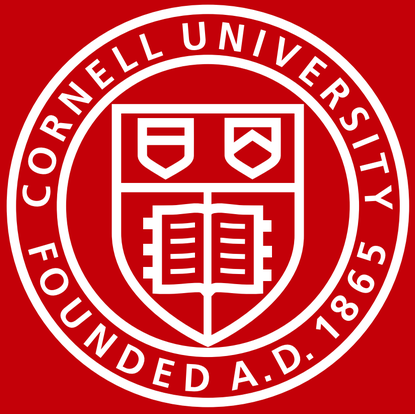
Localization methods

perspectives on optimization and initialization

Anil Damle
IPAM, April 2022



Acknowledgements

Objective functions and virtual orbitals:

Kangbo Li (Cornell)

Robert DiStasio Jr. (Cornell)

Hsin-Yu Ko (Cornell)

SCDM + variational formulations:

Lin Lin (UC Berkeley)

Lexing Ying (Stanford)

Antoine Levitt (Inria Paris)

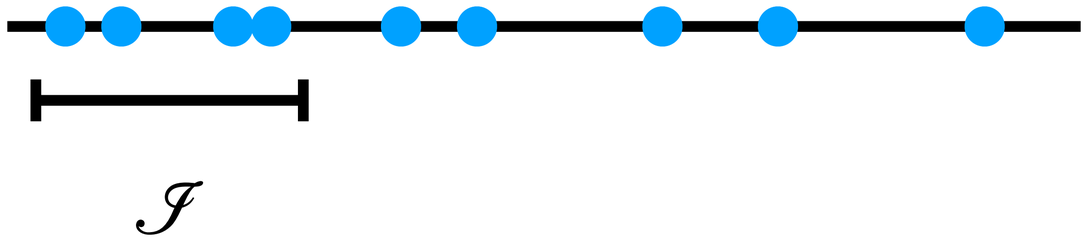
Eric G Fuemmeler (Mayo Clinic)

Funding: NSF, Simons Foundation, and DOE

The set up

Eigenfunctions $\{\psi_j\}$ of a self-adjoint Hamiltonian $\mathscr{H}$ with eigenvalues in $\mathscr{I}$

$$\mathscr{H}[\rho]\psi_j(\mathbf{r}) = \lambda_j\psi_j(\mathbf{r}) \text{ with } j \in \mathscr{I}$$



Denote the set of eigenfunctions: $\{\psi_j\}_{j=1}^{N}$

The localization problem, mathematically

Compute *localized* orbitals (Wannier functions) ϕ_j such that:

Each ϕ_j is “significant” on a small part of the domain

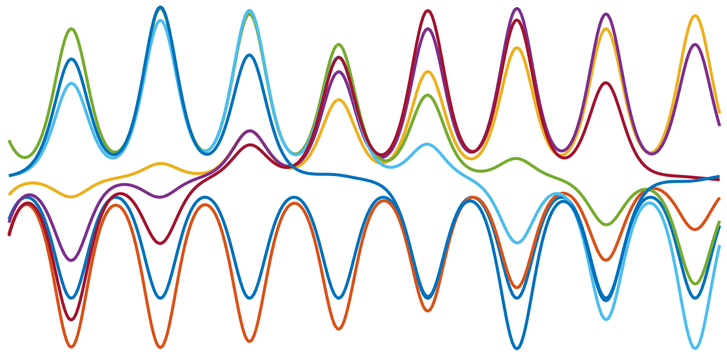
Each ϕ_j decays rapidly away from its “center”

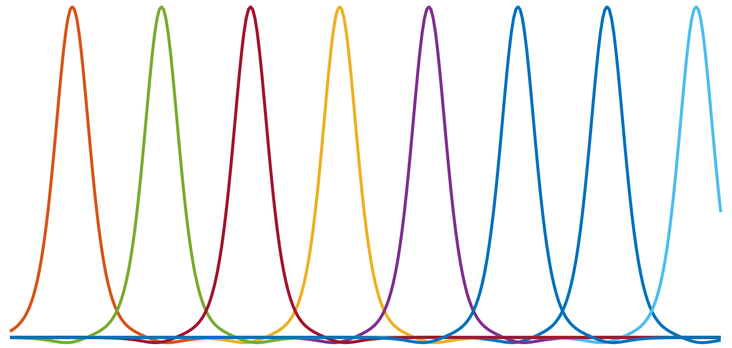
We have N_w localized orbitals with $N_w \leq N$

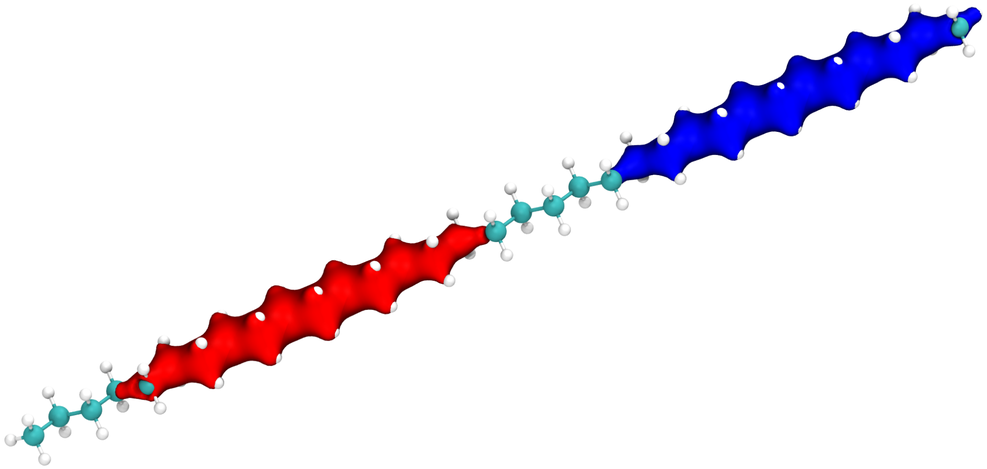
$$\text{span } \{\phi_j\}_{j=1}^{N_w} \subseteq \text{span } \{\psi_j\}_{j=1}^{N}$$

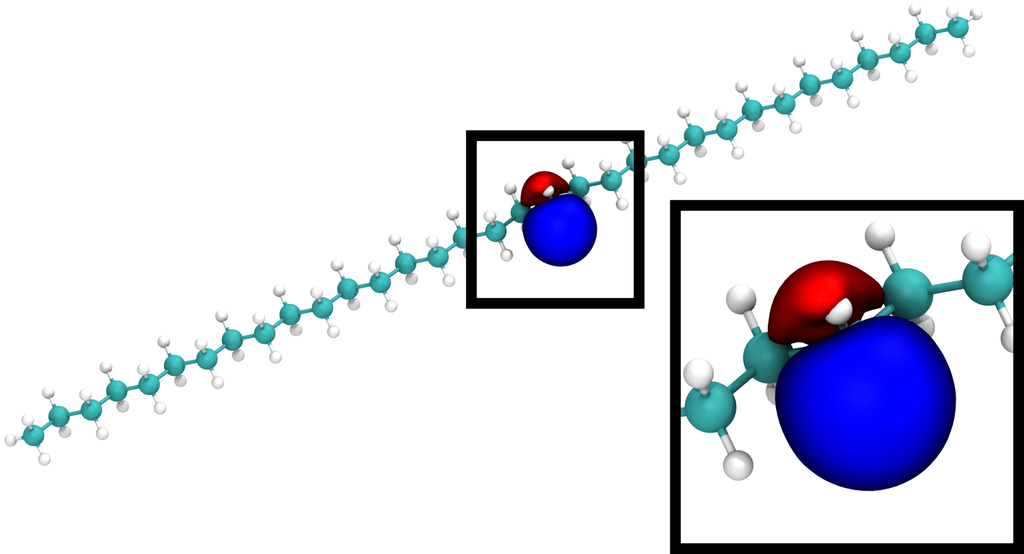
NB: Not immediately clear when/if this is possible—for the moment lets assume it is

Localization, in pictures









Localization, via linear algebra

Upon “discretization” we have: $\Psi = \begin{bmatrix} | & & | \\ \psi_1 & \cdots & \psi_N \\ | & & | \end{bmatrix}$ and $\Phi = \begin{bmatrix} | & & | \\ \phi_1 & \cdots & \phi_{N_w} \\ | & & | \end{bmatrix}$

Find $Q \in \mathbb{O}_{N \times N_w}$ (so $Q^* Q = I$) such that:

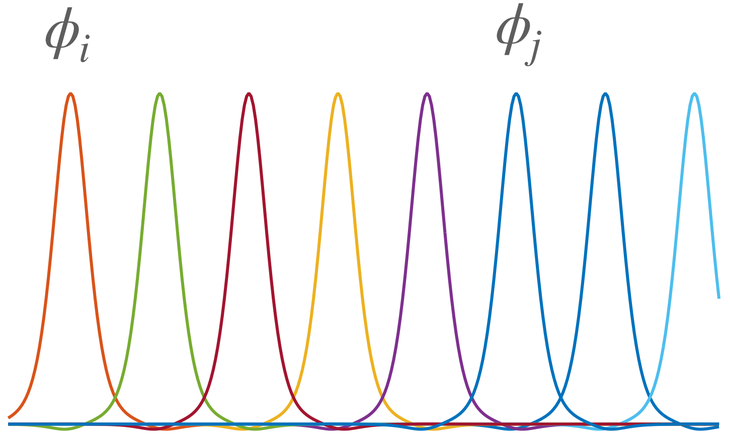
$$\Phi = \Psi Q$$

has localized columns

Why do we want localized basis functions?

