametrized PDE with parameters $\mathbf{R} \in \mathcal{R}$.

Needs to be solved for **many** parameters $\mathbf{R}$.

Offline part : Select accurate solutions for a few **wisely chosen** parameters $\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_K \in \mathcal{R}$

- Generate a training set of snapshots for parameters $\mathbf{R} \in \mathcal{R}_{\text{train}}$

Linear reduced basis method

Problem : Parametrized PDE with parameters $\mathbf{R} \in \mathcal{R}$.

Needs to be solved for **many** parameters $\mathbf{R}$.

Offline part : Select accurate solutions for a few **wisely chosen** parameters $\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_K \in \mathcal{R}$

- Generate a training set of snapshots for parameters $\mathbf{R} \in \mathcal{R}_{\text{train}}$
- Select "good" snapshots for instance with a greedy algorithm (Reduced Basis techniques)
  - Select one parameter $\mathbf{R}_1 \in \mathcal{R}$
  - at each iteration $K \geq 2$, select the snapshot that is worse approximated in the basis of the previously selected snapshots $\Psi_{\mathbf{R}_1}, \ldots, \Psi_{\mathbf{R}_{K-1}}$

Linear reduced basis method

Problem : Parametrized PDE with parameters $\mathbf{R} \in \mathcal{R}$.

Needs to be solved for **many** parameters $\mathbf{R}$.

Offline part : Select accurate solutions for a few **wisely chosen** parameters $\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_K \in \mathcal{R}$

- Generate a training set of snapshots for parameters $\mathbf{R} \in \mathcal{R}_{\text{train}}$
- Select "good" snapshots for instance with a greedy algorithm (Reduced Basis techniques)
  - Select one parameter $\mathbf{R}_1 \in \mathcal{R}$
  - at each iteration $K \geq 2$, select the snapshot that is worse approximated in the basis of the previously selected snapshots $\Psi_{\mathbf{R}_1}, \ldots, \Psi_{\mathbf{R}_{K-1}}$

Online part : compute solutions for **many** parameters $\mathbf{R} \in \mathcal{R}$ in the reduced space spanned by selected snapshots, i.e. in the basis spanned by the solutions $\Psi_{\mathbf{R}_1}, \ldots, \Psi_{\mathbf{R}_K}$.

Barrault, Maday, Nguyen, Patera : An empirical interpolation method : application to efficient reduced-basis discretization of partial differential equations. C. R.(2004)

Hesthaven, Rozza, Stamm : Certified Reduced Basis Methods for Parametrized Partial Differential Equations. Springer (2016)

Kolmogorov n-width

Definition for a Hilbert space $\mathbb{H}$: $\mathcal{M} := \{\Psi_{\mathbf{R}}, \quad \mathbf{R} \in \mathcal{R}\}$

$$d_n(\mathcal{M}, \mathbb{H}) := \inf_{\substack{V_n \subset \mathbb{H} \\ \dim V_n = n}} \sup_{\mathbf{R} \in \mathcal{R}} \|\Psi_{\mathbf{R}} - \mathrm{P}_{V_n} \Psi_{\mathbf{R}}\|.$$

Kolmogorov n-width

Definition for a Hilbert space $\mathbb{H}$: $\mathcal{M} := \{\Psi_{\mathbf{R}}, \quad \mathbf{R} \in \mathcal{R}\}$

$$d_n(\mathcal{M}, \mathbb{H}) := \inf_{\substack{V_n \subset \mathbb{H} \\ \dim V_n = n}} \sup_{\mathbf{R} \in \mathcal{R}} \|\Psi_{\mathbf{R}} - \mathrm{P}_{V_n} \Psi_{\mathbf{R}}\|.$$

- Characterizes if the reduced basis method has a chance to work
- The faster the decay, the better !

Kolmogorov n-width

Definition for a Hilbert space $\mathbb{H}$: $\mathcal{M} := \{\Psi_{\mathbf{R}}, \quad \mathbf{R} \in \mathcal{R}\}$

$$d_n(\mathcal{M}, \mathbb{H}) := \inf_{\substack{V_n \subset \mathbb{H} \\ \dim V_n = n}} \sup_{\mathbf{R} \in \mathcal{R}} \|\Psi_{\mathbf{R}} - \mathrm{P}_{V_n} \Psi_{\mathbf{R}}\|.$$

- Characterizes if the reduced basis method has a chance to work
- The faster the decay, the better !

Typical example where it works : elliptic equation

$$A_{\mathbf{R}} \Psi_{\mathbf{R}} = f,$$

with affine representation of $A_{\mathbf{R}}$:

$$A_{\mathbf{R}} = \sum_{q=1}^{Q} \theta_q(\mathbf{R}) A_q, \quad \text{for some } \theta_q \in \mathbb{R},\ A_q \text{ continuous operators}$$

Exponential decay of the Kolmogorov n-width :

$$d_n(\mathcal{M}, \mathbb{H}) \leq C\ exp(-cn^{1/Q}).$$

Ohlberger, Rave : Reduced Basis Methods : Success, Limitations and Future Challenges (2015)

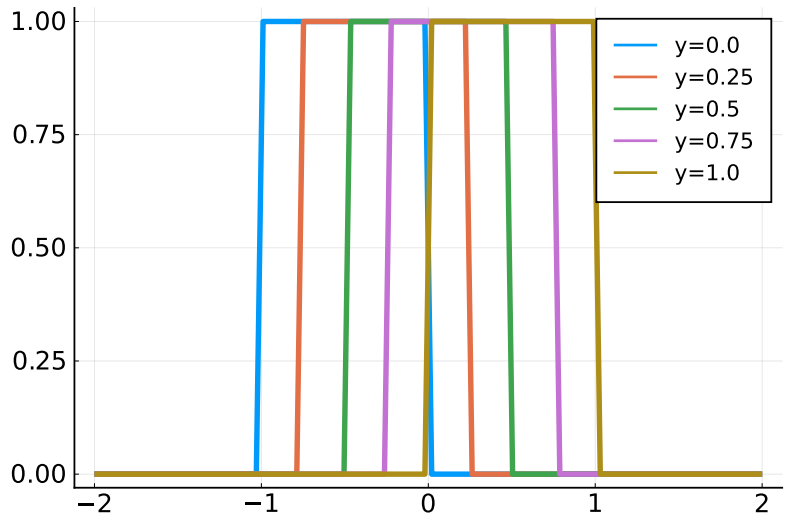
Cohen, DeVore : Approximation of high-dimensional parametric PDEs Acta Numerica (2015)

Less successful examples : Transport problems

Simple example : one-dimensional transport equation, $y \in [0,1]$,

$$\begin{cases} \partial_t \Psi_y(t,x) + y\partial_x \Psi_y(t,x) = 0, \quad x \in \mathbb{R}, t \in \mathbb{R}_+ \\ \Psi_y(0,x) = \mathbf{1}_{[0,1]} \end{cases}$$

At $t = 1$, the solutions are $\Psi_y(t=1,x) = \mathbf{1}_{[y-1,y]}$



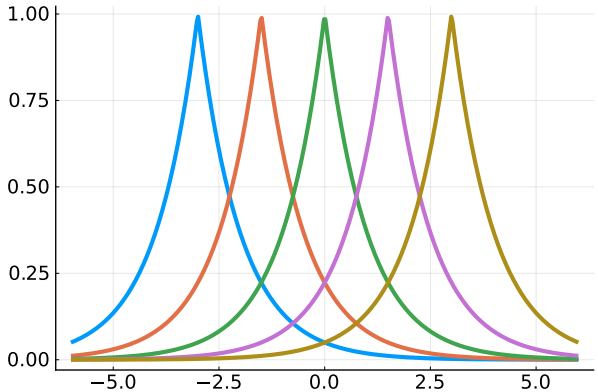
Kolmogorov n-width for $\mathcal{M} := \{\mathbf{1}_{[y-1,y]}, \quad y \in [0,1]\}$:

$$d_n(\mathcal{M}, L^2(\Omega)) \geq cn^{-1/2}$$

DeVore (1998) Ohlberger, Rave (2015)

What about electronic structure ?

Similar behavior as the transport problem



Kolmogorov n-width for the 1D toy problem

For the problem with one nucleus, $\mathcal{M} = \{\Psi_R, R \in [-\bar{R}, \bar{R}]\}$, there exist positive constant $c_{\bar{R}}, C_{\bar{R}}$ such that

$$c_{\bar{R}} n^{-\frac{3}{2}} \leqslant d_n(\mathcal{M}, L^2(\mathbb{R})) \leqslant C_{\bar{R}} n^{-\frac{3}{2}}.$$

For the problem with two nuclei, $\mathcal{M} = \{\Psi_{(R_1,R_2)}, R_1, R_2 \in [-\bar{R}, \bar{R}]^2\}$, there exists a positive constant $c_{\bar{R}}$ such that

