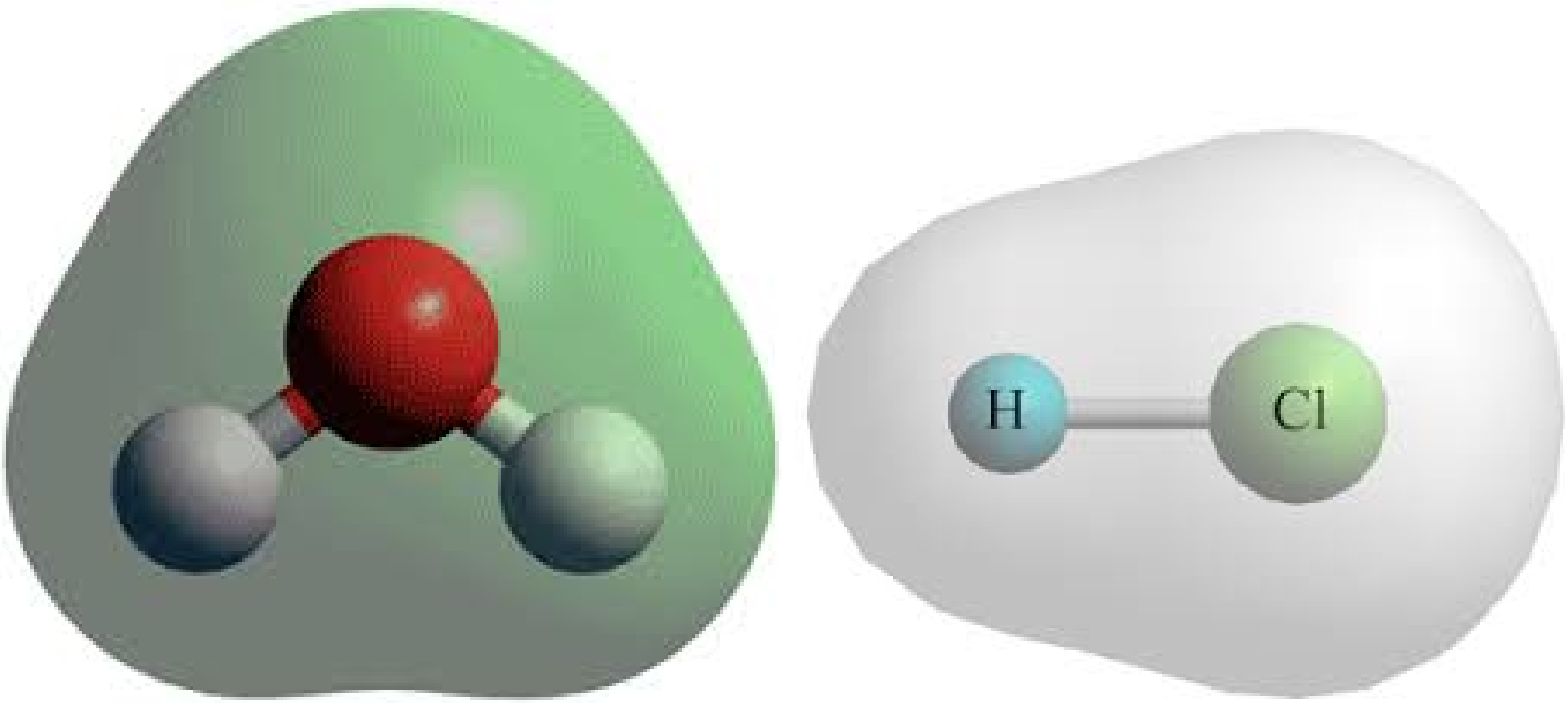
Neural network methods for Schrödinger eigenvalue problems

Mathias Dus, Geneviève Dusson, Virginie Ehrlacher, Clément Guillot, Pascal-Joël Soffo

Ecole nationale des Ponts et Chaussées, IPP & INRIA

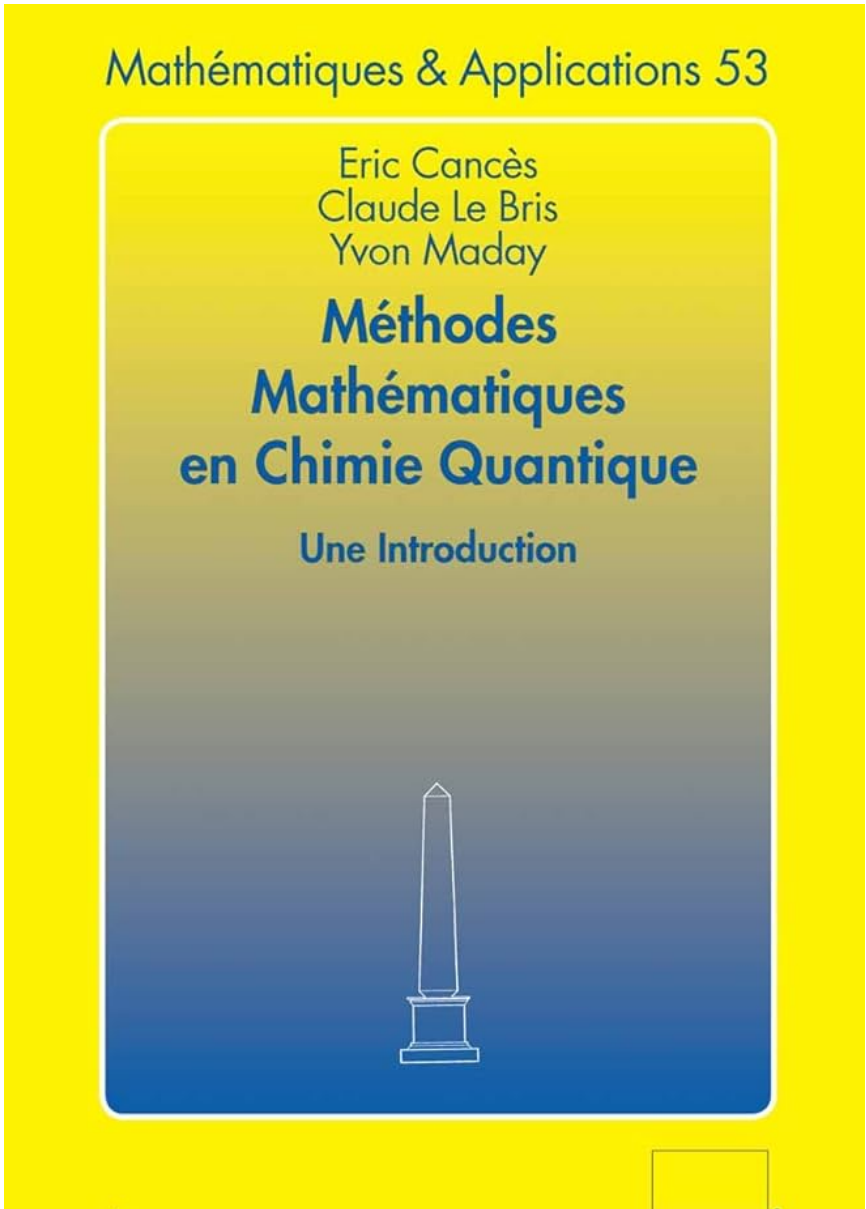
Yvon Maday's symposium, Collège de France, 22nd June 2026

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 molecular dynamics simulations...