Localization methods play a key role in efficient computational methods:

- Induce “sparsity”
- So-called linear-scaling methods
- Exact exchange
- Local electron correlation
- Interpolation (see previous talks)
- and many more…



$$\int \phi_i \phi_j d\mathbf{r} \approx 0$$

See the review on MLWFs by Marzari, et. al. [2012]

Localization, computationally

A prototype scheme (one option)

1. Mathematically define a measure of locality $\mathscr{L}(\phi_i)$
2. “Solve” the optimization problem:

$$\min_{Q^*Q=I} f(\mathscr{L}(\phi_1), \ldots, \mathscr{L}(\phi_{N_w}))$$

Generally not a convex problem, typically exhibits many local minima

For concreteness, can consider

$$f(\mathscr{L}(\phi_1), \ldots, \mathscr{L}(\phi_{N_w})) = \sum \mathscr{L}(\phi_i)$$

Other formulations (simultaneous diagonalization by Gygi et al. [2003], iterated projection by Stubbs et al. [2021],…)

Observations and choices

Choice of f and $\mathscr{L}$ impact how effectively we can solve the problem

Often need a good initialization to find a well localized basis

Choice of f and $\mathscr{L}$ influence the character of the computed orbitals

Need to appeal to manifold optimization

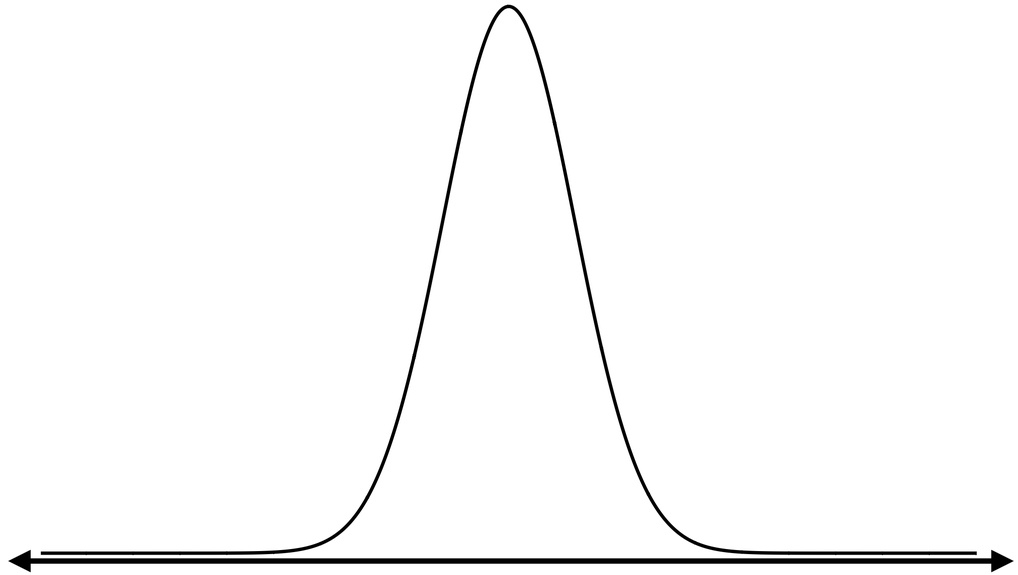
Locality: density convolution

Locality, in an infinite domain

For an infinite system, quite natural to use “variance”:

$$\langle \psi | \, r^2 \, | \psi \rangle - \langle \psi | \, r \, | \psi \rangle^2$$

For molecular systems is essentially the Foster-Boys [1960] criteria

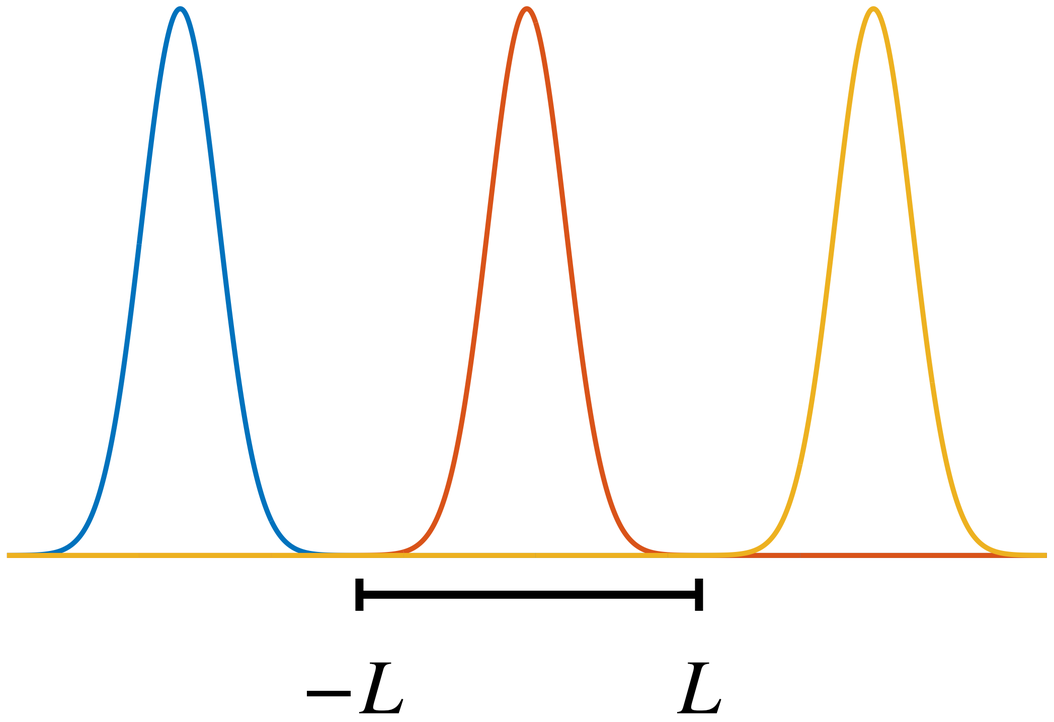


Locality, in a periodic domain?

How do you define the center of an orbital? Which one is it?

Position operator in periodic domains, see [Resta 1998]

How do you define a second moment sensibly?



A simple 1d example

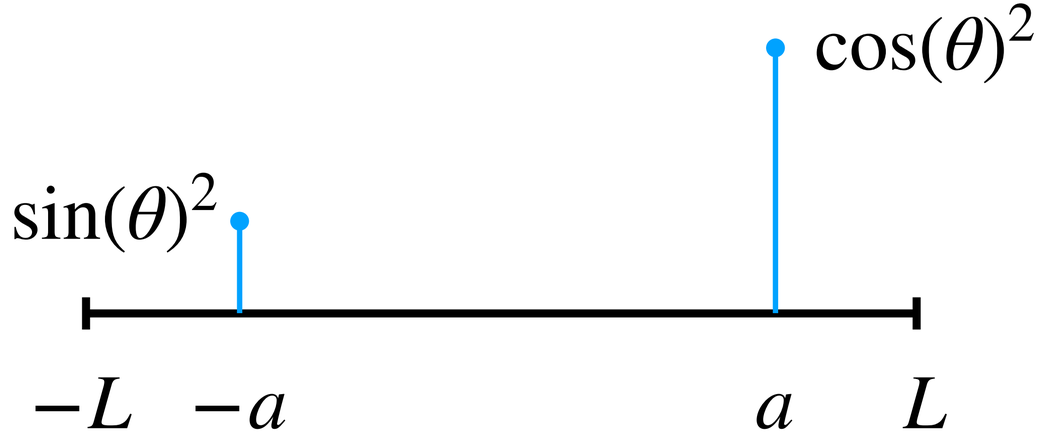
Consider a function with two peaks on $[-L, L)$:

