Stochastic optimization of block Toeplitz matrices

Hylleraas Seminar, 2017

Audun Skau Hansen

The Hylleraas Centre for Quantum Molecular Science
University of Oslo

November 30, 2017

Outline

Project Overview (I)

My PhD project:

- Part of the "Coupled Cluster for Periodic Systems" project.
- Funding from NFR Frirpro.
- Supervised and initiated by Thomas Bondo Pedersen at the Hylleraas Centre, UiO (formerly CTCC).

Project Overview (II)

Some motivation for this presentation:

- Toeplitz matrices:
  - Periodic many-body systems may be neatly described by means of Toeplitz matrices.
  - Optimization and manipulations of such systems may be handled in the same framework.
- Stochastic optimization
  - Generating random unitary matrices
  - Iteratively improve the solution by means of random unitary rotations.

The Infinite Block Toeplitz matrix

Consider a matrix $\mathbf{A}$ with blocks:

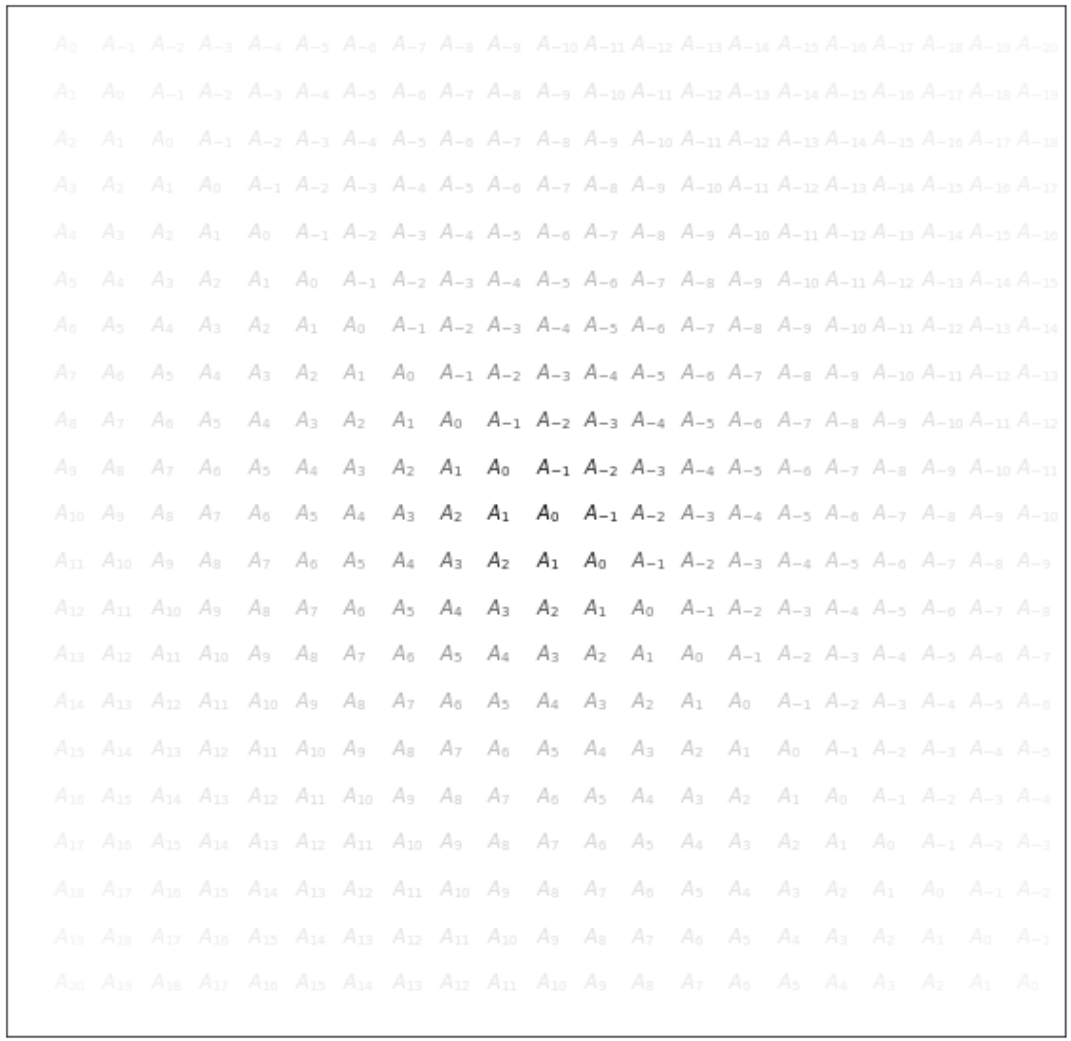
$$\mathbf{A}_{L,M} = A_{L-M}, L, M \in \mathbb{Z} \tag{1}$$

The following symmetry is intrinsic to the definition

$$\mathbf{A}_{L+N,M+N} = A_{L+N-M-N} = A_{M,N}, \tag{2}$$

so that all blocks along the diagonal are the same.