Electronic structure calculations: a rich playground for reduced order models



EMC²

Extreme-scale Mathematically-based Computational Chemistry

Many-body electronic Schrödinger model

For the sake of simplicity, atomic units are used and the influence of spin is neglected.

Born-Oppenheimer approximation: Let us consider a physical system composed of:

Many-body electronic Schrödinger model

For the sake of simplicity, atomic units are used and the influence of spin is neglected.

Born-Oppenheimer approximation: Let us consider a physical system composed of:

- M nuclei, that are assumed to be (fixed) classical point charges, whose positions and electric charges are denoted respectively by $R_1, \ldots, R_M \in \mathbb{R}^3$ and $Z_1, \ldots, Z_M \in \mathbb{N}^*$ respectively;

Many-body electronic Schrödinger model

For the sake of simplicity, atomic units are used and the influence of spin is neglected.

Born-Oppenheimer approximation: Let us consider a physical system composed of:

- M nuclei, that are assumed to be (fixed) classical point charges, whose positions and electric charges are denoted respectively by $R_1, \ldots, R_M \in \mathbb{R}^3$ and $Z_1, \ldots, Z_M \in \mathbb{N}^*$ respectively;
- N electrons, considered as quantum particles, and represented by a complex-valued wavefunction $\psi(x_1, \ldots, x_N)$, where for all $1 \leq i \leq N$, $x_i \in \mathbb{R}^3$. The wavefunction $\psi \in \mathcal{A}$ with

$$\mathcal{A} := \left\{ \psi \in L^2(\mathbb{R}^{3N}; \mathbb{C}),\ \nabla_x \psi \in L^2(\mathbb{R}^{3N}; \mathbb{C}),\ \psi \text{ antisymmetric},\ \|\psi\|_{L^2} = 1 \right\}.$$

Many-body electronic Schrödinger model

For the sake of simplicity, atomic units are used and the influence of spin is neglected.

Born-Oppenheimer approximation: Let us consider a physical system composed of:

- M nuclei, that are assumed to be (fixed) classical point charges, whose positions and electric charges are denoted respectively by $R_1, \ldots, R_M \in \mathbb{R}^3$ and $Z_1, \ldots, Z_M \in \mathbb{N}^*$ respectively;
- N electrons, considered as quantum particles, and represented by a complex-valued wavefunction $\psi(x_1, \ldots, x_N)$, where for all $1 \leq i \leq N$, $x_i \in \mathbb{R}^3$. The wavefunction $\psi \in \mathcal{A}$ with

$$\mathcal{A} := \left\{ \psi \in L^2(\mathbb{R}^{3N}; \mathbb{C}),\ \nabla_x \psi \in L^2(\mathbb{R}^{3N}; \mathbb{C}),\ \psi \text{ antisymmetric},\ \|\psi\|_{L^2} = 1 \right\}.$$

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

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

Many-body electronic problem Schrödinger model

The ground state energy of the system of electrons E_g is given by the electronic problem:

$$E_g = \inf_{\psi \in \mathcal{A}} T[\psi] + V_{ee}[\psi] + V_{ne}[\psi],$$

where

- **kinetic energy:** $T[\psi] := \int_{\mathbb{R}^{3N}} \sum_{i=1}^{N} |\nabla_{x_i} \psi|^2$
- **electron-electron interaction energy:**

  $V_{ee}[\psi] := \int_{\mathbb{R}^{3N}} c(x_1, \cdots, x_N) |\psi(x_1, \cdots, x_N)|^2 \, dx_1 \cdots dx_N$ with

  $$c(x_1, \cdots, x_N) := \sum_{1 \le i < j \le N} \frac{1}{|x_i - x_j|}$$

- **nuclei-electron interaction energy:**

  $V_{ne}[\psi] := \int_{\mathbb{R}^{3N}} \left(\sum_{i=1}^{N} v(x_i) \right) |\psi(x_1, \cdots, x_N)|^2 \, dx_1 \cdots dx_N$, with

  $$v(x) := -\sum_{k=1}^{M} \frac{Z_k}{|R_k - x|}.$$

Eigenvalue problem

The corresponding ground state electronic wavefunction ψ_g is a solution to the eigenvalue problem (electronic Schrödinger problem)

$$H\psi_g = E_g\psi_g,$$

with H the self-adjoint operator on $L^2_{as}(\mathbb{R}^{3N};\mathbb{C})$ with domain $H^2_{as}(\mathbb{R}^{3N};\mathbb{C})$ given by

$$H = -\Delta_{x_1,\ldots,x_N} + \sum_{i=1}^{N} v(x_i) + \sum_{1\leq i<j\leq N} \frac{1}{|x_i - x_j|} = -\Delta_{x_1,\ldots,x_N} + W(x_1,\ldots,x_N)$$

and E_g is the lowest eigenvalue of H

Eigenvalue problem

The corresponding ground state electronic wavefunction ψ_g is a solution to the eigenvalue problem (electronic Schrödinger problem)

$$H\psi_g = E_g\psi_g,$$

with H the self-adjoint operator on $L^2_{as}(\mathbb{R}^{3N};\mathbb{C})$ with domain $H^2_{as}(\mathbb{R}^{3N};\mathbb{C})$ given by

$$H = -\Delta_{x_1,\ldots,x_N} + \sum_{i=1}^{N} v(x_i) + \sum_{1\leq i<j\leq N} \frac{1}{|x_i - x_j|} = -\Delta_{x_1,\ldots,x_N} + W(x_1,\ldots,x_N)$$

and E_g is the lowest eigenvalue of H

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

Eigenvalue problem

The corresponding ground state electronic wavefunction ψ_g is a solution to the eigenvalue problem (electronic Schrödinger problem)

$$H\psi_g = E_g\psi_g,$$

with H the self-adjoint operator on $L^2_{as}(\mathbb{R}^{3N};\mathbb{C})$ with domain $H^2_{as}(\mathbb{R}^{3N};\mathbb{C})$ given by

$$H = -\Delta_{x_1,\dots,x_N} + \sum_{i=1}^{N} v(x_i) + \sum_{1\le i<j\le N} \frac{1}{|x_i - x_j|} = -\Delta_{x_1,\dots,x_N} + W(x_1,\dots,x_N)$$

and E_g is the lowest eigenvalue of H

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

Classical methods (finite volumes, finite elements,...) fail because of the so-called **curse of dimensionality**:
The complexity of these methods scales **exponentially with** N

Why neural networks ?