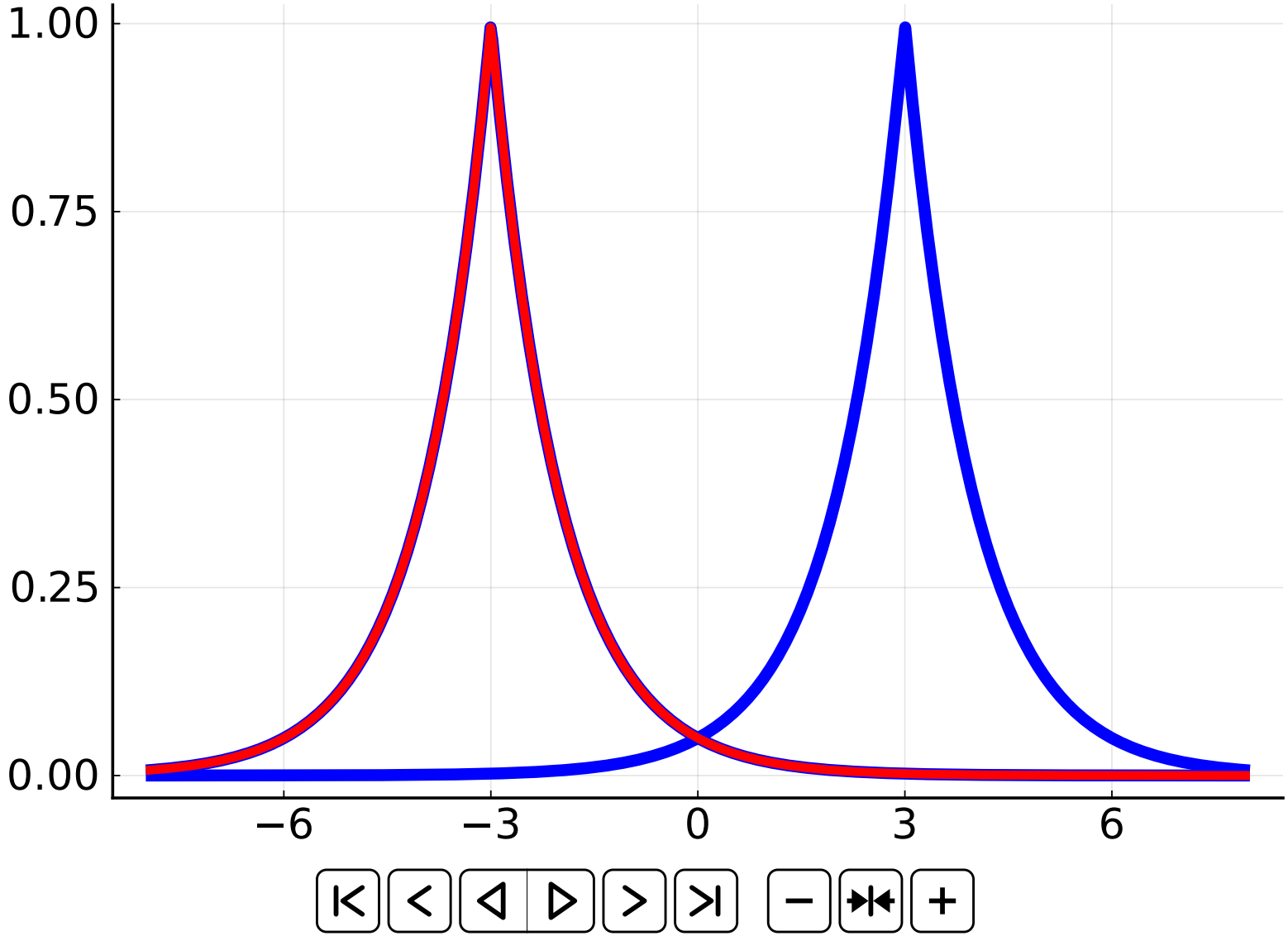
$$c_{\bar{R}} n^{-\frac{3}{2}} \leqslant d_n(\mathcal{M}, L^2(\mathbb{R})).$$

Dalery, Dusson, Ehrlacher, Lozinski : Nonlinear reduced basis using mixture Wasserstein barycenters : application to an eigenvalue problem inspired from quantum chemistry, https ://arxiv.org/abs/2307.15423

Alternative : finding a good nonlinear transformation

Use of **optimal transport** : at minima deals with the translations

Barycenter between two Slater functions : a translated Slater function



Outline

Linear reduced-order modelling

Optimal transport : a few properties

Practical strategy and numerical results for a 1D toy problem

Optimal transport in a nutshell

Let $\mathcal{P}_2(\mathbb{R})$ denotes the set of probability measures on $\mathbb{R}$ with finite second-order moments.

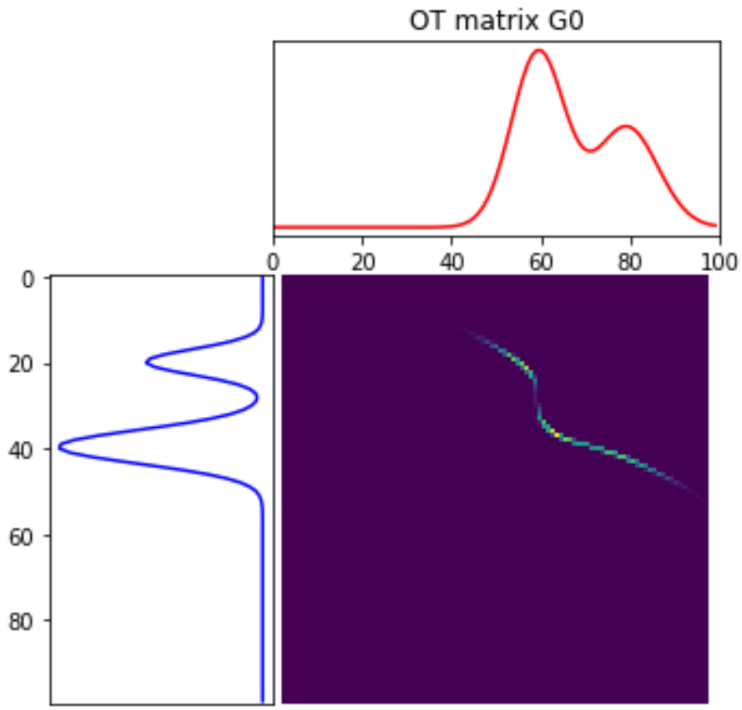
Optimal transport in a nutshell

Let $\mathcal{P}_2(\mathbb{R})$ denotes the set of probability measures on $\mathbb{R}$ with finite second-order moments.

Wasserstein distance : The 2-Wasserstein distance over $\mathcal{P}_2(\mathbb{R})$ is defined for $u, v \in \mathcal{P}_2(\mathbb{R})^2$ as

$$W_2(u,v)^2 := \inf_{\pi \in \Pi(u,v)} \int_{\mathbb{R}^2} (x-y)^2 \, d\pi(x,y),$$

$\Pi(u,v)$: set of probability measures over $\mathbb{R}^2$ with marginals u and v.



Delon, Desolneux : A Wasserstein-Type Distance in the Space of Gaussian Mixture Models. SIAM J. Imaging Sci. (2020).

Wasserstein barycenters

Define

$$T_Q = \left\{ (t_1, \dots, t_Q) \in (\mathbb{R}_+)^Q, \quad \sum_{q=1}^{Q} t_q = 1 \right\}.$$

The **Wasserstein barycenter** $\mathrm{Bar}^{\boldsymbol{t}}_{W_2}(\boldsymbol{u})$ of a collection of Q probability measures $\boldsymbol{u} := (u_1, \dots, u_Q) \in \mathcal{P}_2(\mathbb{R})^Q$ associated to a set of barycentric weights $\boldsymbol{t} := (t_1, \dots, t_Q) \in T_Q$: it is the unique solution to the problem

$$\inf_{u \in \mathcal{P}_2(\mathbb{R})} \sum_{q=1}^{Q} t_q W_2(u, u_q)^2.$$

Wasserstein barycenters

Define
$$T_Q = \left\{ (t_1, \dots, t_Q) \in (\mathbb{R}_+)^Q, \quad \sum_{q=1}^{Q} t_q = 1 \right\}.$$

The **Wasserstein barycenter** $\mathrm{Bar}^{\boldsymbol{t}}_{W_2}(\boldsymbol{u})$ of a collection of Q probability measures $\boldsymbol{u} := (u_1, \dots, u_Q) \in \mathcal{P}_2(\mathbb{R})^Q$ associated to a set of barycentric weights $\boldsymbol{t} := (t_1, \dots, t_Q) \in T_Q$: it is the unique solution to the problem
$$\inf_{u \in \mathcal{P}_2(\mathbb{R})} \sum_{q=1}^{Q} t_q W_2(u, u_q)^2.$$

Remark : The L^2 **barycenter** of a collection of Q probability measures $\boldsymbol{u} := (u_1, \dots, u_Q) \in \mathcal{P}_2(\mathbb{R})^Q$ associated to a set of barycentric weights $\boldsymbol{t} := (t_1, \dots, t_Q) \in T_Q$: it is the unique solution to the problem
$$\inf_{u \in L^2(\mathbb{R})} \sum_{q=1}^{Q} t_q \|u - u_q\|_{L^2}^2, \qquad \text{i.e.} \qquad \sum_{q=1}^{Q} t_q u_q$$

Agueh, Carlier : Barycenters in the Wasserstein Space. SIAM J. Math. Anal. (2011).

Gangbo, Swiech : Optimal maps for the multidimensional Monge–Kantorovich problem. Commun. Pure Appl. Math. (1998)

A few examples : One-dimensional case

Cumulative distribution function (cdf) of an element $u \in \mathcal{P}_2(\mathbb{R})$ is

$$\mathrm{cdf}_u : x \in \mathbb{R} \longmapsto \int_{-\infty}^{x} \mathrm{d}[u],$$

A few examples : One-dimensional case

Cumulative distribution function (cdf) of an element $u \in \mathcal{P}_2(\mathbb{R})$ is

$$\mathrm{cdf}_u : x \in \mathbb{R} \longmapsto \int_{-\infty}^{x} \mathrm{d}[u],$$

Inverse cumulative distribution function (icdf) : generalized inverse of the cdf

$$\mathrm{icdf}_u : s \in [0,1] \longmapsto \mathrm{cdf}_u^{-1} := \inf\{x \in \mathbb{R}, \mathrm{cdf}_u(x) > s\}.$$

Then, for any $(u, v) \in \mathcal{P}_2(\mathbb{R})^2$, there holds

$$W_2(u, v) = \|\mathrm{icdf}_u - \mathrm{icdf}_v\|_{L^2([0,1])},$$

and for any set of barycentric weights $\boldsymbol{t} := (t_1, \ldots, t_Q) \in T_Q$ and $\boldsymbol{u} := (u_1, \ldots, u_Q) \in \mathcal{P}_2(\mathbb{R})^Q$,

A few examples : One-dimensional case

Cumulative distribution function (cdf) of an element $u \in \mathcal{P}_2(\mathbb{R})$ is

$$\mathrm{cdf}_u : x \in \mathbb{R} \longmapsto \int_{-\infty}^{x} \mathrm{d}[u],$$

Inverse cumulative distribution function (icdf) : generalized inverse of the cdf

$$\mathrm{icdf}_u : s \in [0,1] \longmapsto \mathrm{cdf}_u^{-1} := \inf\{x \in \mathbb{R}, \mathrm{cdf}_u(x) > s\}.$$