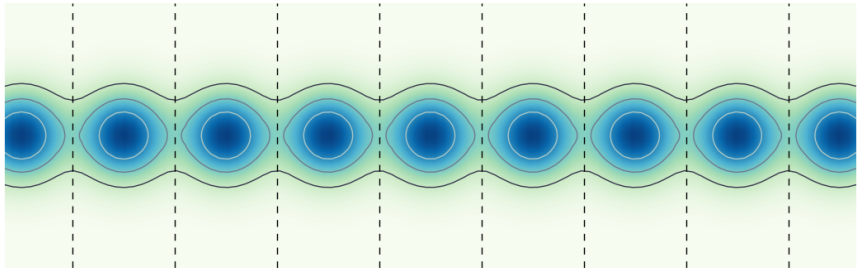
The Infinite Block Toeplitz matrix



The Infinite Block Toeplitz matrix and periodic systems I

Translationally invariant matrix elements in periodic systems may be expressed with infinite block-Toeplitz matrices:

$$\langle L\mu|\hat{A}|M\nu\rangle := \mathbf{A}^{LM}_{\mu\nu} \tag{3}$$



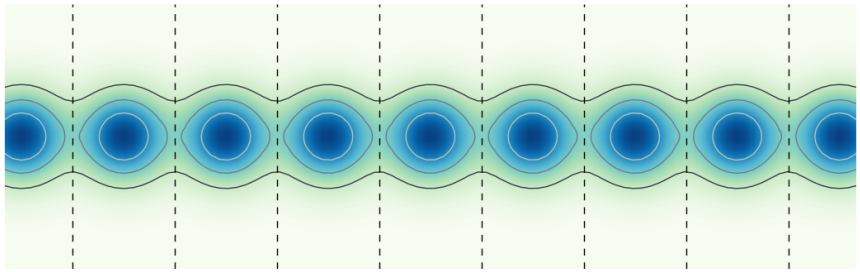
Being *translationally invariant* means that:

$$(\langle L|\hat{A}|M\rangle)_{\mu\nu} = (\langle L|\hat{A}\hat{T}_M|0\rangle)_{\mu\nu} = (\langle L-M|\hat{A}|0\rangle)_{\mu\nu}. \tag{4}$$

The same matrix as a block Toeplitz structure:

$$\mathbf{A}^{LM}_{\mu\nu} = \mathbf{A}^{L-M,0}_{\mu\nu}. \tag{5}$$

The Infinite Block Toeplitz matrix and periodic systems II



A periodic overlap matrix is in principle infinite

$$\mathbf{S}_{\mu\nu}^{0M} \tag{6}$$

$$\mathbf{S}_{\mu\nu}^{L,M} = \mathbf{S}_{\mu\nu}^{L-M,0}. \tag{7}$$

Toeplitz transpose

The blocks in a transposed block Toeplitz matrix $\mathbf{A}^T$ are

$$(\mathbf{A}^T)_{ij} = A^T_{j-i} \tag{8}$$

Some Toeplitz matrices also exhibits symmetries:

$$A_{i-j} = A^T_{j-i}, \tag{9}$$

and may conventionally be referred to as *symmetric* block Toeplitz matrices.

Toeplitz matrix-matrix product I

The matrix-matrix product of two such matrices is

$$(\mathbf{AB})_{ij} = \sum_{k=-\infty}^{\infty} A_{i-k} \cdot B_{k-j}, \tag{10}$$

and since

$$(\mathbf{AB})_{(i+n),(j+n)} = \sum_{k=-\infty}^{\infty} A_{(i+n)-k} B_{k-(j+n)} = \sum_{k'=-\infty}^{\infty} A_{i-k'} B_{k'-j} = (\mathbf{AB})_{ij}, \tag{11}$$

we may conclude that $\mathbf{AB}$ is itself block-Toeplitz.