- Some theory exists to reduce the dimension of the problem : Density Functional Theory, Hartree-Fock, ... but it includes **model approximations**.

Why neural networks ?

- Some theory exists to reduce the dimension of the problem : Density Functional Theory, Hartree-Fock, ... but it includes **model approximations**.
- Some works advocate the relevance of using **neural networks** (Physically Informed Neural Networks) for the electronic Schrödinger problem (FermiNet, PauliNet, DeepErwin...) in high dimension.

Why neural networks ?

- Some theory exists to reduce the dimension of the problem : Density Functional Theory, Hartree-Fock, ... but it includes **model approximations**.
- Some works advocate the relevance of using **neural networks** (Physically Informed Neural Networks) for the electronic Schrödinger problem (FermiNet, PauliNet, DeepErwin...) in high dimension.

Atom / Molecule	DMRG (Energy, Basis)	FermiNet (Energy)	PauliNet (Energy)	Exact
Li	-7.431553 (631g)	-7.473162	**-7.4762**	-7.47806032
C	-37.673536 (631g)	-37.8077	**-37.8176**	-37.8450
O	-74.516816 (sto-6g)	**-75.0551**	-75.008	-75.0673
H_2	-1.014310 (631g)	-1.1378366	-1.02192	×
LiH	-7.949315 (631g)	**-8.073885**	-8.00654	-8.070548
NH_3	-55.713815 (sto-6g)	-56.1439	-56.7146	×
C_2H_4	×	-78.52951	-76.0121	×

TABLE 1. Comparison of the energy computed for the different methods for several atoms and molecules

[Dus, Dusson, VE, Guillot, Soffo, 2025]

Mathematical analysis of a toy problem

Objective: Bring some elements of mathematical analysis on the following toy problem

- Consider the Schrödinger eigenvalue problem with Neumann boundary conditions on the bounded domain $\Omega = [0,1]^d$: find $(\lambda, u^*) \in \mathbb{R} \times H^1(\Omega)$ solution to

$$\begin{cases} -\Delta u^\star + W u^\star = \lambda u^\star \text{ on } \Omega \\ \partial_n u^\star = 0 \text{ on } \partial\Omega \end{cases} \tag{1}$$

  where $W \in L^\infty(\Omega)$ is a bounded interaction potential.

Remark: This toy problem does not cover the case of the true electronic many-body Schrödinger model, but arises in other types of applications for instance in the study of galactic dynamics (Henon-Heiles potential)...

Reformulation as a constrained minimization problem

Key tool: Reformulate the eigenvalue problem as a **constrained optimization problem**

- Consider the Schrödinger eigenvalue problem :

$$u^\star \in \underset{u \in H^1(\Omega);\quad \|u\|_{L^2(\Omega)}=1}{\operatorname{argmin}} \mathcal{E}(u)$$

with

$$\mathcal{E}(u) = \frac{1}{2}\left(\int_\Omega |\nabla u|^2 + W|u|^2\right)$$

Reformulation as a constrained minimization problem

Key tool: Reformulate the eigenvalue problem as a **constrained optimization problem**

- Consider the Schrödinger eigenvalue problem :

$$u^\star \in \underset{u \in H^1(\Omega);\quad \mathcal{C}(u)=0}{\operatorname{argmin}} \mathcal{E}(u)$$

with

$$\mathcal{C}(u) = \|u\|_{L^2(\Omega)} - 1.$$

Barron functional space

[Lu,Lu,Wang, 2021], [Després, 2022], [Barron, 1993]

We need an adapted functional space to study approximation properties of neural networks:

- Spectral definition of $\mathcal{B}^s(\Omega)$:

$$\mathcal{B}^s(\Omega) := \left\{ u \in L^1(\Omega) : \ \|u\|_{\mathcal{B}^s} < \infty \right\} \ \text{ with } \ \|u\|_{\mathcal{B}^s} := \sum_{k \in \mathbb{N}^d} (1 + \pi^s |k|_1^s) |\hat{u}(k)|$$

  with $\hat{u}(k)$ the k^{th} Fourier coefficient of u.

Activation function and parameter set

- Let $\sigma : \mathbb{R} \to \mathbb{R}$ be an **activation function**.

Activation function and parameter set

- Let $\sigma : \mathbb{R} \to \mathbb{R}$ be an **activation function**.
- Consider the **(unbounded) parameter set**

$$\Theta := \left\{ \theta := (a, w, b) : \ a, b \in \mathbb{R}, \ w \in \mathbb{R}^d, |w| = 1 \right\}$$
$$= \mathbb{R} \times S_{\mathbb{R}^d}(1) \times \mathbb{R}$$

Activation function and parameter set

- Let $\sigma : \mathbb{R} \to \mathbb{R}$ be an **activation function**.
- Consider the **(unbounded) parameter set**

$$\Theta := \left\{ \theta := (a, w, b) : \ a, b \in \mathbb{R}, \ w \in \mathbb{R}^d, |w| = 1 \right\}$$
$$= \mathbb{R} \times S_{\mathbb{R}^d}(1) \times \mathbb{R}$$

- and for all $R > 0$, the **bounded parameter set**

$$\Theta_R := \left\{ \theta := (a, w, b) : \ a, b \in \mathbb{R}, \ w \in \mathbb{R}^d, |w| = 1, \ |a| \leq R, \ |b| \leq R \right\}$$
$$= [-R, R] \times S_{\mathbb{R}^d}(1) \times [-R, R]$$