Then, for any $(u, v) \in \mathcal{P}_2(\mathbb{R})^2$, there holds

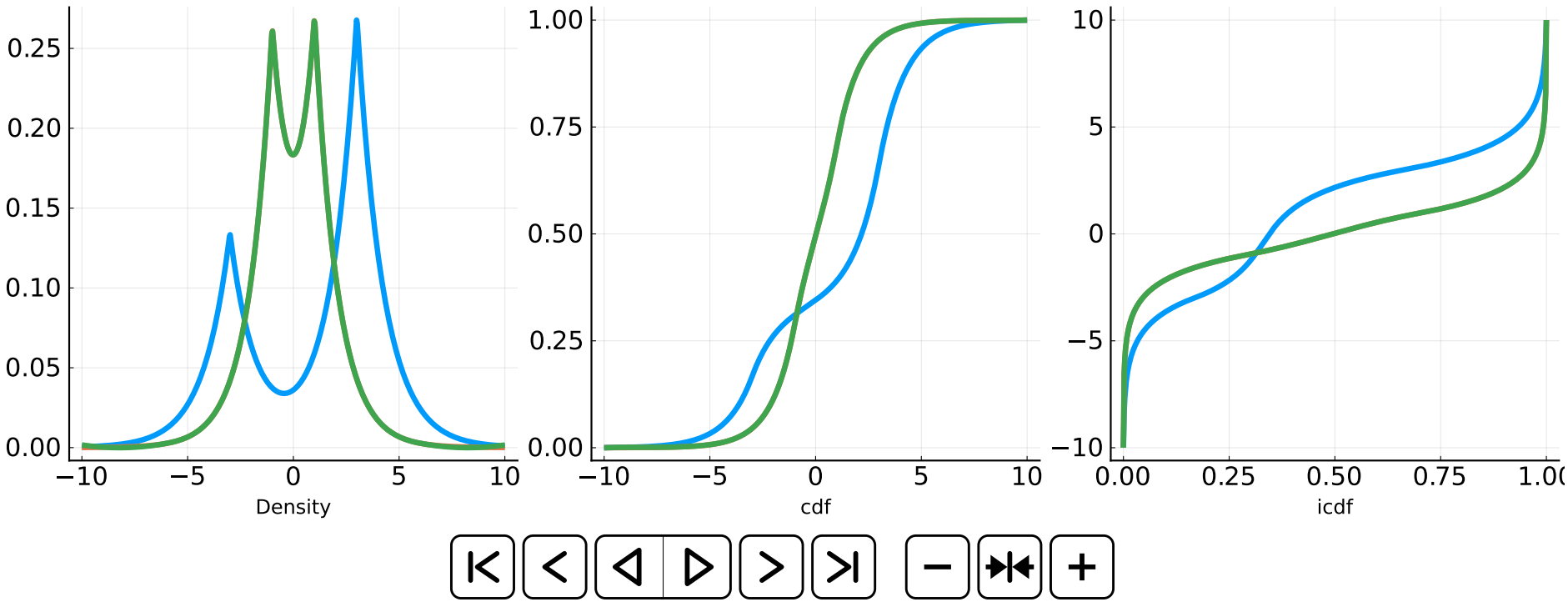
$$W_2(u, v) = \|\mathrm{icdf}_u - \mathrm{icdf}_v\|_{L^2([0,1])},$$

and for any set of barycentric weights $\boldsymbol{t} := (t_1, \ldots, t_Q) \in T_Q$ and $\boldsymbol{u} := (u_1, \ldots, u_Q) \in \mathcal{P}_2(\mathbb{R})^Q$,

the icdf of the barycenter $\mathrm{Bar}_{W_2}^{\boldsymbol{t}}(\boldsymbol{u})$ satisfies

$$\mathrm{icdf}_{\mathrm{Bar}_{W_2}^{\boldsymbol{t}}(\boldsymbol{u})} = \sum_{q=1}^{Q} t_n \mathrm{icdf}_{u_q}.$$

Illustration



$$\mathrm{icdf}_{\mathrm{Bar}^{\boldsymbol{t}}_{W_2}(\boldsymbol{u})} = \lambda\,\mathrm{icdf}_{\boldsymbol{u}_1} + (1-\lambda)\,\mathrm{icdf}_{\boldsymbol{u}_2}.$$

A few examples : Location-scatter transforms

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

All measures generated with translation and dilations of a single measure

A few examples : Location-scatter transforms

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

All measures generated with translation and dilations of a single measure

Here, a will be given as a Slater distribution $da(x) = \frac{1}{Z} e^{-|x|}\, dx$

A few examples : Location-scatter transforms

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

All measures generated with translation and dilations of a single measure

Here, a will be given as a Slater distribution $da(x) = \frac{1}{Z} e^{-|x|}\, dx$

Wasserstein distance : Let $a_0, a_1 \in \mathcal{A}$ havings means m_0, m_1 and covariance matrices Σ_0, Σ_1. Then

$$W_2^2(a_0, a_1) = \|m_0 - m_1\|^2 + \mathrm{Tr}\left(\Sigma_0 + \Sigma_1 - 2(\Sigma_0^{1/2}\Sigma_1\Sigma_0^{1/2})^{1/2}\right),$$

A few examples : Location-scatter transforms

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

All measures generated with translation and dilations of a single measure

Here, a will be given as a Slater distribution $da(x) = \frac{1}{Z} e^{-|x|}\, dx$

Wasserstein distance : Let $a_0, a_1 \in \mathcal{A}$ havings means m_0, m_1 and covariance matrices Σ_0, Σ_1. Then

$$W_2^2(a_0, a_1) = \|m_0 - m_1\|^2 + \mathrm{Tr}\left(\Sigma_0 + \Sigma_1 - 2(\Sigma_0^{1/2}\Sigma_1\Sigma_0^{1/2})^{1/2}\right),$$

Wasserstein barycenter : Let $a_1, \ldots, a_Q \in \mathcal{A}$ such that for all $1 \leq q \leq Q$, a_q has mean m_q and covariance matrix Σ_q. For weights $\boldsymbol{t} := (t_1, \ldots, t_Q) \in T_Q$, the Wassertein barycenter belongs to $\mathcal{A}$ with mean m and covariance matrix S given by

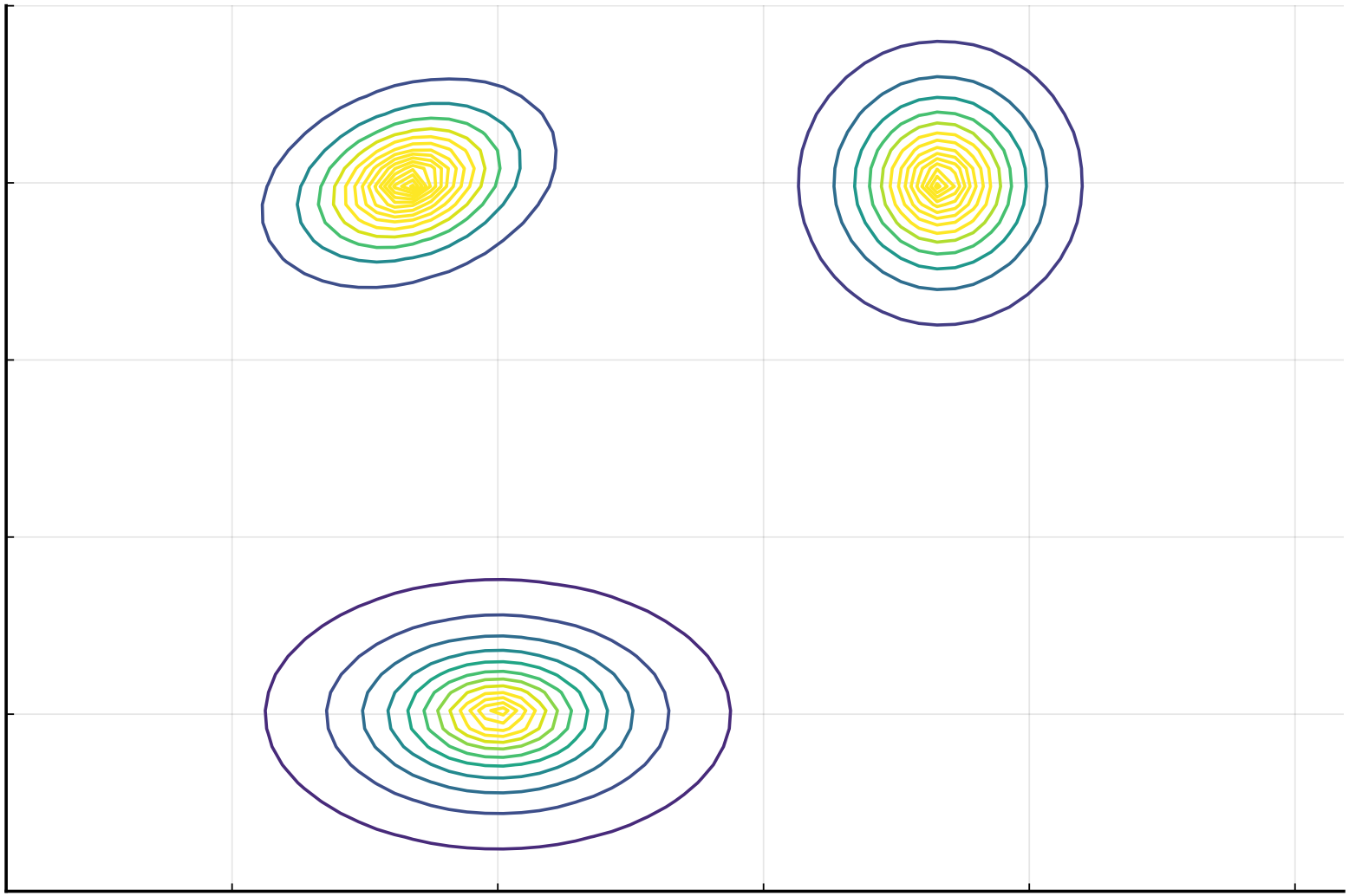
$$S = \sum_{q=1}^{Q} t_q (S^{1/2}\Sigma_q S^{1/2})^{1/2}, \quad \text{and} \quad m = \sum_{q=1}^{Q} t_q m_q.$$

Center is the mean of centers, small equation to solve for the covariance.