$$\rho(r) = \sin(\theta)^2\delta(r + a) + \cos(\theta)^2\delta(r - a))$$

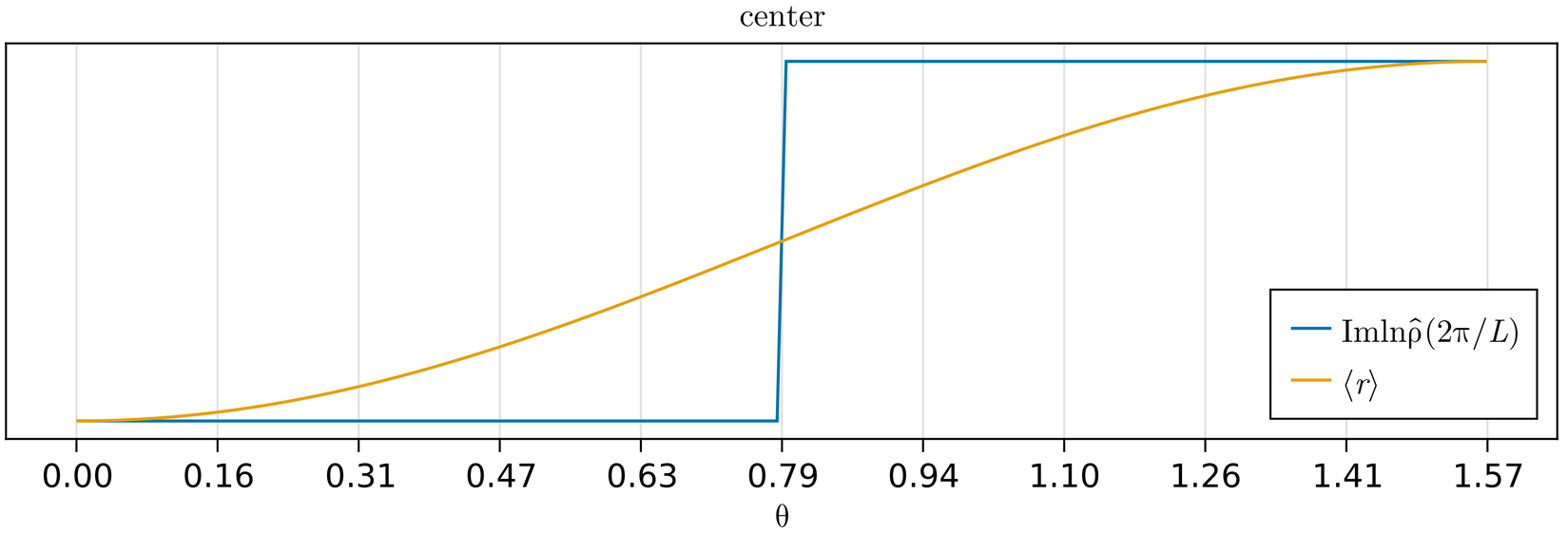
What should the center be as θ varies?

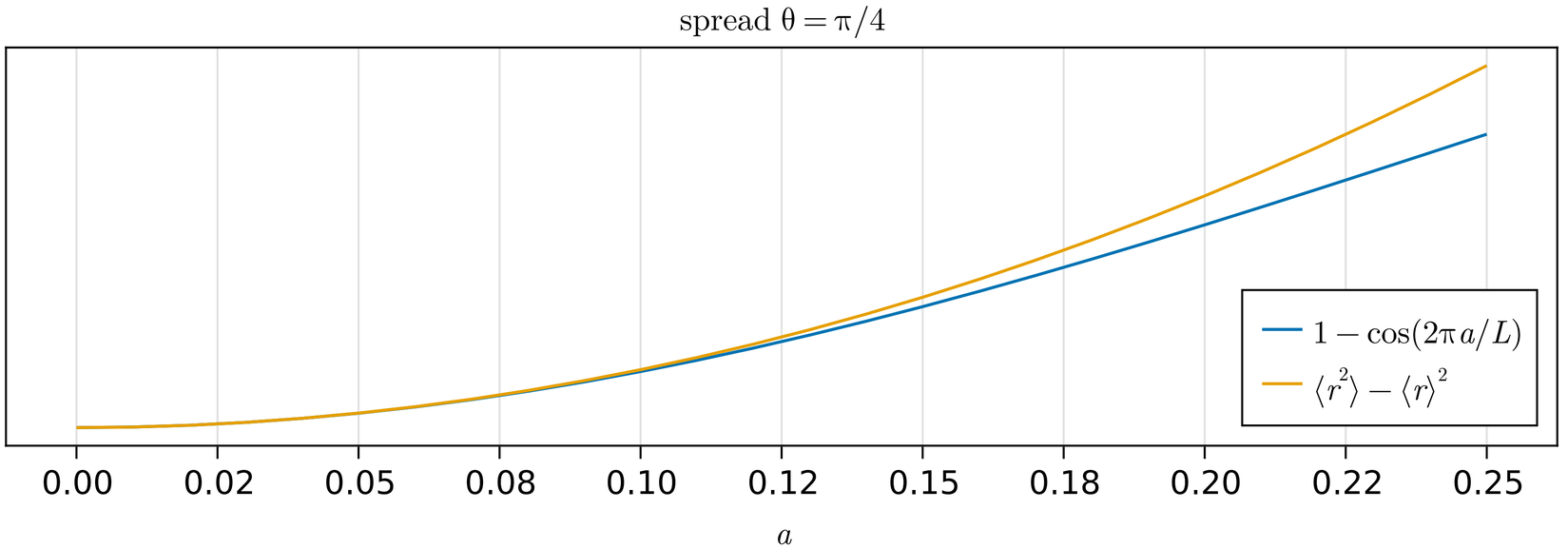
What about the spread?

Admittedly a bit contrived, but illustrative

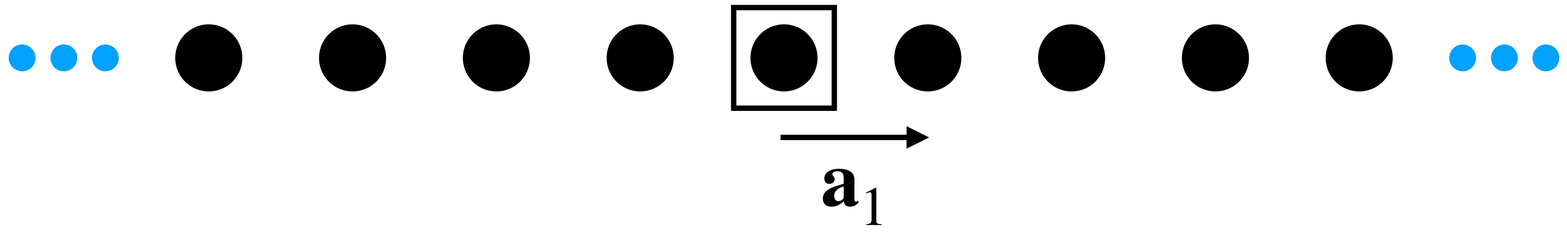


Real space vs periodic approximation





Some necessary notation



Periodic atomic structure

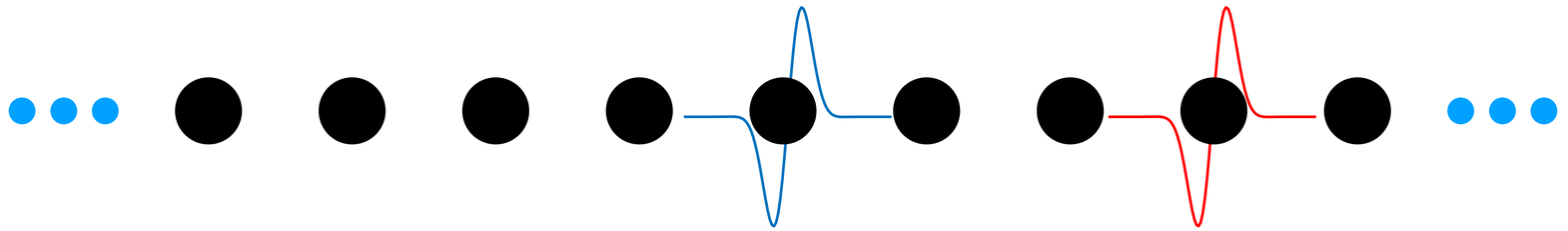
Potential $v(\mathbf{r})$ satisfies $v(\mathbf{r} + \mathbf{a}_1 n_1) = v(\mathbf{r})\ \forall \mathbf{r} \in \mathbb{R}, n_i \in \mathbb{Z}$

So-called Bravis lattice: $\mathbb{L} = \{\mathbf{R} \mid \mathbf{R} = \mathbf{a}_1 n_1, n_1 \in \mathbb{Z}\}$

Unit cell: $\Gamma = \{\mathbf{r} \mid \mathbf{r} = c\mathbf{a}_1, -1/2 \leq c < 1/2\}$

Block orbitals $\psi_{j,\mathbf{k}} = e^{i\mathbf{r}\cdot\mathbf{k}} u_{j,\mathbf{k}}(\mathbf{r})$; $u_{j,\mathbf{k}}$ is $\mathbb{L}$ periodic

Periodic copies and Fourier transforms



Define Wannier functions spatially

$$\phi_{j,\mathbf{R}}(\mathbf{r}) = \frac{1}{|\Gamma^*|} \int_{\Gamma^*} \tilde{\psi}_{j,\mathbf{k}}(\mathbf{r}) e^{-i\mathbf{k}\cdot\mathbf{R}} d\mathbf{k}$$