Representation of a Barron function by a two-layer neural network

Question: Given $u \in \mathcal{B}^s(\Omega)$ for some $s > 0$, at which rate does the lowest possible error

$$\inf_{u_m \in \mathcal{N}_{\sigma,m}} \|u - u_m\|_{H^1(\Omega)}$$

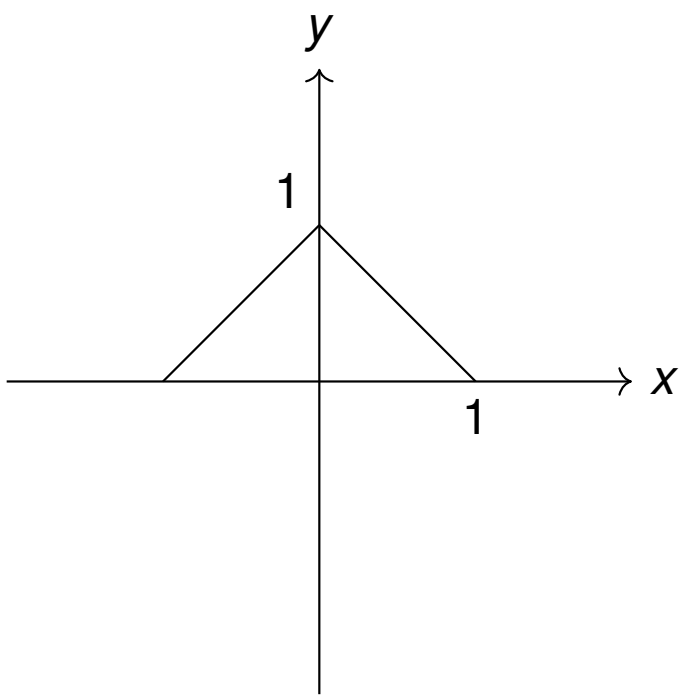
decay with m when $u_m \in \mathcal{N}_{\sigma,m}$ ranges over the set of functions that can be represented by a **two-layer neural network** with m neurons, i.e. when u_m can be written as

$$\forall x \in \Omega, \quad u_m(x) = \frac{1}{m} \sum_{i=1}^{m} a_i \sigma(w_i \cdot x + b_i)$$

with $\theta_i := (a_i, w_i, b_i) \in \Theta$ for all $1 \leq i \leq m$.

The hat activation function

$$\sigma_H(x) := \begin{cases} 0 & \text{si } |x| \geq 1 \\ x+1 & \text{si } -1 \leq x \leq 0 \\ 1-x & \text{si } 0 < x \leq 1. \end{cases}$$



Fundamental properties

[Lu, Lu, Wang, 2021], [Lu,Lu, 2021]

Theorem

- *Let* $u \in \mathcal{B}^2(\Omega)$*, then for all* $m \in \mathbb{N}$*, there exists* $(\theta_i)_{1 \leq i \leq m} \subset \Theta_R$ *with* $R = 4\|u\|_{\mathcal{B}^2(\Omega)}$ *such that :*

$$\|u - u_m\|_{H^1(\Omega)} \leq C \frac{\|u\|_{\mathcal{B}^2(\Omega)}}{\sqrt{m}}$$

with $\theta_i = (a_i, w_i, b_i)$ *and*

$$u_m(x) = \frac{1}{m} \sum_{i=1}^{m} a_i \sigma_H(w_i \cdot x + b_i) \in \mathcal{N}_{\sigma_H, m}$$

Fundamental properties

[Lu, Lu, Wang, 2021], [Lu,Lu, 2021]

Theorem

- *Let* $u \in \mathcal{B}^2(\Omega)$*, then for all* $m \in \mathbb{N}$*, there exists* $(\theta_i)_{1 \leq i \leq m} \subset \Theta_R$ *with* $R = 4\|u\|_{\mathcal{B}^2(\Omega)}$ *such that :*

$$\|u - u_m\|_{H^1(\Omega)} \leq C \frac{\|u\|_{\mathcal{B}^2(\Omega)}}{\sqrt{m}}$$

with $\theta_i = (a_i, w_i, b_i)$ *and*

$$u_m(x) = \frac{1}{m} \sum_{i=1}^{m} a_i \sigma_H(w_i \cdot x + b_i) \in \mathcal{N}_{\sigma_H, m}$$

- *If* $W \in L^\infty(\Omega) \cap \mathcal{B}^0(\Omega)$ *is such that the lowest eigenvalue* λ *of* $-\Delta + W$ *is simple, then any solution* u^* *to the Schrödinger eigenvalue problem belongs to* $\mathcal{B}^2(\Omega)$*.*

Physically Informed Neural Network approach

The PINN approach then consists in finding $(\theta_i^\star)_{1\leq i\leq m} \subset \Theta$ solution to

$$(\theta_i^\star)_{1\leq i\leq m} \in \underset{(\theta_i)_{1\leq i\leq m}\subset\Theta;\quad \mathcal{C}(u_m)=0}{\operatorname{argmin}} \mathcal{E}(u_m)$$

with $\theta_i = (a_i, w_i, b_i) \in \Theta$ and

$$u_m(x) = \frac{1}{m}\sum_{i=1}^{m} a_i \sigma_H(w_i \cdot x + b_i)$$

by means of (stochastic) **gradient descent algorithms** in the parameter space Θ^m.

Physically Informed Neural Network approach

The PINN approach then consists in finding $(\theta_i^\star)_{1\leq i\leq m} \subset \Theta$ solution to

$$(\theta_i^\star)_{1\leq i\leq m} \in \underset{(\theta_i)_{1\leq i\leq m}\subset\Theta;\quad \mathcal{C}(u_m)=0}{\operatorname{argmin}} \mathcal{E}(u_m)$$

with $\theta_i = (a_i, w_i, b_i) \in \Theta$ and

$$u_m(x) = \frac{1}{m}\sum_{i=1}^{m} a_i \sigma_H(w_i \cdot x + b_i)$$

by means of (stochastic) **gradient descent algorithms** in the parameter space Θ^m.

Highly non-convex problem with multiple local minima!