Alvarez-Esteban, del Barrio, Cuesta-Albertos, Matran : A fixed-point approach to barycenters in Wasserstein space. (2016).

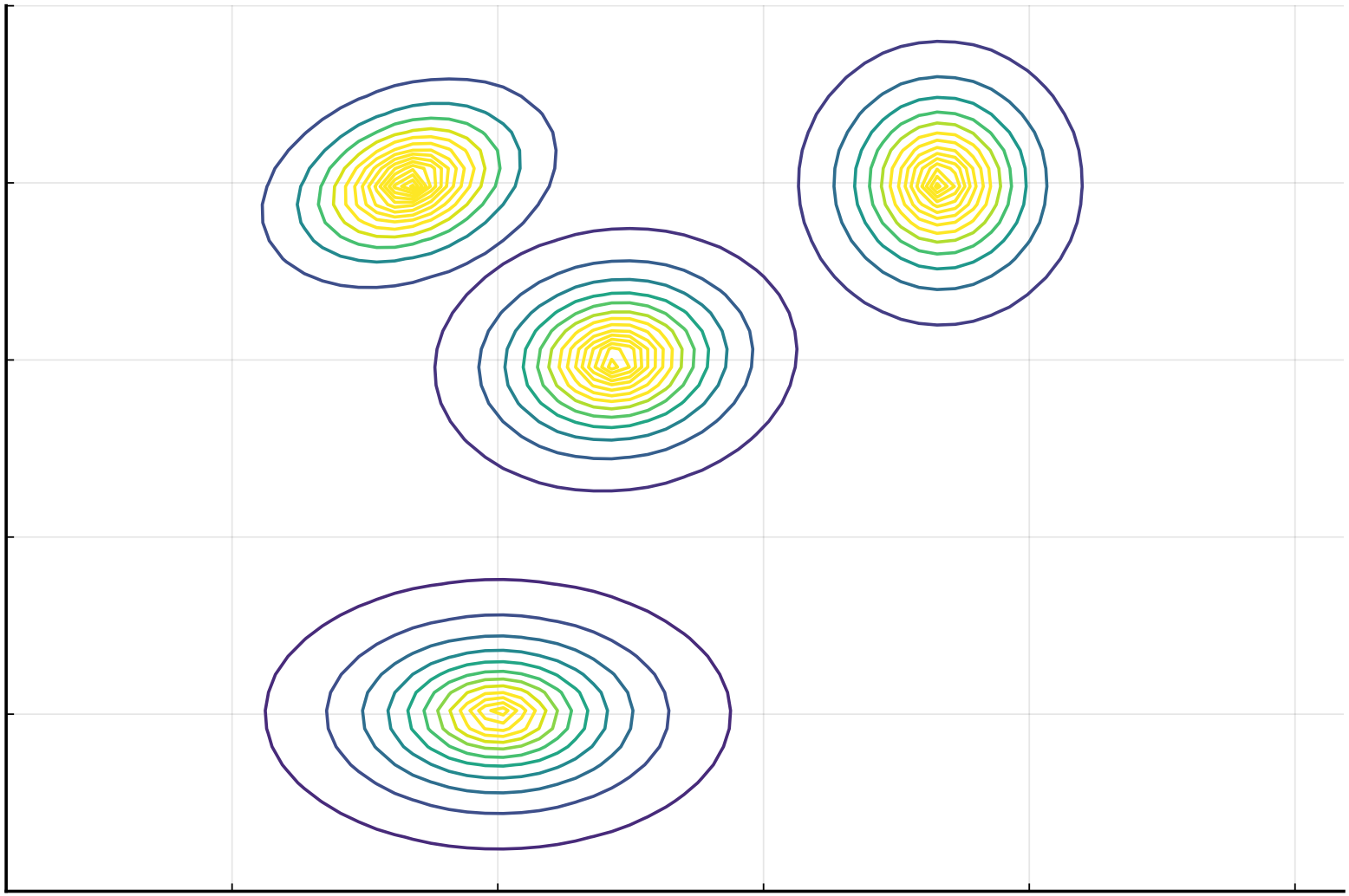
Illustration

Barycenter between three Slater distributions



Illustration

Barycenter between three Slater distributions



Kolmogorov n-width for the Wasserstein distance

Solution manifold : $\mathcal{M} = \{\Psi_{\mathbf{R}}, \mathbf{R} \in [-\bar{R}, \bar{R}]^M\}$

Since $W_2(u, v) = \|\mathrm{icdf}_u - \mathrm{icdf}_v\|_{L^2(0,1)}$, we consider

$$d_n(\mathcal{M}, W_2) := d_n(\mathrm{icdf}(\mathcal{M}), L^2(0, 1))$$

Kolmogorov n-width for the Wasserstein distance

Solution manifold : $\mathcal{M} = \{\Psi_{\mathbf{R}}, \mathbf{R} \in [-\bar{R}, \bar{R}]^M\}$

Since $W_2(u, v) = \|\mathrm{icdf}_u - \mathrm{icdf}_v\|_{L^2(0,1)}$, we consider

$$d_n(\mathcal{M}, W_2) := d_n(\mathrm{icdf}(\mathcal{M}), L^2(0, 1))$$

Simple one-nucleus problem : $\forall n > 1, d_n(\mathcal{M}, W_2) = 0$

Kolmogorov n-width for the Wasserstein distance

Solution manifold : $\mathcal{M} = \{\Psi_{\mathbf{R}}, \mathbf{R} \in [-\bar{R}, \bar{R}]^M\}$

Since $W_2(u, v) = \|\mathrm{icdf}_u - \mathrm{icdf}_v\|_{L^2(0,1)}$, we consider

$$d_n(\mathcal{M}, W_2) := d_n(\mathrm{icdf}(\mathcal{M}), L^2(0, 1))$$

Simple one-nucleus problem : $\forall n > 1, d_n(\mathcal{M}, W_2) = 0$

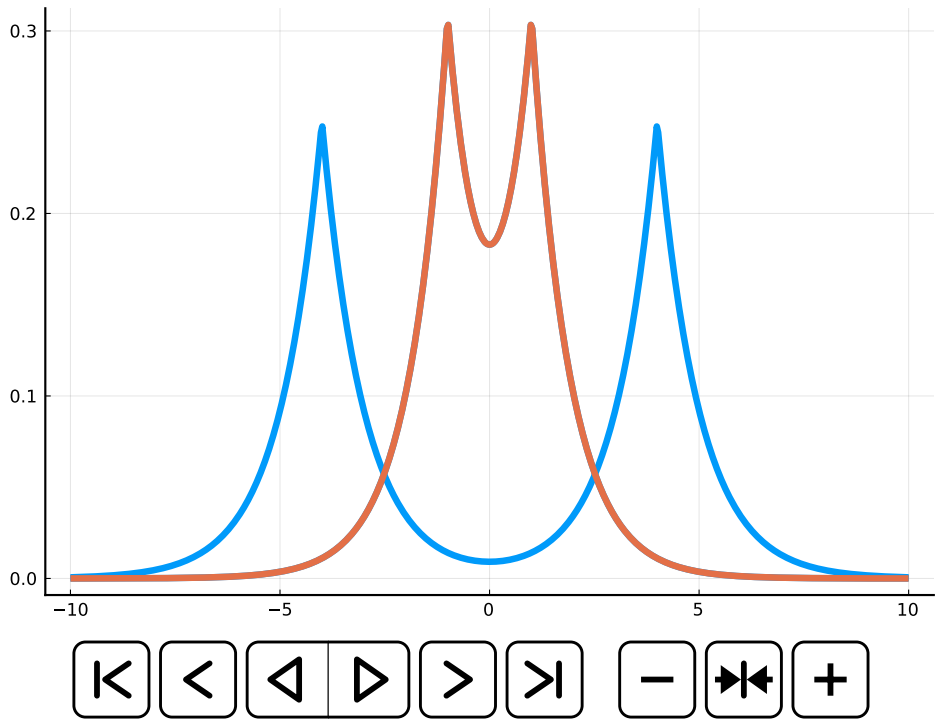
Case of two nuclei in $[-\bar{R}, \bar{R}]^2$ **:** There exists a constant $C_{\bar{R}} > 0$ such that for all $n \geq 1$,

$$d_n(\mathcal{M}, W_2) \leq C_{\bar{R}} n^{-5/2}.$$

To compare with the linear Kolmogorov width :

$$c_{\bar{R}} n^{-3/2} \leqslant d_n(\mathcal{M}, L^2(\mathbb{R}))$$

Wasserstein barycenter between two solutions



Limitations :

- High computational cost
- Smoothing of the barycenter with Sinkhorn algorithm
- Bad scaling with the dimension
- Approximation properties of the solution set not optimal

A modified distance

$\mathcal{A} \subset \mathcal{P}(\mathbb{R})$: set of location-scatter probability measures (Slater functions, gaussians, etc.)

Modified distance between two mixtures : $\lambda_0^j \geq 0$, $\sum_{j=1}^J \lambda_0^j = 1$, $\lambda_1^k \geq 0$, $\sum_{k=1}^K \lambda_1^k = 1$

$$\mu_0 = \sum_{j=1}^{J} \lambda_0^j a_0^j \quad \text{and } \mu_1 = \sum_{k=1}^{K} \lambda_1^k a_1^k$$

A modified distance

$\mathcal{A} \subset \mathcal{P}(\mathbb{R})$: set of location-scatter probability measures (Slater functions, gaussians, etc.)