Smooth in $\mathbf{k}$ $\tilde{\psi}_{j,\mathbf{k}}$ (similarly $\tilde{u}_{j,\mathbf{k}}$) implies $\phi_{j,\mathbf{k}}$ is highly localized

The MLWF approach (abridged history)

Marzari and Vanderbilt [1997]: optimize *localization criteria* Ω inspired by Foster-Boys

Per Blount [1962] can reformulate Ω in $\mathbf{k}$ space via:

$$\langle u_{m,\mathbf{k}} \,|\, \nabla_{\mathbf{k}} \,|\, u_{n,\mathbf{k}} \rangle \text{ and } \langle u_{m,\mathbf{k}} \,|\, \nabla^2_{\mathbf{k}} \,|\, u_{n,\mathbf{k}} \rangle$$

Propose using finite differences in $\mathbf{k}$ space via "shells" defined by $\mathbf{b}$

Can compute everything using $M_{mn}^{\mathbf{k},\mathbf{b}} = \langle u_{m,\mathbf{k}} \,|\, u_{n,\mathbf{k}} \rangle$

Resta [1998] defined a position operator for periodic boundary conditions; appeared earlier, e.g., Selloni et al. [1987] and Fois et al. [1988]

Consistent definition in the thermodynamic limit $L \to \infty$

Key challenges / shortcomings

Results in a center computed as $\frac{1}{N}\sum_{\mathbf{k},\mathbf{b}} w_{\mathbf{b}}\mathbf{b}\ \mathrm{Im}\log M_{n,n}^{\mathbf{k},\mathbf{b}}$

Discontinuous, can have “false” local minima (e.g., branch cut)

Slow convergence(?) of center and spread as $L \to \infty$ (equivalently, $\|\mathbf{b}\| \to 0$)

Does not admit error bounds, can be quite inaccurate in practice

Requires a “special” reduction to the Γ-point only case

Optimization may converge slowly

A new formulation: density convolution

Center of a density ρ: the point around which the second moment is minimized

$$\mathbf{c}^* \triangleq \arg\min_{\mathbf{c}} \int_{\infty}^{\infty} \rho(\mathbf{r})(\mathbf{r}-\mathbf{c})^2 \mathrm{d}\mathbf{r} = \int_{-\infty}^{\infty} \rho(\mathbf{r})\mathbf{r}\mathrm{d}\mathbf{r}$$

"Truncate" to get a spread definition with periodic boundary conditions

$$s_\rho(\mathbf{c}) \triangleq \int_{\mathbb{S}_\mathbf{c}} \rho(\mathbf{r})(\mathbf{r}-\mathbf{c})^2 \mathrm{d}\mathbf{r} = \int_{\mathbb{S}_\mathbf{0}} \rho(\mathbf{r}+\mathbf{c})\mathbf{r}^2 \mathrm{d}\mathbf{r}$$

Finally,

$$\mathbf{c}^* \triangleq \arg\min_{\mathbf{c}} s_\rho(\mathbf{c}) \quad \text{and} \quad \mathbf{s}^* \triangleq s_\rho(\mathbf{c}^*)$$

Spread in the Fourier domain

In an infinite domain, natural representation in Fourier space

$$\langle \phi_{j,\mathbf{R}} \,|\, r \,|\, \phi_{j,\mathbf{R}} \rangle \text{ to } \langle u_{j,\mathbf{k}} \,|\, \nabla_{\mathbf{k}} \,|\, u_{j,\mathbf{k}} \rangle$$

and

$$\langle \phi_{j,\mathbf{R}} \,|\, r^2 \,|\, \phi_{j,\mathbf{R}} \rangle \text{ to } \langle u_{j,\mathbf{k}} \,|\, \nabla^2_{\mathbf{k}} \,|\, u_{j,\mathbf{k}} \rangle$$

In a finite domain, simply truncate r and r^2 to $[-L, L)$

A “spectrally accurate” representation of the differentiation

Properties

Consistent in the limit $L \to \infty$

Has proper translation symmetry

The spread is continuous even if the center is not

A nice formulation, but hard to optimize directly (“non local”)

Admits systematic truncation to construct (recover) simpler formula

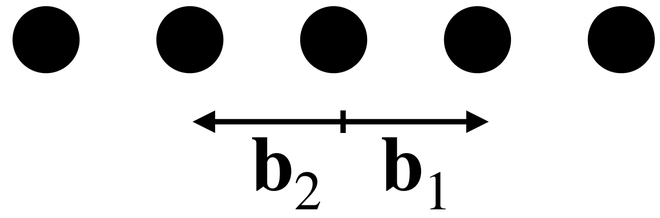
Essentially a spectral definition of the differentiation operators

Foster-Boys $\geq$ density convolution $\geq$ truncated density convolution

Systematic approximation

Following Marzari Vanderbilt pick “shells” $\mathbf{b}$ such that

$$\sum_{\mathbf{b}} w_b \mathbf{b}\mathbf{b}^T = I$$



Make the approximation (valid for local orbitals)

$$\mathbf{r}^2 = \sum_{\mathbf{b}} w_{\mathbf{b}} (\mathbf{b}^T \mathbf{r})^2 \approx \sum_{\mathbf{b}} 2 w_{\mathbf{b}} \mathrm{Re}(1 - e^{-i\mathbf{b}^T \mathbf{r}})$$

Yields the spread approximation (don’t know the center)

$$s_\rho(\mathbf{c}) = \sum_{\mathbf{b}} 2 w_{\mathbf{b}} \mathrm{Re}(1 - \hat{\rho}(\mathbf{b}) e^{i\mathbf{b}^T \mathbf{c}})$$

Truncated density convolution (TDC)

Solve for the center

$$\mathbf{c}^* = -\sum_{\mathbf{b}} w_{\mathbf{b}} \mathbf{b}\ \mathrm{Im} \log \hat{\rho}(\mathbf{b})$$

and compute the spread

$$s^* = \sum_{\mathbf{b}} 2w_{\mathbf{b}}(1 - |\hat{\rho}(\mathbf{b})|)$$

To connect with prior work

$$\hat{\rho} = \frac{1}{N} \sum_{\mathbf{k}} M_{n,n}^{\mathbf{k},\mathbf{k}+\mathbf{b}}$$

The spread formula appeared in Stengel and Spaldin [2006]

The approximation, informally

Roughly (but not exactly) corresponds to the approximations:

$$r \approx \sin(r)$$

$$1 - r^2 \approx \cos(r)$$

Good near the origin, but does deteriorate

Could be systematically improved (higher order truncation)

Similar to Von Neumann analysis for finite difference schemes

Comparison with Marzari Vanderbilt

How different is this? We have

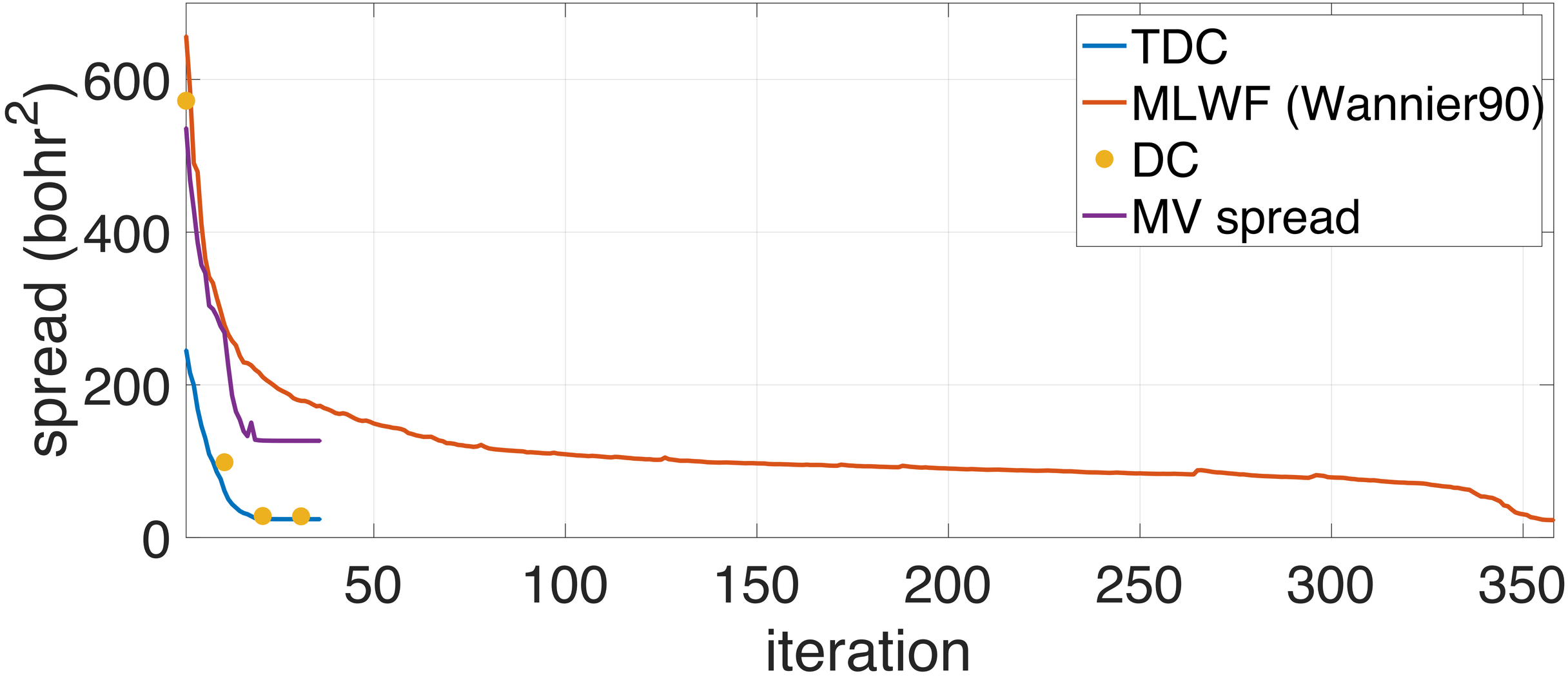
$$\mathbf{c}^* = -\sum_{\mathbf{b}} w_{\mathbf{b}} \mathbf{b}\, \mathrm{Im} \log \frac{1}{N} \sum_{\mathbf{k}} M_{n,n}^{\mathbf{k},\mathbf{k}+\mathbf{b}}$$