Measure representation

[Bach, Chizat, 2018], [Bach, Chizat, 2019]

Corollary

There exists $(\theta_i)_{1 \leq i \leq m} \subset \Theta$ *such that :*

$$\|u^\star - u_m\|_{H^1(\Omega)} \leq C \frac{\|u^\star\|_{\mathcal{B}(\Omega)}}{\sqrt{m}}$$

with

$$u_m(x) = \int_\Theta \Phi(\theta; x) d\mu_m(\theta)$$

with

$$\Phi(\theta; x) = \Phi((a, w, b); x) = a\sigma_H(w \cdot x + b)$$

and

$$\mu_m := \frac{1}{m} \sum_{i=1}^{m} \delta_{\theta_i} \in \mathcal{P}_2(\Theta)$$

where $\mathcal{P}_2(\Theta)$ *is the set of probability measures defined on the parameter set* Θ *with finite second-order moments.*

Fundamental idea and starting point

This suggests to study the following constrained minimization problem defined on the set of probability measures on the parameter set $\mathcal{P}_2(\Theta)$:

$$\mu^\star \in \underset{\mu \in \mathcal{P}_2(\Theta);\quad \mathcal{C}(u_\mu)=0}{\operatorname{argmin}} \mathcal{E}(u_\mu)$$

where

$$u_\mu(x) = \int_\Theta \Phi(\theta; x)\, d\mu(\theta),$$

Denoting by

$$E(\mu) = \mathcal{E}(u_\mu) \qquad C(\mu) = \mathcal{C}(u_\mu)$$

we then equivalently have

$$\mu^\star \in \underset{\mu \in \mathcal{P}_2(\Theta);\quad C(\mu)=0}{\operatorname{argmin}} E(\mu)$$

Gradient descent algorithms (for unconstrained problems)

- **in the parameter set:** For some small step size $\Delta t > 0$, at iteration $k \in \mathbb{N}^*$,

$$\theta_i^{k+1} \approx \theta_i^k - \Delta t \partial_{\theta_i} \mathcal{E}(u_m^k)$$

Gradient descent algorithms (for unconstrained problems)

- **in the parameter set:** For some small step size $\Delta t > 0$, at iteration $k \in \mathbb{N}^*$,

$$\theta_i^{k+1} \approx \theta_i^k - \Delta t \partial_{\theta_i} \mathcal{E}(u_m^k)$$

More precisely, consider the gradient descent algorithm (minimizing movement):

$$\left(\theta_i^{k+1}\right)_{1 \leq i \leq m} = \underset{(\theta_i)_{1 \leq i \leq m} \in \Theta^m}{\operatorname{argmin}} \frac{1}{2\Delta t} \sum_{i=1}^{m} d(\theta_i, \theta_i^k)^2 + \mathcal{E}(u_m)$$

where d is the geodesic distance on $\Theta = \mathbb{R} \times S_{\mathbb{R}^d}(1) \times \mathbb{R}$.

Gradient descent algorithms (for unconstrained problems)

- **in the parameter set:** For some small step size $\Delta t > 0$, at iteration $k \in \mathbb{N}^*$,

$$\theta_i^{k+1} \approx \theta_i^k - \Delta t \partial_{\theta_i} \mathcal{E}(u_m^k)$$

More precisely, consider the gradient descent algorithm (minimizing movement):

$$\left(\theta_i^{k+1}\right)_{1 \le i \le m} = \underset{(\theta_i)_{1 \le i \le m} \in \Theta^m}{\operatorname{argmin}} \frac{1}{2\Delta t} \sum_{i=1}^{m} d(\theta_i, \theta_i^k)^2 + \mathcal{E}(u_m)$$

where d is the geodesic distance on $\Theta = \mathbb{R} \times S_{\mathbb{R}^d}(1) \times \mathbb{R}$.

- **in the probability measure set:**

$$\mu^{k+1} = \underset{\mu \in \mathcal{P}_2(\Theta)}{\operatorname{argmin}} \frac{1}{2\Delta t} W_2(\mu, \mu^k)^2 + \mathcal{E}(u_\mu)$$

where W_2 is the so-called Wasserstein distance.

Definition of the Wasserstein distance

- The set of probability measures $\mathcal{P}_2(\Theta)$ is embedded with the 2-Wasserstein distance:

$$\forall \mu, \nu \in \mathcal{P}_2(\Theta), \quad W_2^2(\mu, \nu) = \inf_{\gamma \in \Gamma(\mu,\nu)} \int_{\Theta \times \Theta} d(\theta, \widetilde{\theta})^2 \, d\gamma(\theta, \widetilde{\theta})$$

where $\Gamma(\mu, \nu)$ is the set of probability measures $\gamma \in \mathcal{P}_2(\Theta \times \Theta)$ with marginals respectively $\mu \in \mathcal{P}_2(\Theta)$ and $\nu \in \mathcal{P}_2(\Theta)$;

Definition of the Wasserstein distance

- The set of probability measures $\mathcal{P}_2(\Theta)$ is embedded with the 2-Wasserstein distance:

$$\forall \mu, \nu \in \mathcal{P}_2(\Theta), \quad W_2^2(\mu, \nu) = \inf_{\gamma \in \Gamma(\mu,\nu)} \int_{\Theta \times \Theta} d(\theta, \widetilde{\theta})^2 \, d\gamma(\theta, \widetilde{\theta})$$

  where $\Gamma(\mu, \nu)$ is the set of probability measures $\gamma \in \mathcal{P}_2(\Theta \times \Theta)$ with marginals respectively $\mu \in \mathcal{P}_2(\Theta)$ and $\nu \in \mathcal{P}_2(\Theta)$;

- If $(\theta_i)_{1 \leq i \leq m}$ and $(\theta_i^k)_{1 \leq i \leq m}$ are elements of Θ^m that are close enough, it holds that

$$W_2(\mu, \mu^k)^2 = \sum_{i=1}^{m} d(\theta_i, \theta_i^k)^2,$$

  where

$$\mu = \frac{1}{m} \sum_{i=1}^{m} \delta_{\theta_i} \quad \text{and} \quad \mu^k == \frac{1}{m} \sum_{i=1}^{m} \delta_{\theta_i^k}$$

Continuous in time dynamics

In the parameter space, if we were considering an unconstrained optimization problem, letting Δt go to 0 would yield to the following continuous in time dynamics (coupled system of ODEs):

$$\dot{\theta}_i = -P_{T_{\theta_i}\Theta} \nabla_{\theta_i} \mathcal{E}(u_m) \qquad i = 1, \ldots, m$$

Continuous in time dynamics

In the parameter space, if we were considering an unconstrained optimization problem, letting Δt go to 0 would yield to the following continuous in time dynamics (coupled system of ODEs):

$$\dot{\theta}_i = -P_{T_{\theta_i}\Theta} \nabla_{\theta_i} \mathcal{E}(u_m) \qquad i = 1, \dots, m$$