Modified distance between two mixtures : $\lambda_0^j \geq 0$, $\sum_{j=1}^J \lambda_0^j = 1$, $\lambda_1^k \geq 0$, $\sum_{k=1}^K \lambda_1^k = 1$

$$\mu_0 = \sum_{j=1}^J \lambda_0^j a_0^j \quad \text{and } \mu_1 = \sum_{k=1}^K \lambda_1^k a_1^k$$

$$mW_2(\mu_0, \mu_1)^2 := \min_{w:=(w_{jk})_{\substack{1\leq j\leq J,\\ 1\leq k\leq K}} \in \Pi(\Lambda_0,\Lambda_1)} \sum_{j=1}^J \sum_{k=1}^K w_{jk} W_2^2(a_0^j, a_1^k),$$

$$\text{with} \quad \Pi(\Lambda_0, \Lambda_1) := \Bigg\{ w := (w_{jk})_{1\leq j\leq J, 1\leq k\leq K} \in \mathbb{R}_+^{J\times K},$$

$$\forall 1 \leq j \leq J, \ \sum_{k=1}^K w_{jk} = \lambda_0^j, \quad \forall 1 \leq k \leq K, \ \sum_{j=1}^J w_{jk} = \lambda_1^k \Bigg\}.$$

A modified distance

$\mathbb{M}$ set of mixtures of location-scatter probability measures obtained from the Slater distribution

For the 1D toy model test case for electronic structure calculations,

$$\mathcal{M} \subset \mathbb{M}$$

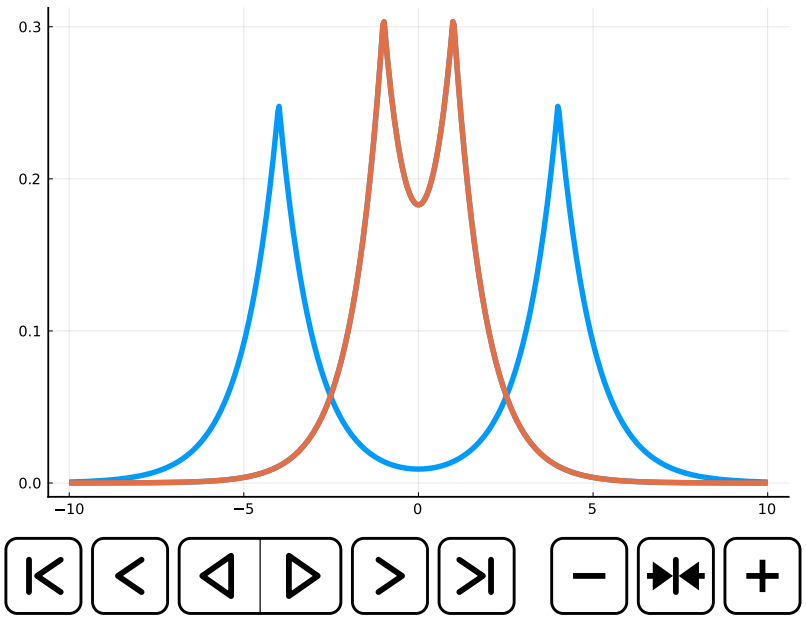
Aim : Exploit the compact form of the solution as a mixture of Slater functions

Delon, Desolneux : A Wasserstein-Type Distance in the Space of Gaussian Mixture Models. SIAM J. Imaging Sci. (2020).
Dusson, Ehrlacher, Nouaime : A Wasserstein-type metric for generic mixture models, including location-scatter and group invariant measures (2023)

Mixture barycenter between two solutions

Mixture Wasserstein barycenter of a collection of Q probability measures $\boldsymbol{u} := (u_1, \ldots, u_Q) \in \mathcal{P}_2(\mathbb{R})^Q$ associated to a set of barycentric weights $\boldsymbol{t} := (t_1, \ldots, t_Q) \in T_Q$: it is the unique solution to the problem

$$\inf_{u \in \mathcal{M}(\mathcal{A})} \sum_{q=1}^{Q} t_q m W_2^2(u, u_q)^2.$$



Interesting features :

- Way better approximation compared to W_2 barycenter
- Computational cost independent of the dimension

Formula for the barycenter :

$$\mathrm{Bar}^{\boldsymbol{t}}_{\mathrm{MW}_2}(u_1, \ldots, u_Q) = \sum_{\mathbf{k} \in \mathbf{K}} w^*_{\mathbf{k}} \, \mathrm{Bar}^{\boldsymbol{t}}_{W_2}\left(u_1^{k^1}, \ldots, u_Q^{k^Q}\right),$$

A few properties

Valid for a large number of probability distributions

Mathematically :

- needs a geodesic space for the atoms (space with distance + geodesic)
- identifiability

A few properties

Valid for a large number of probability distributions

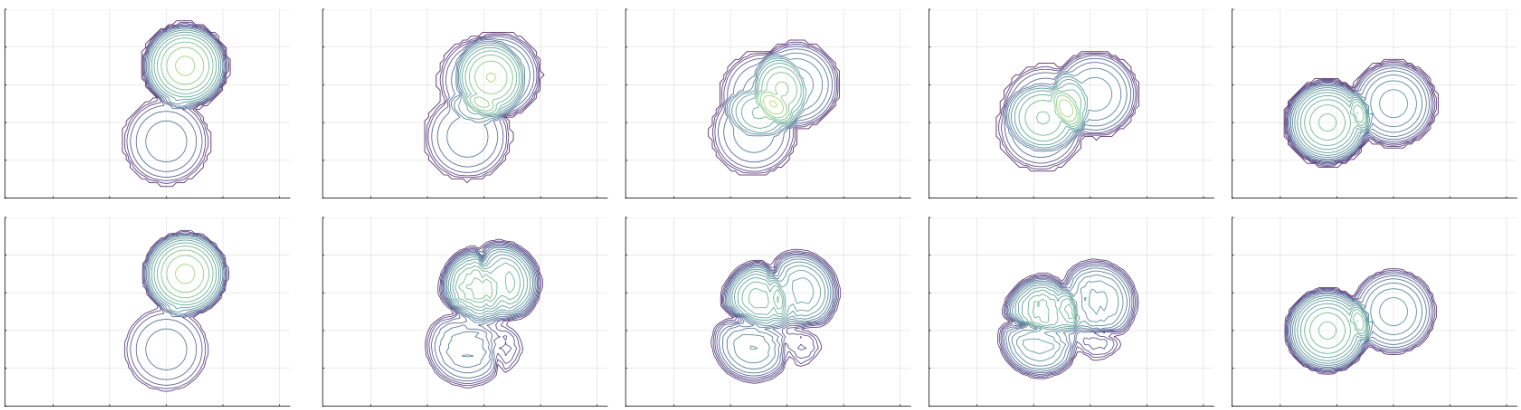
Mathematically :

- needs a geodesic space for the atoms (space with distance + geodesic)
- identifiability

Computationally : barycenters need to be easily computable (best if explicit !)

Examples :

- Elliptic distributions (Slater, gaussians, Wigner semicircle)
- Location-scatter (dilations+translations)



(a) $t = 0$ (b) $t = 0.25$ (c) $t = 0.5$ (d) $t = 0.75$ (e) $t = 1$

Dusson, Ehrlacher, Nouaime : A Wasserstein-type metric for generic mixture models, including location-scatter and group invariant measures (2023)

Back to Kolmogorov n-width

Definition of nonlinear Kolmogorov n-width :
The Kolmogorov n-width of the set $\mathcal{M} \subset \mathbb{M}$ is defined by

$$d_n(\mathcal{M}, \mathbb{M}) = \inf_{\boldsymbol{u} \in \mathbb{M}^n} \sup_{\mathbf{R} \in \mathcal{R}} \inf_{\boldsymbol{t} \in T_n} mW_2(\Psi_{\mathbf{R}}, \mathrm{Bar}^{\boldsymbol{t}}_{\mathrm{MW}_2}(\boldsymbol{u})).$$

Theorem : (Dalery, Dusson, E., Lozinski) for a system with two nuclei with identical charges, for $n \geq 2$,

$$d_n(\mathcal{M}, \mathbb{M}) = 0.$$

- Exact representation of all solutions in this case

Dalery, Dusson, Ehrlacher, Lozinski : Nonlinear reduced basis using mixture Wasserstein barycenters : application to an eigenvalue problem inspired from quantum chemistry

Outline

Linear reduced-order modelling

Optimal transport : a few properties

Practical strategy and numerical results for a 1D toy problem

Practical strategy

Nonlinear version

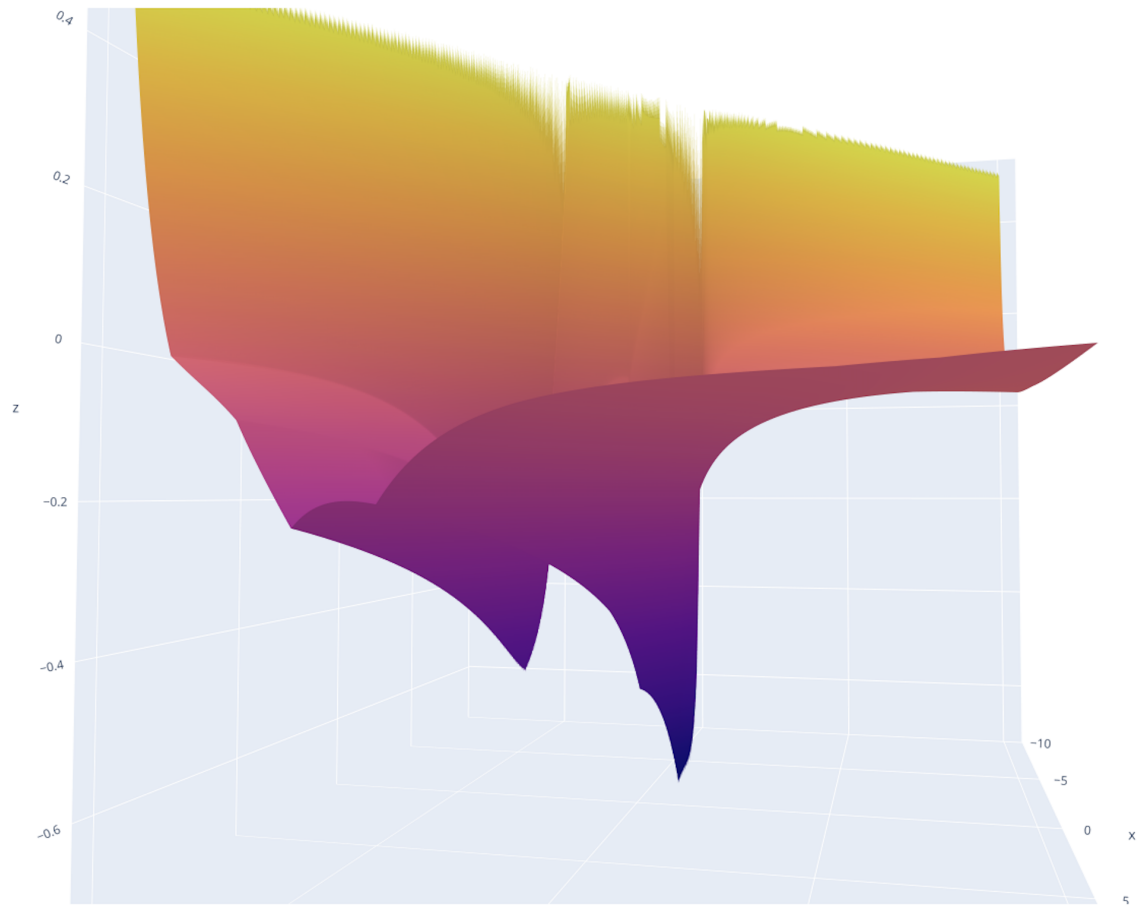
Offline part : Select accurate solutions for a few **wisely chosen** parameters $\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_K \in \mathcal{R}$