and compare with

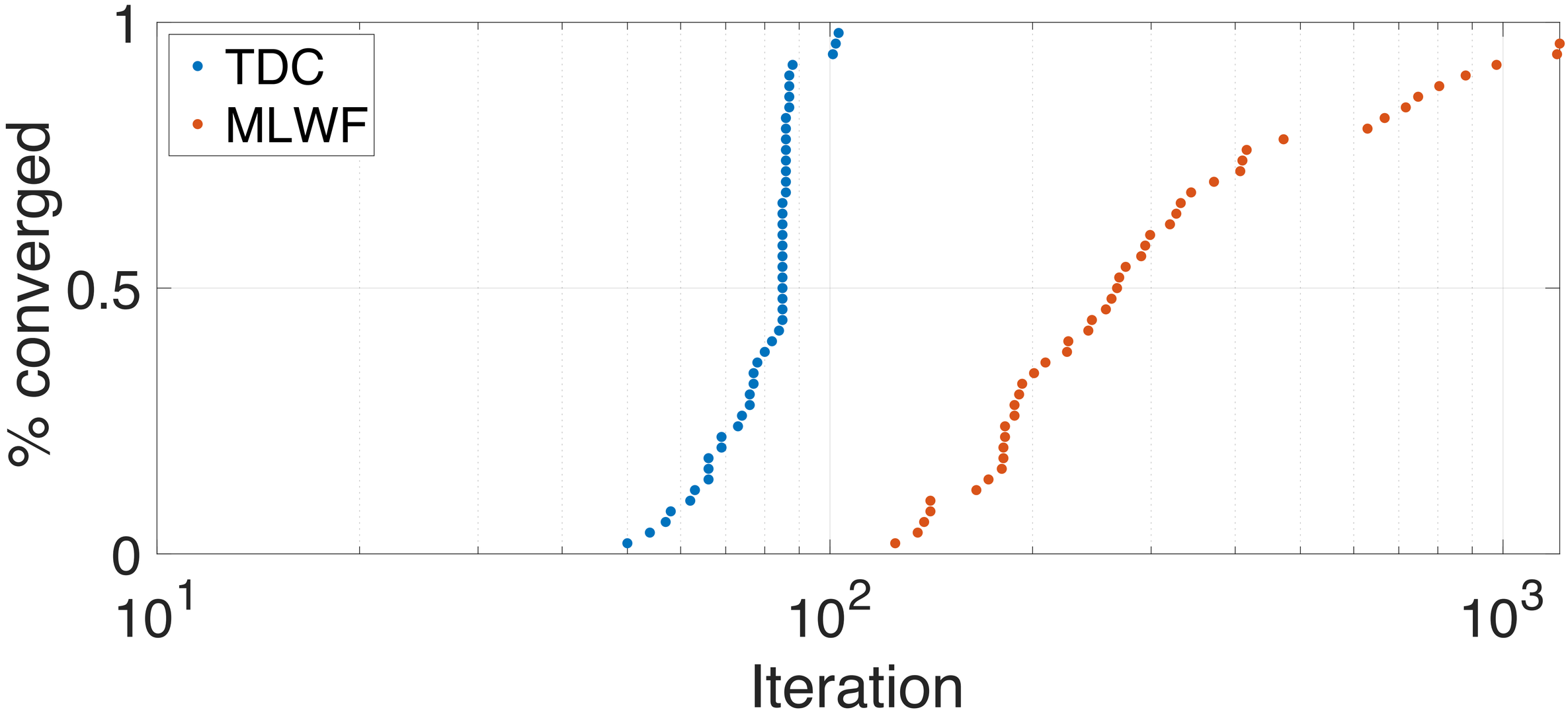
$$\mathbf{c}^*_{\mathrm{MLWF-FD}} \approx -\frac{1}{N} \sum_{\mathbf{k}} \sum_{\mathbf{b}} w_{\mathbf{b}} \mathbf{b}\, \mathrm{Im} \log M_{n,n}^{\mathbf{k},\mathbf{k}+\mathbf{b}}$$

Not the same, but similar for local functions—spread formula matters more

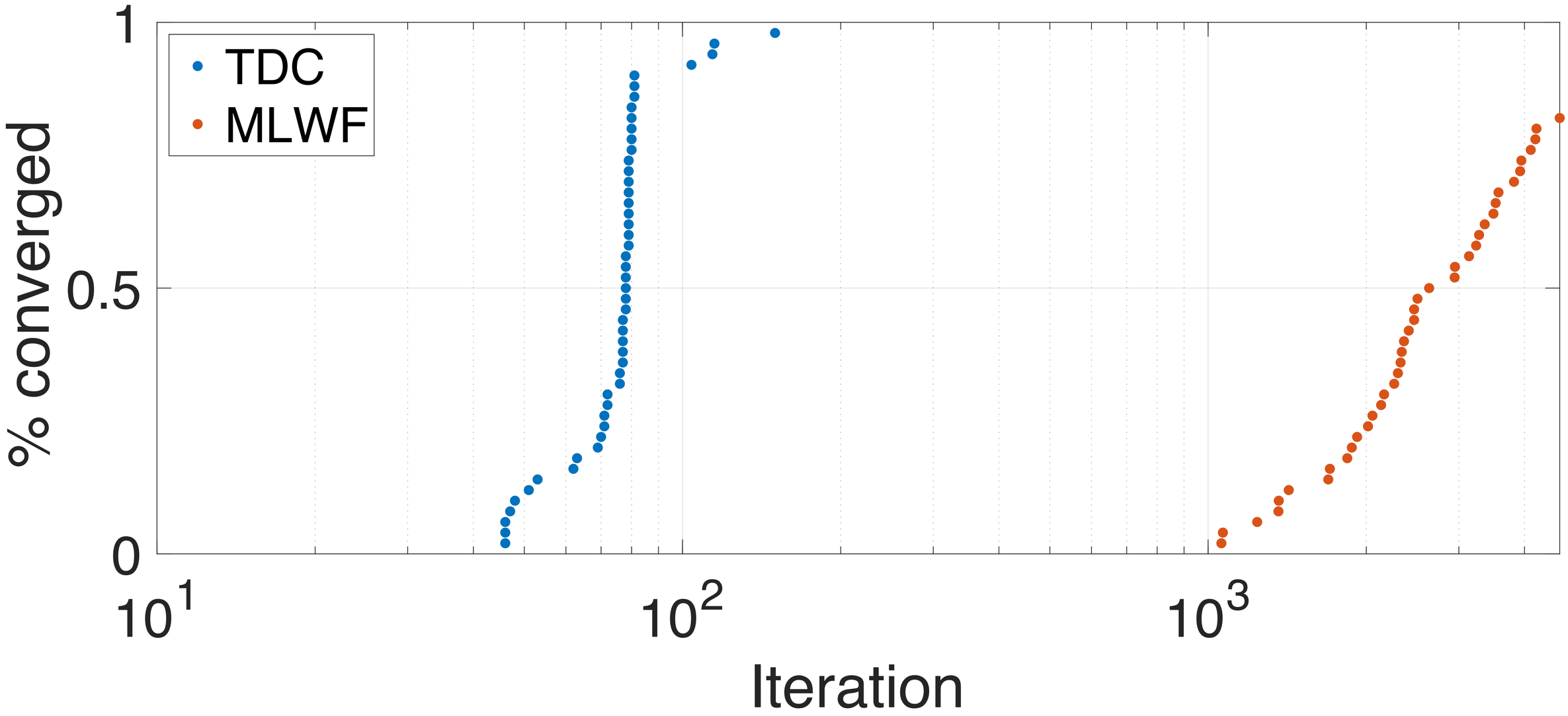
A single optimization trajectory



Random initializations $BaTiO_3$



Random initializations Cr_2O_3



MLWF didn't really converge here, this is to within 1% of the best spread

Outlook

Experiments suggest the TDC provides a more robust objective for localization

Converges as it should for a smooth objective

Julia codes forthcoming: WTP.jl and SCDM.jl

Some variations of the formula we recovered have appeared, but always seemed to be considered “equivalent”

Initialization: SCDM

Do we always need to optimize?