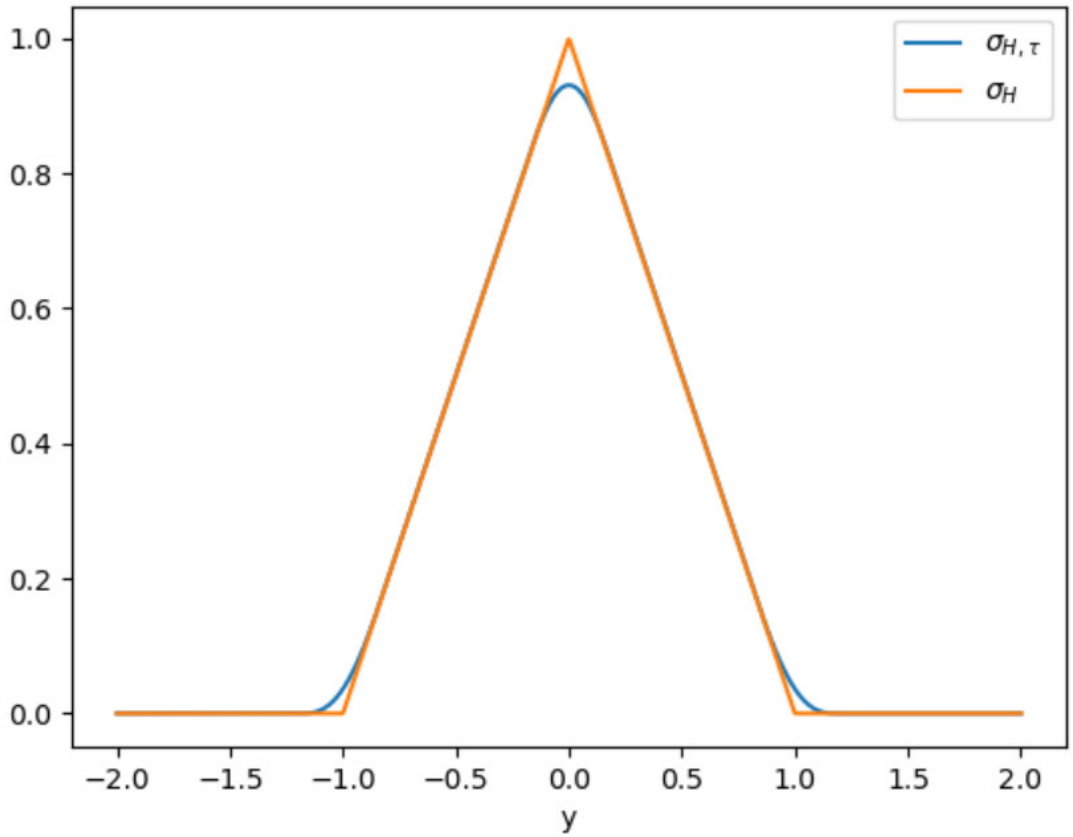
Questions:

$$\mu^{k+1} = \underset{\mu \in \mathcal{P}_2(\Theta);\ \mathcal{C}(u_\mu)=0}{\operatorname{argmin}} \frac{1}{2\Delta t} W_2(\mu, \mu^k)^2 + \mathcal{E}(u_\mu)$$

- Is there a continuous in time dynamics **in the space of measures** (taking into account the constraint)?
- Can we characterize it with some PDE?
- Does it converge in the long-time limit?
- If yes, what is its long-time limit?

Regularized hat activation function

- Need to consider a regularized version of the hat activation function $\sigma_{H,\tau}$ for $\tau > 0$.



Remark: Actually, the following theoretical results hold for any activation function σ that is C^∞ with compact support.

Existence of the constrained continuous in time dynamics

Theorem (Dus, VE, 2026)

Let $\mu_0 \in \mathcal{P}_2(\Theta)$ be compactly supported and such that $\|u_{\mu_0}\|_{L^2(\Omega)} = 1$. Then there exists a locally absolutely continuous curve $(\mu_t)_{t\geq 0} \subset \mathcal{P}_2(\Theta)$ such that $\|u_{\mu_t}\|_{L^2(\Omega)} = 1$ for all $t \geq 0$ and such that, for every $T > 0$,

$$\partial_t \mu_t + \operatorname{div}\big((-\nabla_\Theta V^C_{\mu_t})\mu_t\big) = 0 \qquad \text{in } \mathcal{D}'((0,T)\times\Theta), \tag{2}$$

$$\int_0^T \|\nabla_\Theta V^C_{\mu_t}\|^2_{L^2(\Theta;\mu_t)}\, dt < +\infty. \tag{3}$$

Moreover, there exists $R_T > 0$ such that

$$\operatorname{Supp}(\mu_t) \subset \Theta_{R_T} \qquad \text{for every } t \in [0,T]. \tag{4}$$

Moreover, if $\chi(t,s;\theta)$ denotes the flow associated with the vector field $-\nabla_\Theta V^C_{\mu_t}$, then

$$\mu_t = \chi(t,0;\cdot)_\# \mu_0 \qquad \text{for all } t \geq 0. \tag{5}$$

Main steps of the proof

- Existence of a continuous in time dynamics constrained to remain in a bounded parameter domain (Evolutionary Variational Inequality)
- Identification of the velocity field under the assumption that the support of μ_t remains strictly included in the bounded domain
- Extension of the dynamics to the whole parameter domain Θ using estimates on the growth of the size of the support of μ_t with t

Velocity field: For a regular function $F : \Theta \to \mathbb{R}$, we denote by

$$\nabla_\Theta F(\theta) := P_{T_\theta \Theta} \nabla F(\theta)$$

$$C_\mu(\theta) := \frac{\langle u_\mu, \Phi(\theta; \cdot) \rangle_{L^2(\Omega)}}{\|u_\mu\|_{L^2(\Omega)}}, \tag{6}$$

$$V_\mu(\theta) := \langle \nabla u_\mu, \nabla \Phi(\theta; \cdot) \rangle_{L^2(\Omega)} + \langle W u_\mu, \Phi(\theta; \cdot) \rangle_{L^2(\Omega)}, \tag{7}$$

$$\lambda_\mu := \frac{\langle \nabla_\Theta V_\mu, \nabla_\Theta C_\mu \rangle_{L^2(\Theta;\mu)}}{\|\nabla_\Theta C_\mu\|^2_{L^2(\Theta;\mu)}}, \tag{8}$$

$$V_\mu^C(\theta) := V_\mu(\theta) - \lambda_\mu C_\mu(\theta) = a g_\mu(w, b). \tag{9}$$

(10)

Main theorem of convergence

Hypothesis (1)

Assume that $\mu_0 \in \mathcal{P}_2(\Theta)$ *with compact support,* $\|u_\mu\|_{L^2(\Omega)} = 1$ *and*

$$\{0\} \times B \subset \mathrm{Supp}(\mu_0),$$

where

$$B := S_{\mathbb{R}^d}(1) \times [-\sqrt{d} - 2, \sqrt{d} + 2].$$

Theorem (Dus, VE, 2026)

Assume Hypothesis 1.