- Generate a training set of snapshots for parameters $\mathbf{R} \in \mathcal{R}_{\text{train}}$
- Select "good" snapshots with a greedy algorithm adapted to the miwture Wasserstein metric
  - Select two parameters $\mathbf{R}_1, \mathbf{R}_2 \in \mathcal{R}$ that are as far as possible
  - at each iteration $K \geq 3$, select the snapshot that is worse approximated **as a barycenter** in the **set** of the previously selected snapshots $\Psi_{\mathbf{R}_1}, \ldots, \Psi_{\mathbf{R}_{K-1}}$

Online part : compute solutions for **many** parameters $\mathbf{R} \in \mathcal{R}$ **as a barycenter of** selected snapshots, i.e. in the set of the solutions $\Psi_{\mathbf{R}_1}, \ldots, \Psi_{\mathbf{R}_K}$.

Online algorithm : energy minimization

- Nonlinear problem, but in low dimension
- Using quasi-Newton method starting from different initial guesses

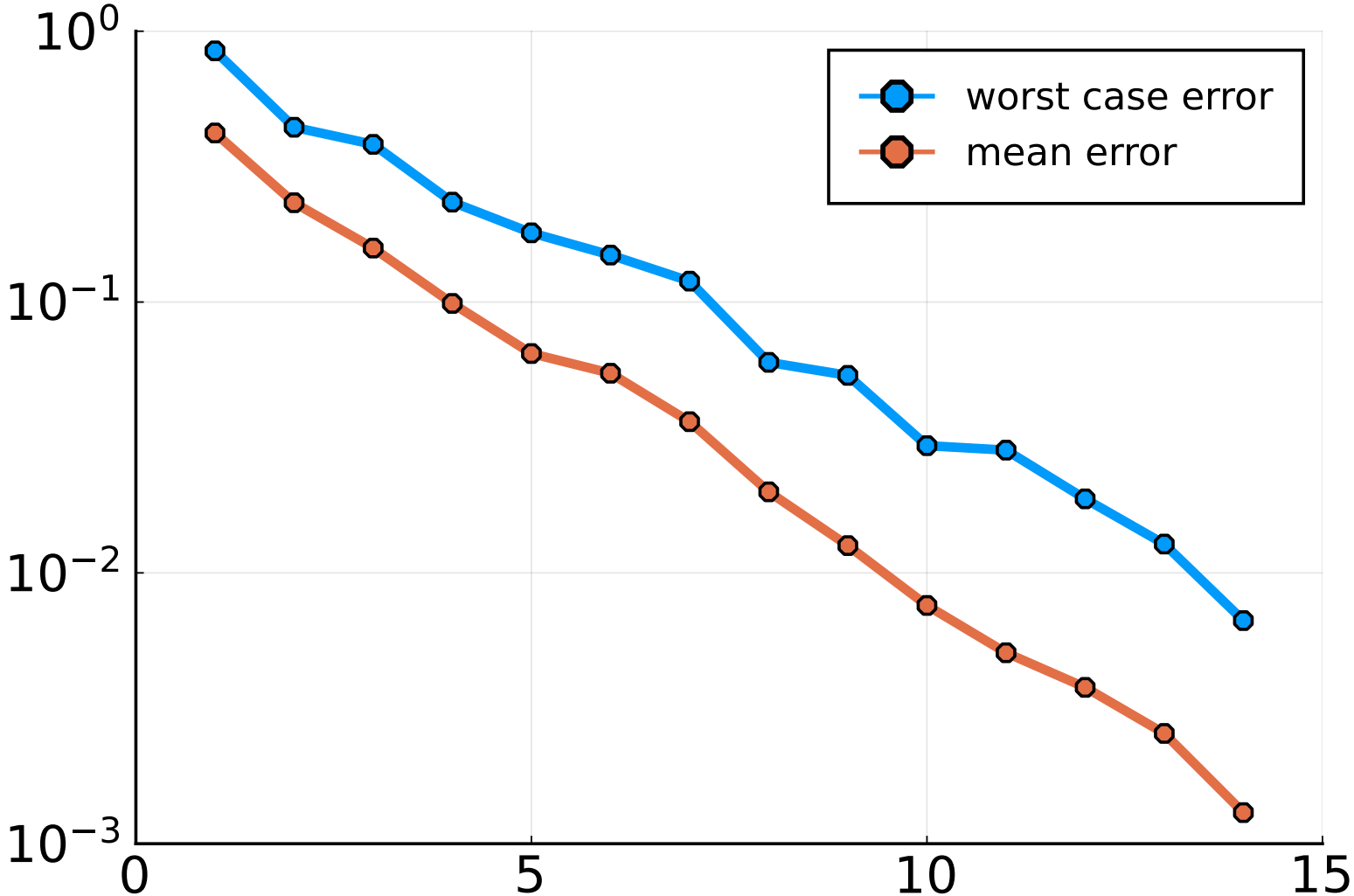


Numerical results : greedy algorithm

Charges : $(0.8, 1.1)$.

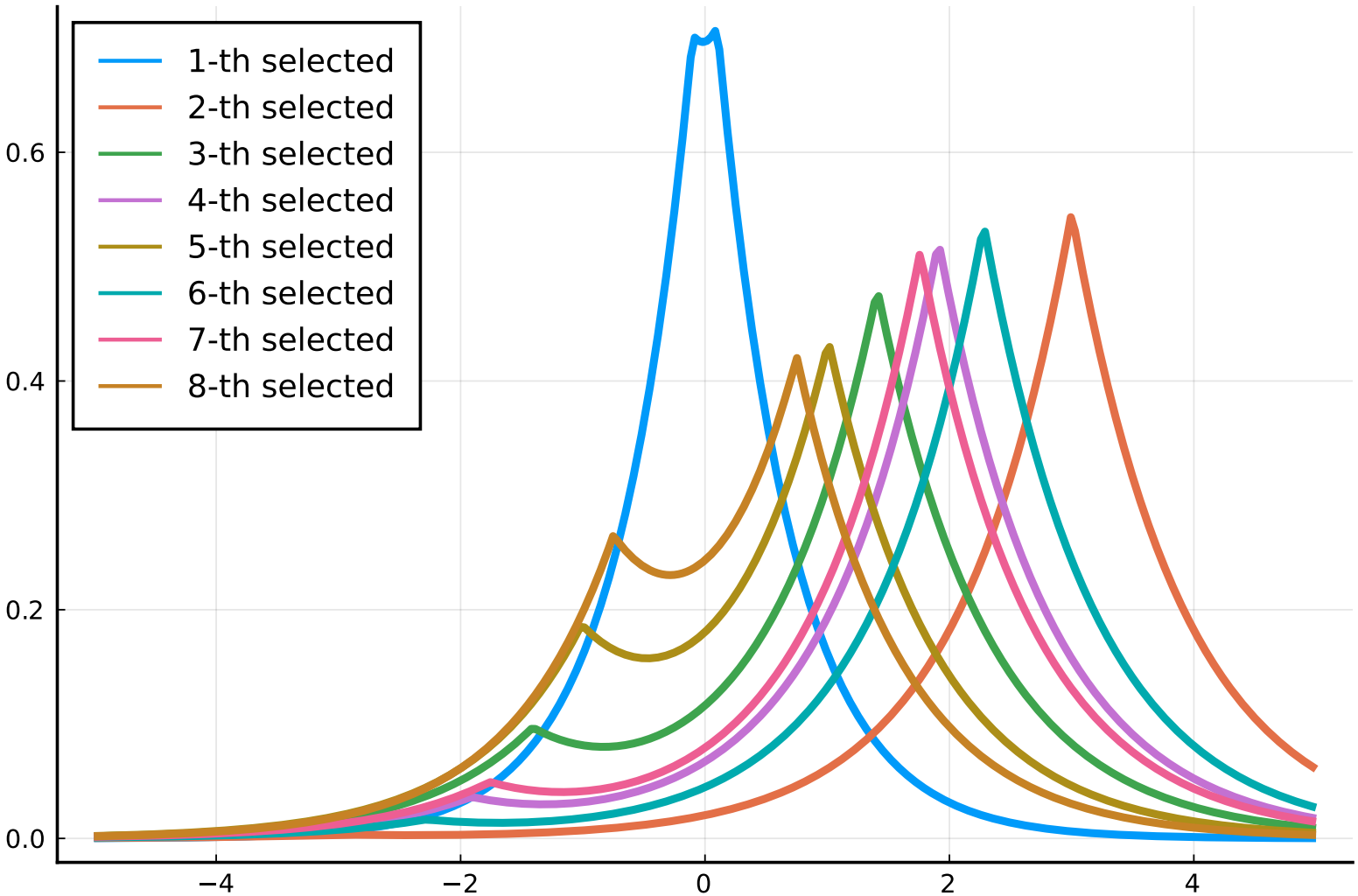
291 solutions in the training set.