Betteridge’s law of headlines: no

Selected columns of the density matrix (SCDM):

- A direct method for localization
- Can serve as an initial guess for optimization
- Condensed phase (+ k-points and/or entangled bands) and molecular systems
- Robust and efficient: part of QE + Wannier90 and more
- Parameter free (except in the entangled case)

SCDM—the algorithm ($N = N_w$ and Γ-point only)

1. Compute the column-pivoted QR factorization: $\Psi^*\Pi = QR$
2. Let $\mathscr{C}$ denote the first N elements selected by the permutation Π
3. Solve $\min_{Q^*Q=I} \|\Psi Q - \Psi(\Psi^*)_{:,\mathscr{C}}\|_F$ (this is Löwdin orthogonalization)

$\Phi = \Psi Q$ will be well localized

QRCP paper [Golub and Businger 1965], textbook [Golub and Van Loan 1996]

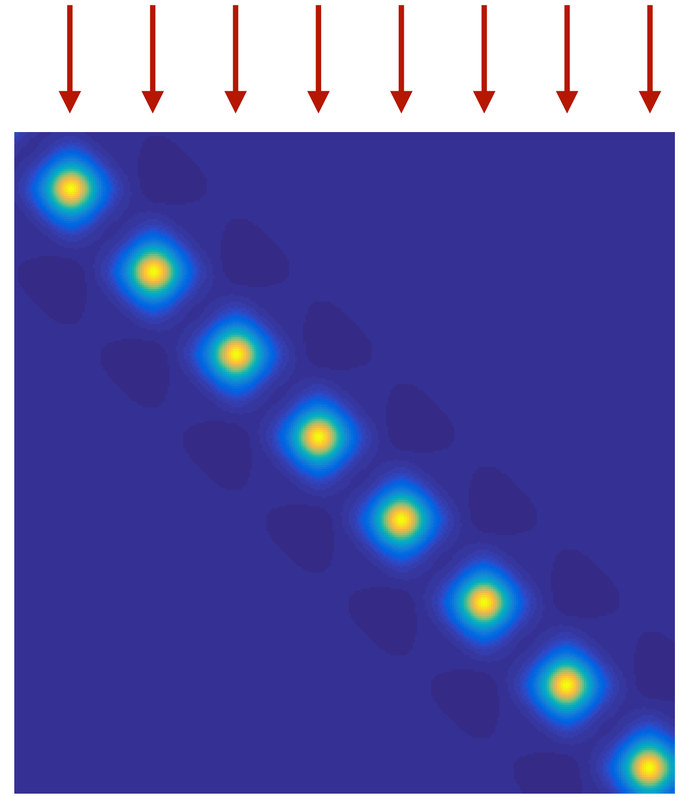
SCDM—why it works

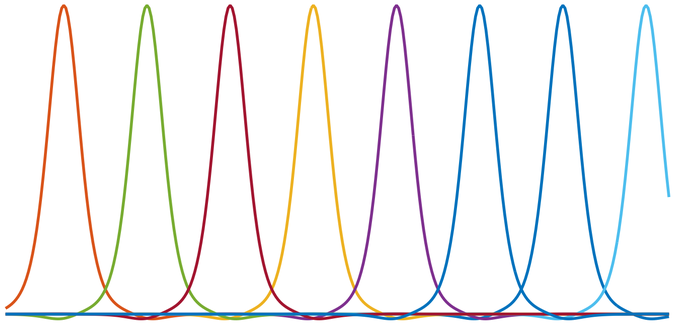
For insulating systems $P = \Psi\Psi^*$ has localized columns [Kohn 1959 and 1996]

Pivoted QR identifies N columns of P to serve as models for localized orbitals

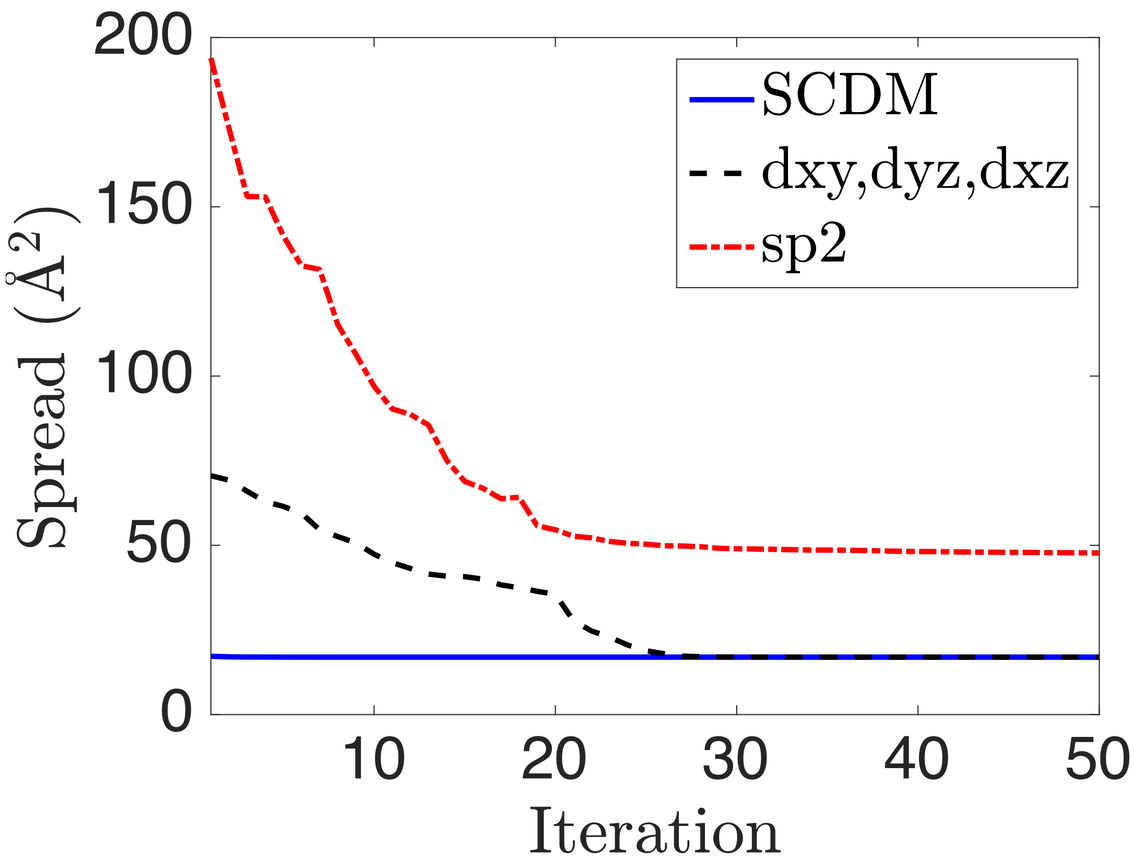
> All columns are well localized, also need the set to be well conditioned

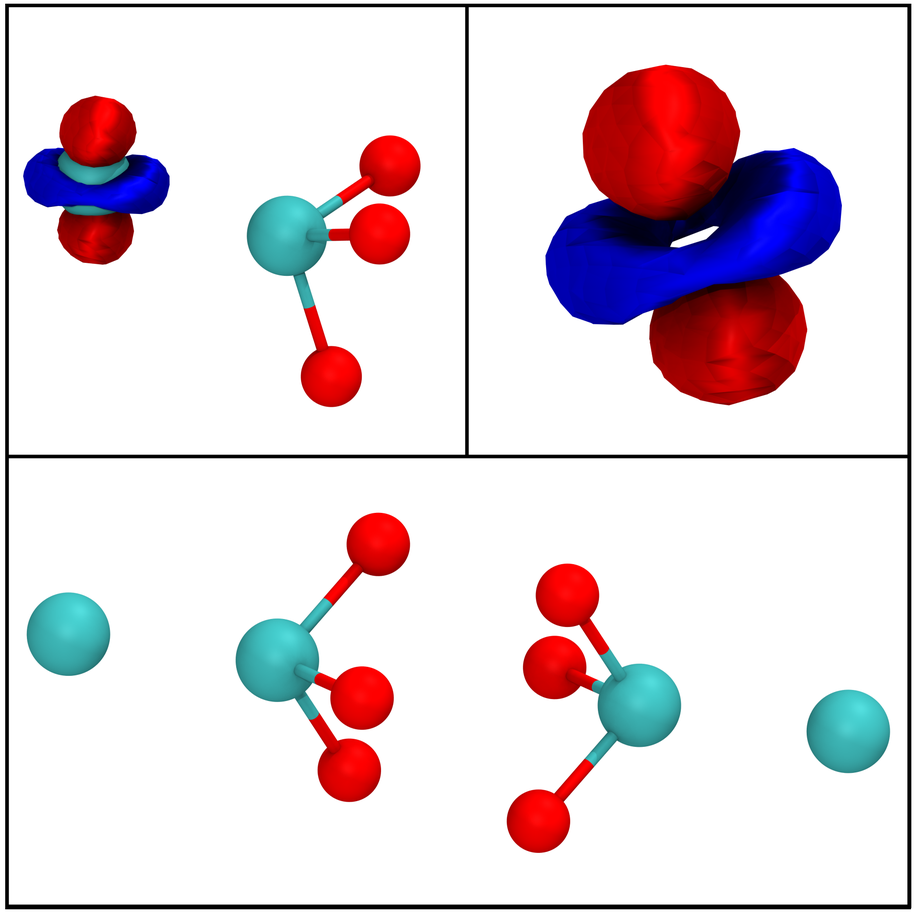
Orthogonalization of a well-conditioned subset doesn't delocalize the functions much





SCDM—in practice for Cr_2O_3





What if $\mathscr{I}$ is not isolated?