- *Let* $(\mu_t)_{t\geq 0}$ *be the solution given by Theorem 1, and assume that* $\mu_t \longrightarrow \mu^\star$ *in* $W_2(\Theta)$*as* $t \to +\infty$*. Then* $V^C_{\mu^\star} \equiv 0$ *on* Θ*.*
- *If* $\mu \in \mathcal{P}_2(\Theta)$ *is such that* $\|u_\mu\|_{L^2(\Omega)} = 1$ *and* $V^C_\mu = 0$ *everywhere, then* u_μ *is an eigenfunction of the Schrödinger-Neumann problem with* λ_μ *as eigenvalue.*

Main theorem of convergence

Hypothesis (1)

Assume that $\mu_0 \in \mathcal{P}_2(\Theta)$ *with compact support,* $\|u_\mu\|_{L^2(\Omega)} = 1$ *and*

$$\{0\} \times B \subset \mathrm{Supp}(\mu_0),$$

where

$$B := S_{\mathbb{R}^d}(1) \times [-\sqrt{d} - 2, \sqrt{d} + 2].$$

Theorem (Dus, VE, 2026)

Assume Hypothesis 1.

- *Let* $(\mu_t)_{t\geq 0}$ *be the solution given by Theorem 1, and assume that* $\mu_t \longrightarrow \mu^\star$ *in* $W_2(\Theta)$*as* $t \to +\infty$*. Then* $V^C_{\mu^\star} \equiv 0$ *on* Θ*.*
- *If* $\mu \in \mathcal{P}_2(\Theta)$ *is such that* $\|u_\mu\|_{L^2(\Omega)} = 1$ *and* $V^C_\mu = 0$ *everywhere, then* u_μ *is an eigenfunction of the Schrödinger-Neumann problem with* λ_μ *as eigenvalue.*

Take-away message: For infinite-width two-layer NNs, if the gradient descent algorithm converges, it converges towards a solution of the Schrödinger eigenvalue problem.

Asymptotic sign separation on the transported slice

Theorem (Dus, VE, 2026)

Assume Hypothesis 1. Let $(\mu_t)_{t\geq 0}$ be the solution given by Theorem 1, and assume that

$$\mu_t \longrightarrow \mu^\star \qquad \text{in } W_2(\Theta) \quad \text{as } t \to +\infty.$$

Write

$$V^C_{\mu^\star}(a,\omega) = a\, g_{\mu^\star}(\omega), \qquad \omega = (w,b) \in B.$$

Let $\eta_- > 0$ and $\eta_+ > 0$ be such that $-\eta_-$ and η_+ are regular values of $g_{\mu^\star}|_{\mathrm{int}(B)}$, and define

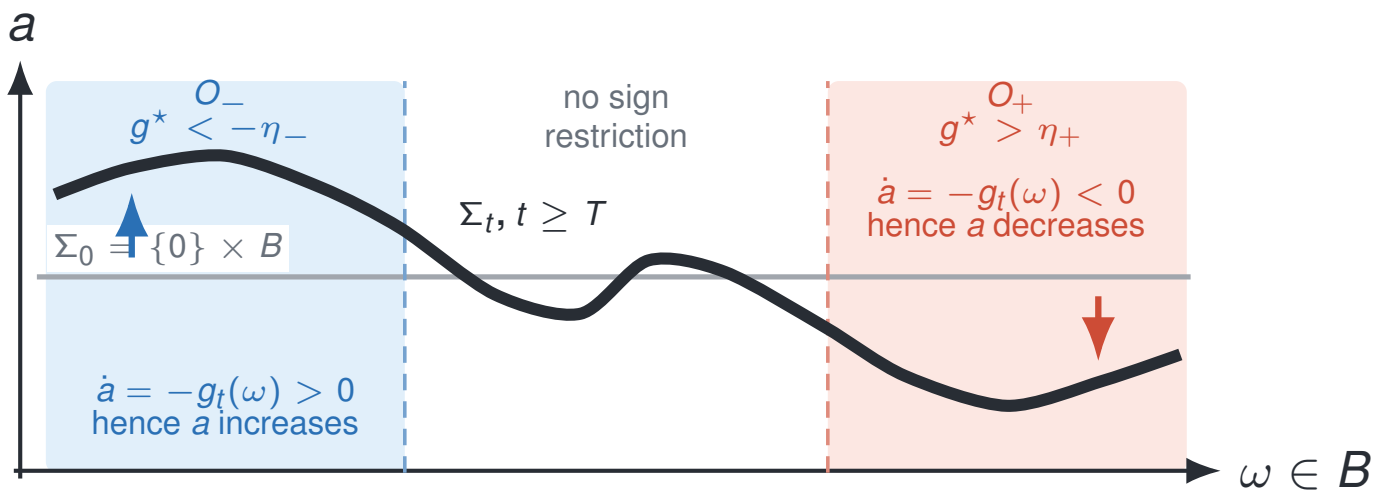
$$O_- := \{\omega \in B\,;\; g_{\mu^\star}(\omega) < -\eta_-\}, \qquad O_+ := \{\omega \in B\,;\; g_{\mu^\star}(\omega) > \eta_+\}.$$

Then there exists $T \geq 0$ such that, for all $t \geq T$,

$$\Sigma_t \cap (\mathbb{R}\times O_-) \subset [0,+\infty)\times O_-, \qquad \text{and} \qquad \Sigma_t \cap (\mathbb{R}\times O_+) \subset (-\infty,0]\times O_+,$$

where $\Sigma_t := \chi(t, \{0\}\times B)$.