Error decrease with respect to number of selected snapshots



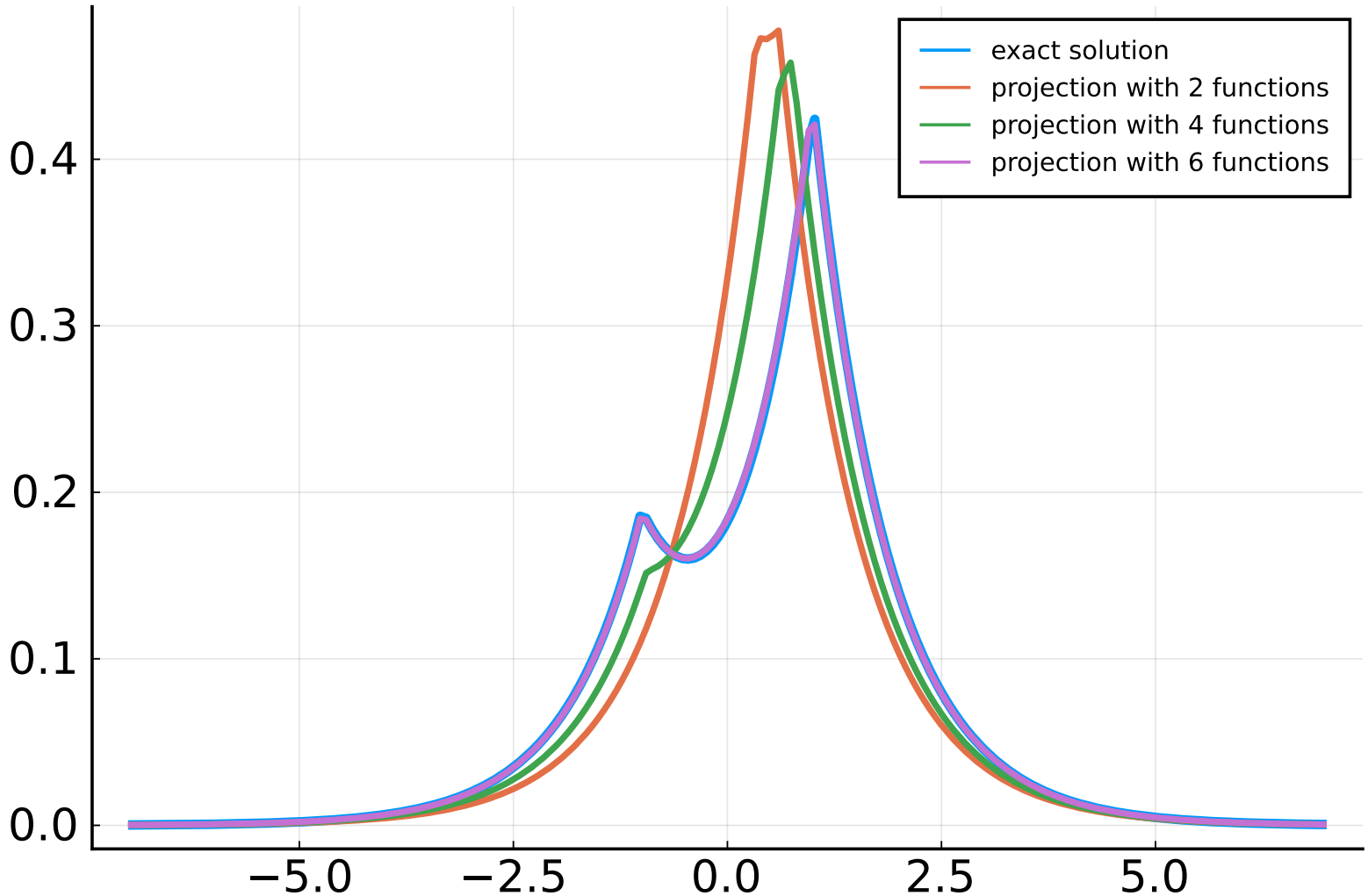
Numerical results : greedy algorithm

First eight selected solutions

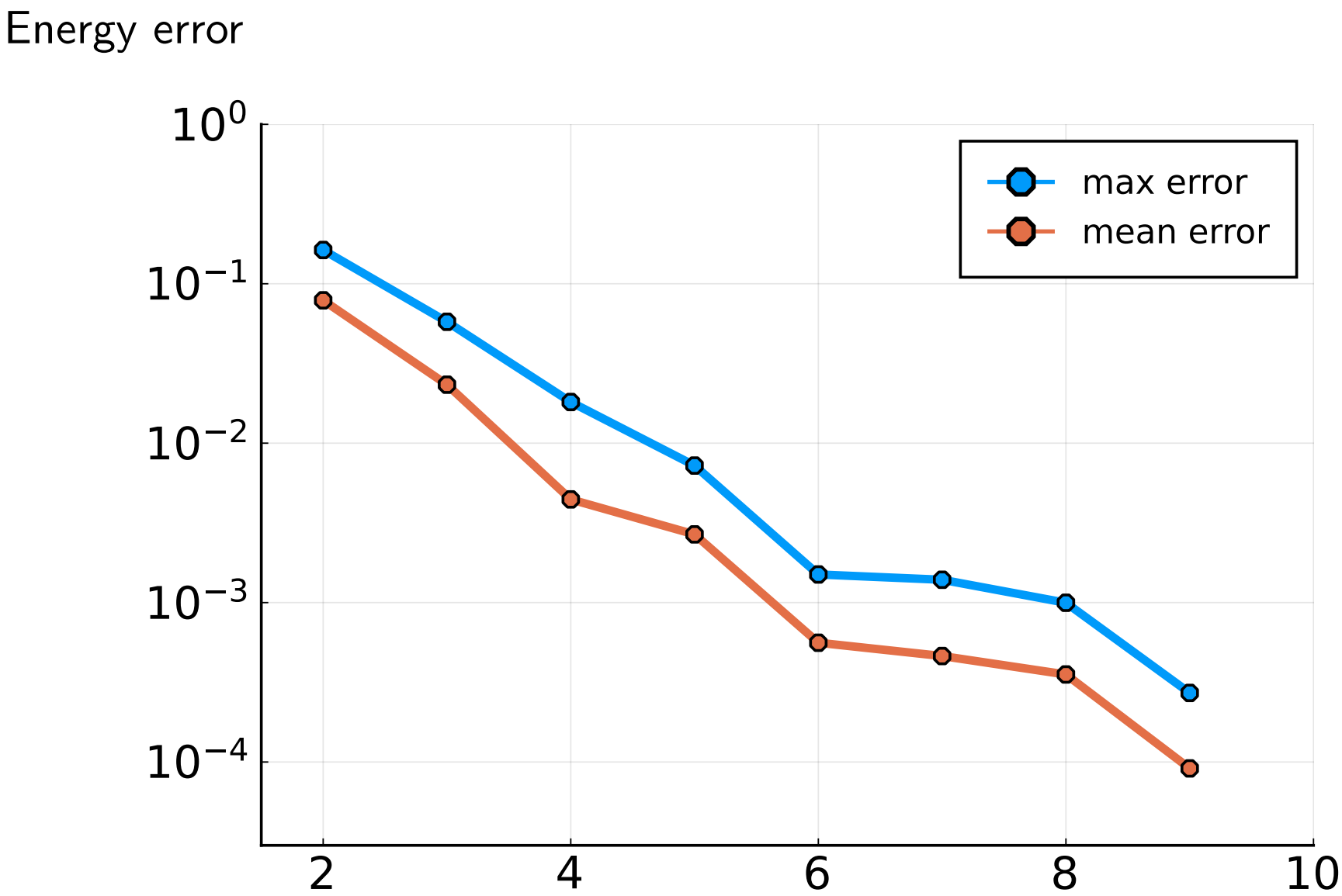


Numerical results : greedy algorithm

Projection example

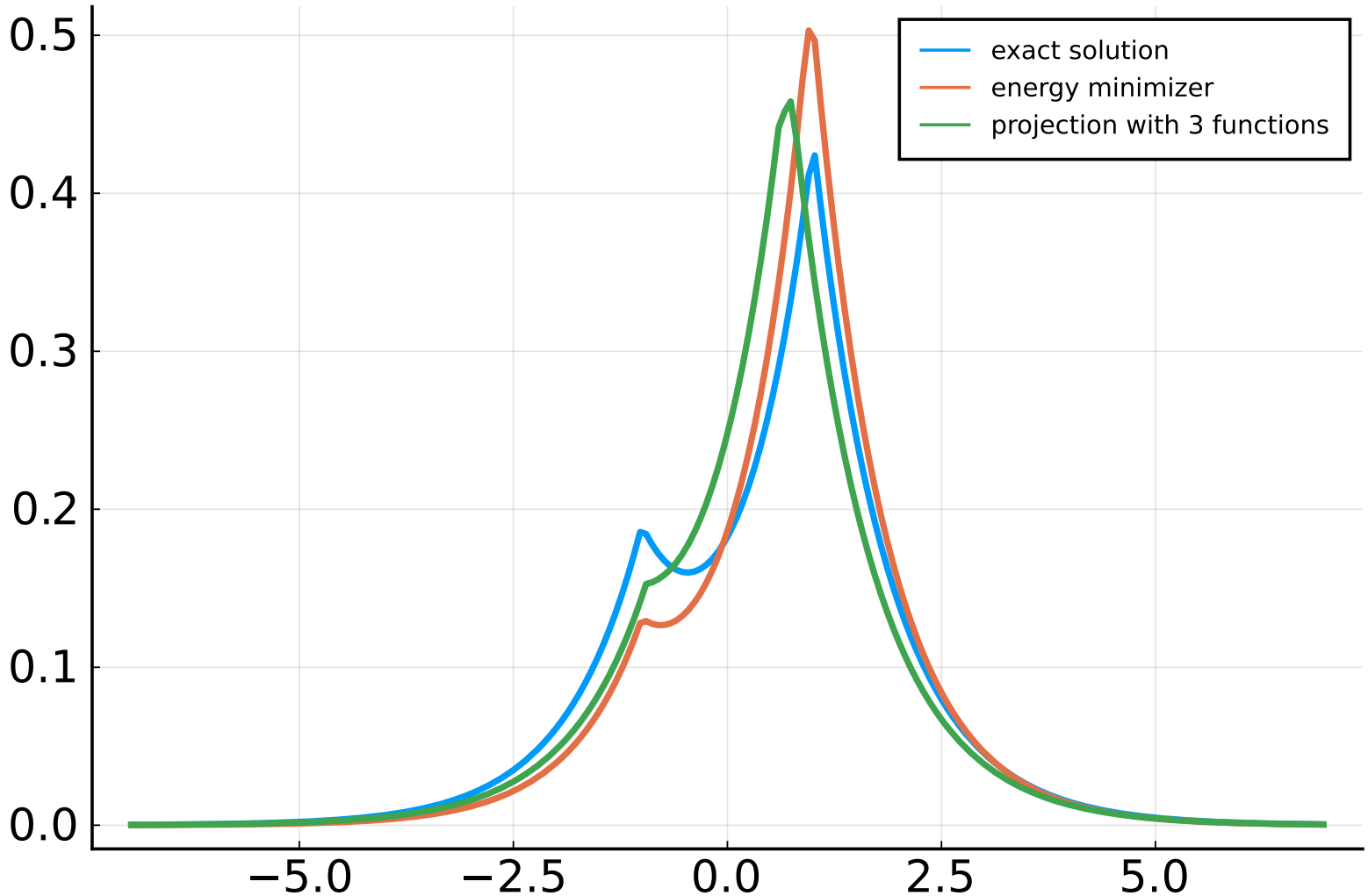


Numerical results : online energy minimization



Numerical results

Comparison between projection and energy minimization



Limitations and extensions

Key points :