Use a quasi-density matrix

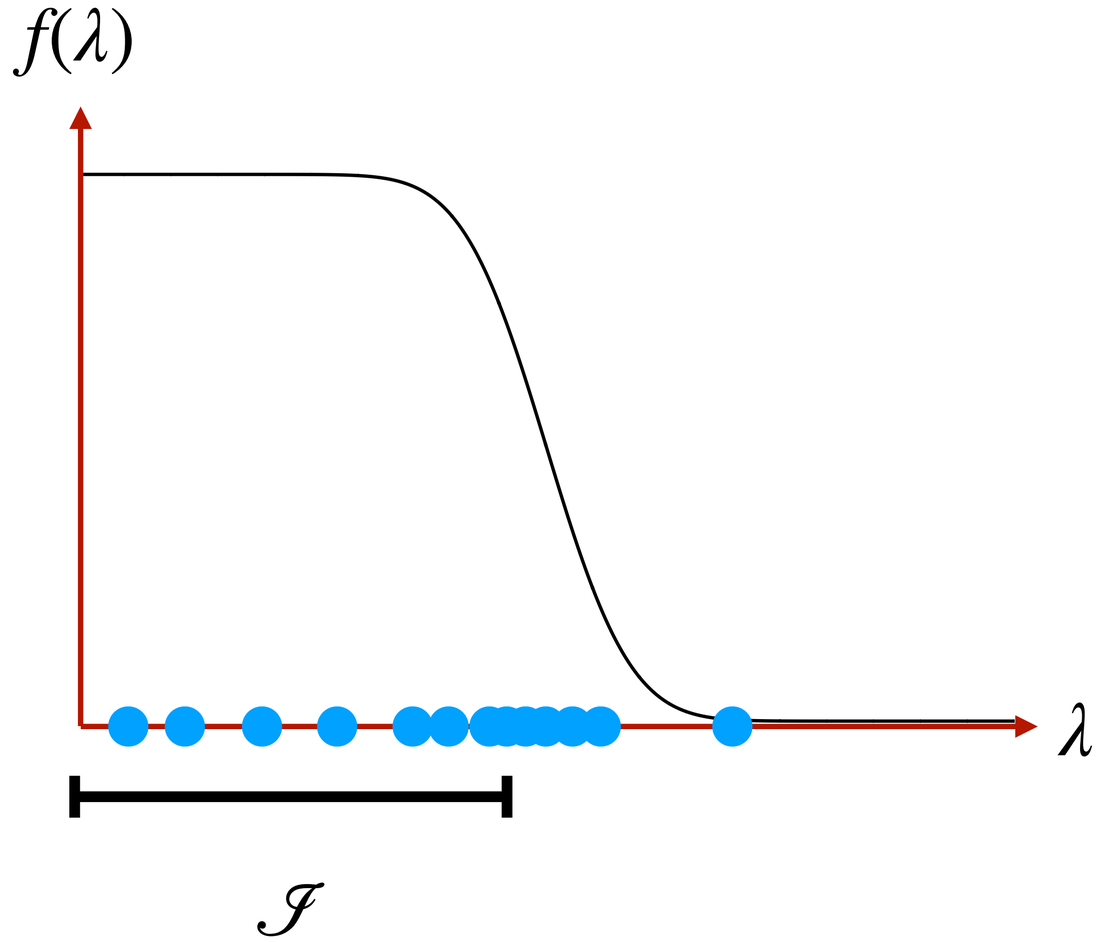
$$\widehat{P} = \sum_i \psi_i f(\lambda_i) \psi^*$$

Compute the CPQR

$$f(\Lambda)\Psi^*\Pi = QR$$

Let $\mathscr{C}$ be the first N_w selected columns

Solve $\min_{Q^*Q=I} \|\Psi Q - \Psi(f(\Lambda)\Psi^*)_{:,\mathscr{C}}\|_F$



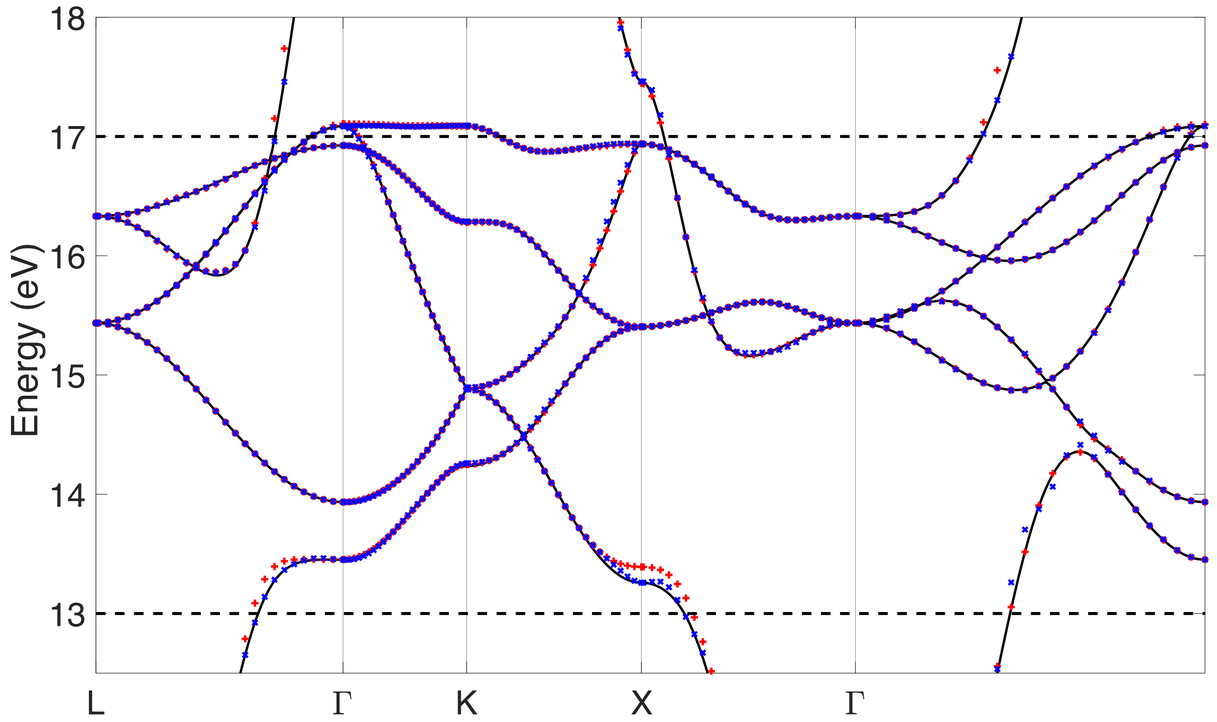
In practice, only keep eigenvalues in Λ such that $|f(\lambda_i)| > \epsilon$; quasi-density matrix will have well-localized columns

Extension to $\mathbf{k}$-points

Pick the columns at the Γ-point and use them for all $\mathbf{k}$ (so one QRCP)

Solve for the gauge at each $\mathbf{k}$ using those same columns

Can also be paired with the entangled version



10x10x10 k-point grid, reference calculation in black, SCDM in blue, wannier90 in red

Justified by Panati and Pisante [2013], Nenciu [1991], des Cloizeaux [1964], etc.

What about an atomic orbital basis?