Asymptotic sign separation on the transported slice



LaSalle principle

Theorem (Dus, VE, 2026)

Assume that g_μ takes a negative value, and let $-\eta < 0$ be a regular value in the range of $g_\mu|_{\operatorname{int} B}$. For $\alpha > 0$, define

$$O_- := \{\omega \in B : g_\mu(\omega) < -\eta\}, \qquad A^-_{\alpha,\eta} := (\alpha, +\infty) \times O_-.$$

Then there exists $\varepsilon_{\mu,\alpha,\eta} > 0$ with the following property. For every $t_0 \geq 0$ and every locally W_2-absolutely continuous curve $(\nu_t)_{t \geq t_0} \subset X_\tau$ solving

$$\partial_t \nu_t + \operatorname{div}(b_t \nu_t) = 0, \qquad b_t(\theta) := -\nabla_\Theta V^C_{\nu_t}(\theta),$$

if

$$W_2(\nu_{t_0}, \mu) \leq \varepsilon_{\mu,\alpha,\eta} \quad \text{and} \quad \nu_{t_0}(A^-_{\alpha,\eta}) > 0,$$

then there exists $t_1 > t_0$ such that

$$W_2(\nu_{t_1}, \mu) > \varepsilon_{\mu,\alpha,\eta}.$$

Support condition

Theorem (Dus, VE, 2026)

Assume Hypothesis 1. Let $(\mu_t)_{t\geq 0}$ be the solution given by Theorem 1, and assume that

$$\mu_t \to \mu^\star \qquad \text{in } W_2(\Theta) \quad \text{as } t \to +\infty.$$

Write

$$V^C_{\mu^\star}(a,\omega) = a\, g_{\mu^\star}(\omega).$$

Assume that $g_{\mu^\star} < 0$ somewhere, let $\eta > 0$ be such that $-\eta$ is a regular value of $g_{\mu^\star}$, and define

$$O_- := \{\omega \in B\ ;\ g_{\mu^\star}(\omega) < -\eta\}.$$

Then, for every $\alpha > 0$, there exists $T_\alpha \geq 0$ such that

$$\mu_t\big((\alpha,+\infty)\times O_-\big) = \mu_t\big(A^-_{\alpha,\eta}\big) > 0 \qquad \forall t \geq T_\alpha.$$

The proof of this result stems from homotopy properties of the unit sphere $S_{\mathbb{R}^d}(1)$.

Conclusions

Summary:

- Elements of mathematical analysis on the convergence of gradient descent for infinite-width two-layer neural networks for Schrödinger eigenvalue problems with bounded potentials.
- For inifinite-width two-layer NNs, under appropriate assumptions on the support of the initial measure, if the gradient descent algorithm converges, it converges towards a solution of the Schrödinger eigenvalue problem.

Conclusions

Summary:

- Elements of mathematical analysis on the convergence of gradient descent for infinite-width two-layer neural networks for Schrödinger eigenvalue problems with bounded potentials.
- For inifinite-width two-layer NNs, under appropriate assumptions on the support of the initial measure, if the gradient descent algorithm converges, it converges towards a solution of the Schrödinger eigenvalue problem.

Perspectives and open questions:

- Does the dynamics converge in the long-time limit?
- If it converges, does it converge to a solution of the eigenvalue problem associated with the lowest eigenvalue problem?

Conclusions

Summary:

- Elements of mathematical analysis on the convergence of gradient descent for infinite-width two-layer neural networks for Schrödinger eigenvalue problems with bounded potentials.
- For inifinite-width two-layer NNs, under appropriate assumptions on the support of the initial measure, if the gradient descent algorithm converges, it converges towards a solution of the Schrödinger eigenvalue problem.

Perspectives and open questions:

- Does the dynamics converge in the long-time limit?
- If it converges, does it converge to a solution of the eigenvalue problem associated with the lowest eigenvalue problem?
- Propagation of chaos results: can we quantify the speed of convergence of dynamics of finite-width two-layer NNs to the infinite width limit with respect to the number of parameters m?

Conclusions

Summary:

- Elements of mathematical analysis on the convergence of gradient descent for infinite-width two-layer neural networks for Schrödinger eigenvalue problems with bounded potentials.
- For inifinite-width two-layer NNs, under appropriate assumptions on the support of the initial measure, if the gradient descent algorithm converges, it converges towards a solution of the Schrödinger eigenvalue problem.

Perspectives and open questions:

- Does the dynamics converge in the long-time limit?
- If it converges, does it converge to a solution of the eigenvalue problem associated with the lowest eigenvalue problem?
- Propagation of chaos results: can we quantify the speed of convergence of dynamics of finite-width two-layer NNs to the infinite width limit with respect to the number of parameters m?
- Towards more complex architectures ? [Barboni, Peyré, Vialard, 2025], [Nenna, Pass, 2024]

Back to electronic structure calculations

What about potentials with Coulombic singularities ?

Back to electronic structure calculations

What about potentials with Coulombic singularities ?

- The ground state wavefunction $\psi_g \in \mathcal{B}^s$ for all $s < 1$ and that is optimal. [Yserentant, 2025]

Back to electronic structure calculations

What about potentials with Coulombic singularities ?

- The ground state wavefunction $\psi_g \in \mathcal{B}^s$ for all $s < 1$ and that is optimal. [Yserentant, 2025]
- Actually, one can write

$$\psi_g = J\phi_g,$$

where J is an explicit quantity named the Jastrow factor and ϕ_g is more regular that ψ_g.

Back to electronic structure calculations

What about potentials with Coulombic singularities ?

- The ground state wavefunction $\psi_g \in \mathcal{B}^s$ for all $s < 1$ and that is optimal. [Yserentant, 2025]
- Actually, one can write

$$\psi_g = J\phi_g,$$

where J is an explicit quantity named the Jastrow factor and ϕ_g is more regular that ψ_g.

- For the Hydrogen atom, $\phi_g \in \mathcal{B}^s$ for all $s < 2$.

Back to electronic structure calculations

What about potentials with Coulombic singularities ?

- The ground state wavefunction $\psi_g \in \mathcal{B}^s$ for all $s < 1$ and that is optimal. [Yserentant, 2025]
- Actually, one can write

$$\psi_g = J\phi_g,$$

  where J is an explicit quantity named the Jastrow factor and ϕ_g is more regular that ψ_g.

- For the Hydrogen atom, $\phi_g \in \mathcal{B}^s$ for all $s < 2$.
- **Open question/Conjecture:** Does $\phi_g \in \mathcal{B}^s$ for all $s < 2$ for a general molecule?

Back to electronic structure calculations

What about potentials with Coulombic singularities ?

- The ground state wavefunction $\psi_g \in \mathcal{B}^s$ for all $s < 1$ and that is optimal. [Yserentant, 2025]
- Actually, one can write

$$\psi_g = J\phi_g,$$

  where J is an explicit quantity named the Jastrow factor and ϕ_g is more regular that ψ_g.

- For the Hydrogen atom, $\phi_g \in \mathcal{B}^s$ for all $s < 2$.
- **Open question/Conjecture:** Does $\phi_g \in \mathcal{B}^s$ for all $s < 2$ for a general molecule?

Thank you for your attention!