- Independent of the underlying dimension
- Problem size depends on number of functions in the mixtures

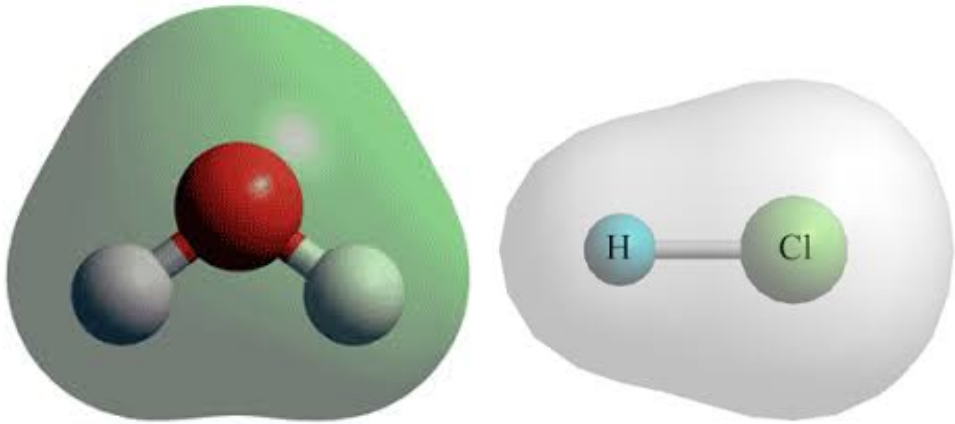
Limitations :

- Needs to consider probability distributions with specific structure

Extensions :

- consider orthogonal projectors problems using Quantum Optimal transport
- accelerate calculations via learning of the parameter map

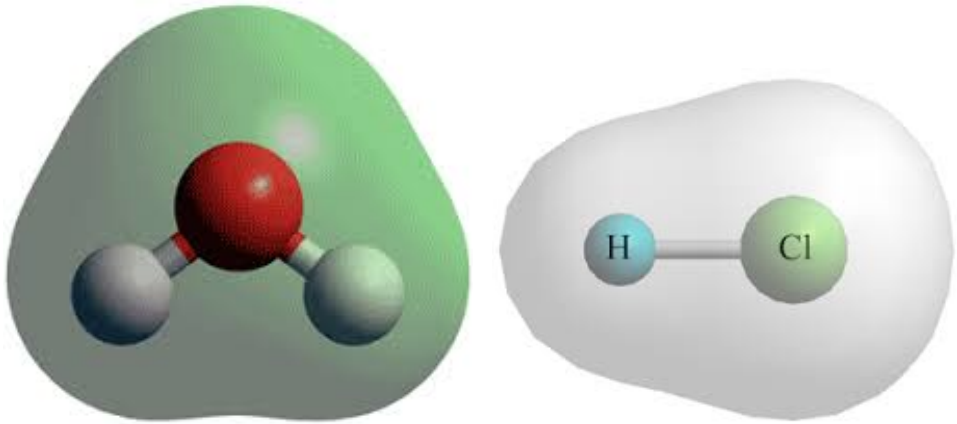
Towards calculations for molecules



Electronic (one or two-body) density :

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

Towards calculations for molecules

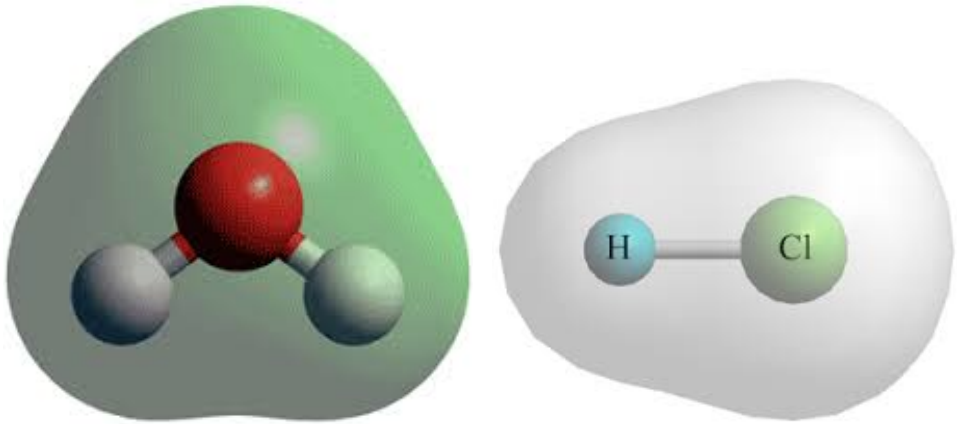


Electronic (one or two-body) density :

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

ELectronic structure models which yield approximations of the **one-body density** ρ are much faster (but less accurate) than models which yield approximations of the **two-body density** τ

Towards calculations for molecules



Electronic (one or two-body) density :

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

ELectronic structure models which yield approximations of the **one-body density** ρ are much faster (but less accurate) than models which yield approximations of the **two-body density** τ

Aim : accelerate two-body density calculations

Targeted workflow

Offline part :

- Select accurate solutions $\rho_{\mathbf{R}_1}, \ldots, \rho_{\mathbf{R}_k}$ and $\tau_{\mathbf{R}_1}, \ldots \tau_{\mathbf{R}_k}$ for a few **wisely chosen** parameters $\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_K \in \mathcal{R}$

Targeted workflow

Offline part :

- Select accurate solutions $\rho_{\mathbf{R}_1}, \ldots, \rho_{\mathbf{R}_k}$ and $\tau_{\mathbf{R}_1}, \ldots \tau_{\mathbf{R}_k}$ for a few **wisely chosen** parameters $\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_K \in \mathcal{R}$
- Approximate them as mixtures of Gaussians or Slater distributions $\overline{\rho}_{\mathbf{R}_1}, \ldots, \overline{\rho}_{\mathbf{R}_k}$ and $\overline{\tau}_{\mathbf{R}_1}, \ldots \overline{\tau}_{\mathbf{R}_k}$

  Bachmayr, Chen, Schneider : Error estimates for Hermite and even-tempered Gaussian approximations in quantum chemistry, Numerische Mathematik, (2014).

Targeted workflow

Online part :

- Compute only an approximation of the one-body density matrix $\rho_{\mathbf{R}}$ for $\mathbf{R} \in \mathcal{R}$

Targeted workflow

Online part :

- Compute only an approximation of the one-body density matrix $\rho_{\mathbf{R}}$ for $\mathbf{R} \in \mathcal{R}$
- Approximate it as mixtures of Gaussians or Slater distributions $\overline{\rho}_{\mathbf{R}}$

Targeted workflow

Online part :

- Compute only an approximation of the one-body density matrix $\rho_{\mathbf{R}}$ for $\mathbf{R} \in \mathcal{R}$
- Approximate it as mixtures of Gaussians or Slater distributions $\overline{\rho}_{\mathbf{R}}$
- Compute optimal barycentric weights $(t_1, \ldots, t_K) \in T_K$ such that

$$\overline{\rho}_{\mathbf{R}} \approx \mathrm{Bar}^{\boldsymbol{t}}_{\mathrm{MW}_2}\left(\overline{\rho}_{\mathbf{R}_1}, \ldots, \overline{\rho}_{\mathbf{R}_K}\right)$$

Targeted workflow

Online part :

- Compute only an approximation of the one-body density matrix $\rho_{\mathbf{R}}$ for $\mathbf{R} \in \mathcal{R}$
- Approximate it as mixtures of Gaussians or Slater distributions $\overline{\rho}_{\mathbf{R}}$
- Compute optimal barycentric weights $(t_1, \ldots, t_K) \in T_K$ such that

$$\overline{\rho}_{\mathbf{R}} \approx \mathrm{Bar}^{\boldsymbol{t}}_{\mathrm{MW}_2}\left(\overline{\rho}_{\mathbf{R}_1}, \ldots, \overline{\rho}_{\mathbf{R}_K}\right)$$

- Compute an approximation $\overline{\tau}_{\mathbf{R}}$ of the two-body density matrix $\tau_{\mathbf{R}}$ as

$$\overline{\tau}_{\mathbf{R}} = \mathrm{Bar}^{\boldsymbol{t}}_{\mathrm{MW}_2}\left(\overline{\tau}_{\mathbf{R}_1}, \ldots, \overline{\tau}_{\mathbf{R}_K}\right)$$

Conclusion

Conclusion :

- Nonlinear reduced model based for electronic structure calculations on mixture Wasserstein barycenters
- So far, theoretical and numerical results on 1D toy model
- Scales with the dimension

Perspectives :

- Extensions to density matrices (quantum optimal transport)
- 3D simulations and larger systems
- Other metrics than Wasserstein ?

Bon, Pai, Bellaard,Mula and Duits : Optimal Transport on the Lie group of Roto-Translations, 2024