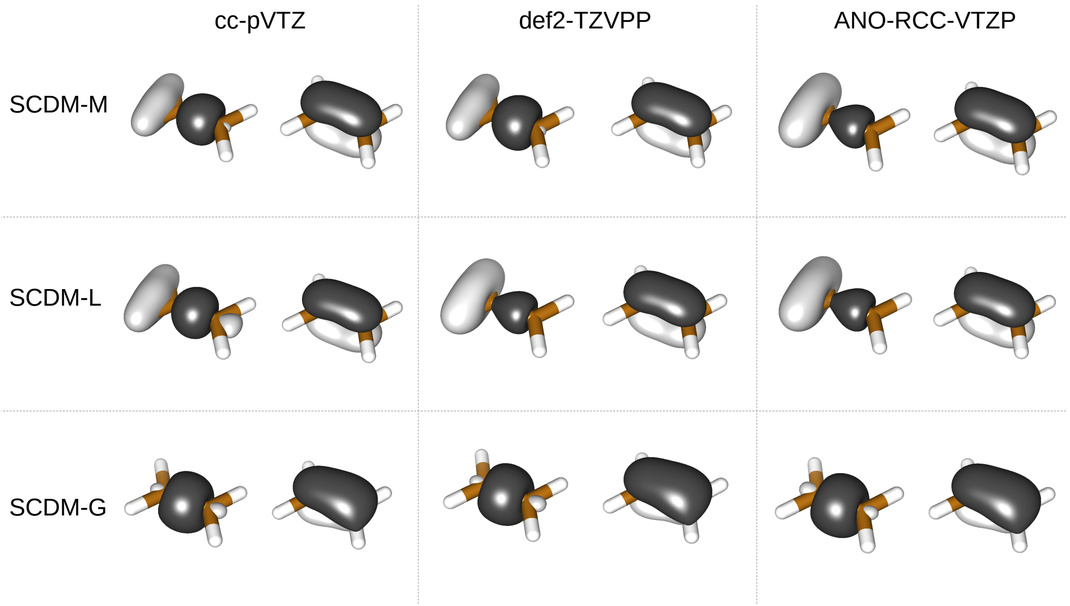
Two natural extensions:

- Apply SCDM to P in an AO basis (really PS or $S^{1/2}PS^{1/2}$

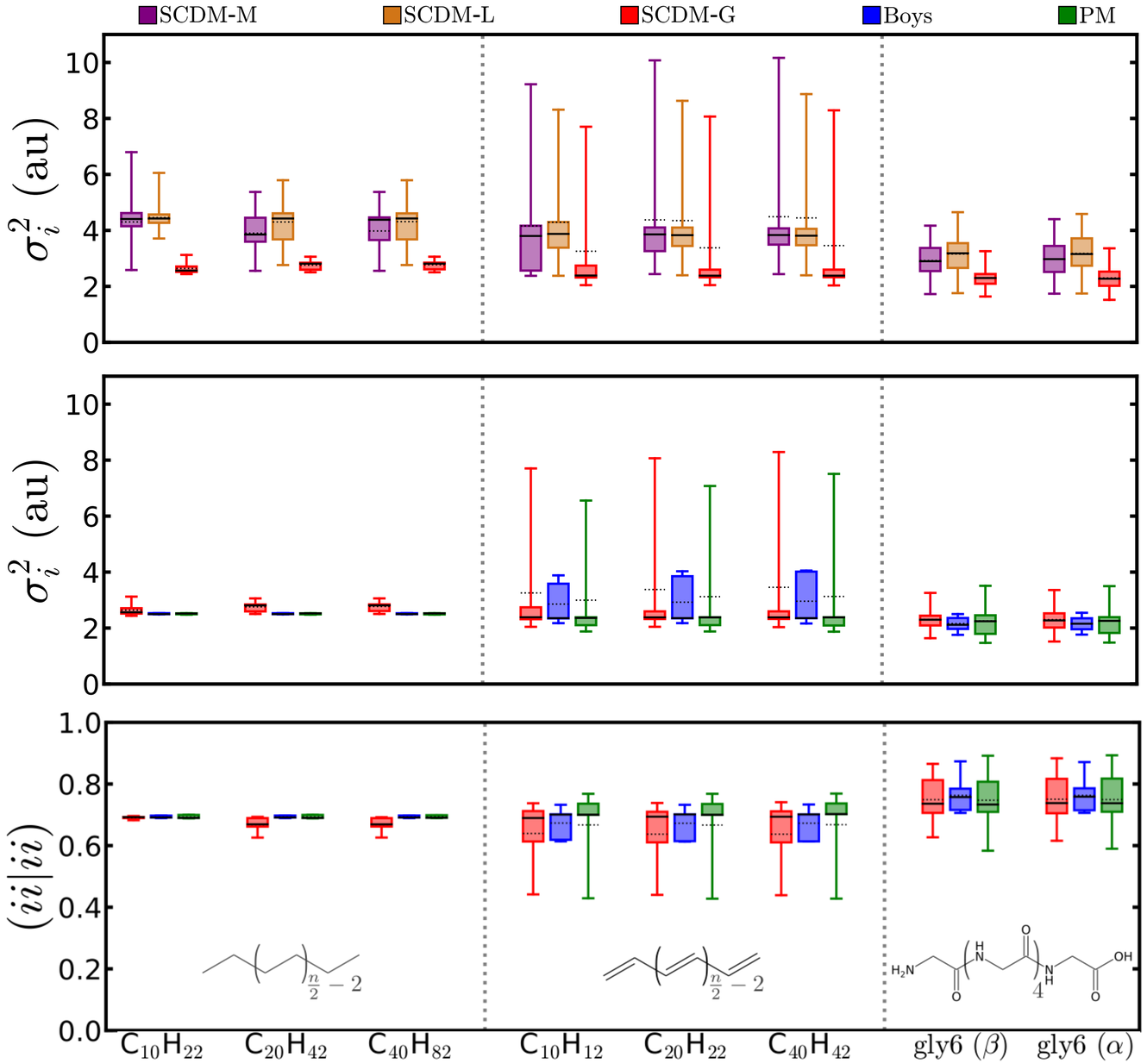
  - Corresponds to picking well-conditioned *projected atomic orbitals*

- Sparse, real space grid evaluation of Ψ plus "standard" SCDM

  - Can easily transform back into AO basis



Atomic orbital basis sets



Discussion

Orbital centers / spreads in periodic systems are ambiguous

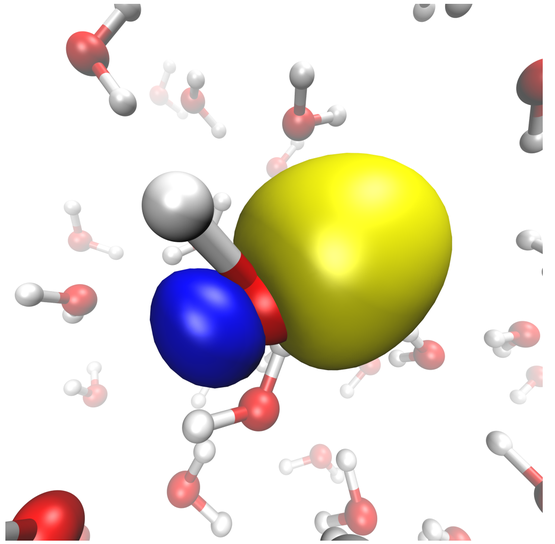
Careful definition provides a starting point for localization

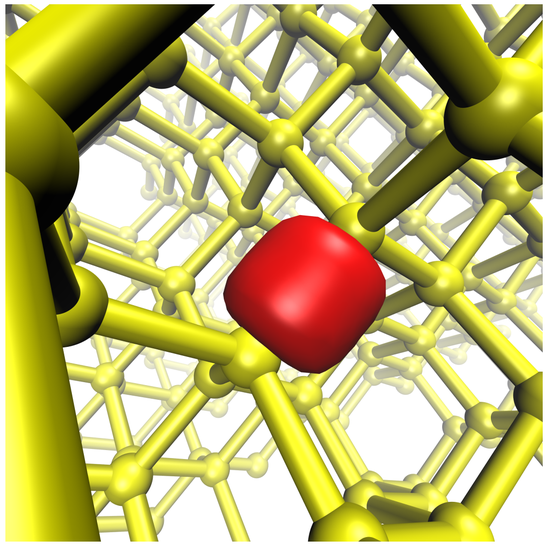
The objective function can have a significant impact on convergence of the optimization procedure

SCDM provides a robust initialization for localization

- Can even be used in isolation

Robust computation of localized orbitals is important in many situations where automation is important, e.g., high-throughput calculations (Vitale et al. [2020])





Selected references

cs.cornell.edu/~damle

1. Kangbo Li, Hsin-Yu Ko, Robert DiStasio Jr, and Anil Damle *in preparation*

2. Eric G Fuemmeler, Anil Damle, Robert A DiStasio Jr "Selected columns of the density matrix in an atomic orbital basis I: an intrinsic and non-iterative orbital localization scheme for the occupied space," *accepted* arXiv:2108.06399

3. Anil Damle and Lin Lin "Disentanglement via entanglement: A unified method for Wannier localization," SIAM Multiscale Modeling and Simulation, 2018, 16 (3), 1392-1410

4. Anil Damle, Lin Lin, Lexing Ying, "SCDM-k: Localized orbitals for solids via selected columns of the density matrix," Journal of Computational Physics Volume 334, 1 April 2017, 1-15