Toeplitz matrix-matrix product II

The matrix-matrix product may now be expressed simply as

$$(AB)_i = \sum_{k=-\infty}^{\infty} (A_{i-k}) \cdot B_k. \tag{12}$$

Any truncation in the number of non-zero blocks in A and B above will make the summation finite and thus computationally feasible.

Spectral decomposition of infinite block Toeplitz matrices

In the case of a truncation, there will be a finite number M of non-zero blocks in the matrix A_{i-j}. The transformation to reciprocal space is achieved through the DFT (Discrete Fourier Transform), where $m \in \mathbb{Z}$

$$\widetilde{A}_m = \sum_n e^{-i(\frac{2\pi}{M})mn} A_n \tag{13}$$

Spectral decomposition of infinite block Toeplitz matrices

Equivalently, the DFT-transformation may be seen as a matrix-vector product:

$$\widetilde{A}_m = \sum_n W_{mn} A_n, \tag{14}$$

where the elements of W are

$$W_{mn} = \left(e^{-i\frac{2\pi}{M}}\right)^{mn}. \tag{15}$$

$$W^{-1}_{nm} = \frac{1}{M}\left(e^{i\frac{2\pi}{M}}\right)^{mn}. \tag{16}$$

(Strictly speaking, the structure we get back from k-space is a infinite block *circulant* matrix.)

Truncated Toeplitz visual

Unitary infinite block Toeplitz matrices

Certain infinite block Toeplitz matrices fullfill the following relation (assuming the blocks are real):

$$\mathbf{U}^T\mathbf{U} = \mathbf{U}\mathbf{U}^T = \mathbf{1}, \tag{17}$$

meaning they are unitary.

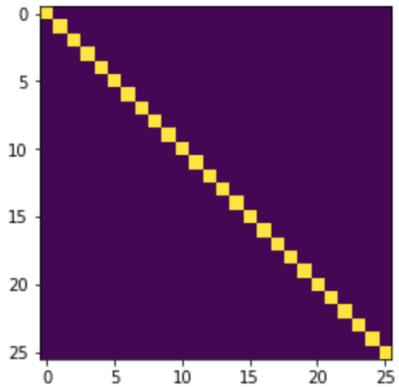
Unitary infinite block Toeplitz matrices

Some notable things about these Toeplitz unitary transformations:

- An optimization procedure resulting in a new set of orbitals may be expressed as a block Toeplitz unitary transformation of the original set.
- The translational symmetry is conserved in the transformation (product of two Toeplitz matrices is itself a Toeplitz matrix).
- Restrictions may be imposed, for example to conserve the Hartree-Fock condition ($f^{ia} = 0$), reducing the number of possible choices of U.

Toeplitz implementation

```
import common_tests as ct

project_folder = "/Users/audunhansen/papers/DEC-POC/ethylene/1d_rotations/ethylene_column"

C = ct.latmat()

C.npload(project_folder + "/crystal_reference_state.npy", project_folder + "/crystal_reference_coords.npy")

S = ct.latmat()

S.npload(project_folder + "/crystal_overlap_matrix.npy", project_folder + "/crystal_overlap_coords.npy")

Smo = C.tT()*S.t()*C

plt.imshow(Smo.get([0,0,0])) #cell coordinate: 0,0,0
plt.show()
```



K-space transform

In [29]:
```
C = ct.latmat()

C.npload(project_folder + "/lsdalton_reference_state.npy", project_folder + "/lsdalton_reference_coords.npy")

C_k = C.dft() #the k-space transofmr

print(C_k.get([2,0,0]))
```

```
[[  9.93294597e-04 +9.10583270e-04j   5.02290668e-03 -9.10582957e-04j
   -2.57337598e-02 -6.70714665e-06j   1.96540735e-02 +6.70727943e-06j
   -3.33215269e-01 +6.70720913e-06j  -1.39268316e-02 -1.30658731e-03j
    1.88805823e-02 -6.70726743e-06j   9.54411830e-03 -1.30658711e-03j
    6.23943016e-02 -1.53236788e-04j  -3.60904307e-02 -1.71795853e-05j
   -7.34000101e-01 +1.71747898e-05j   3.30726315e-02 +2.96364873e-05j
   -3.04107987e-02 +2.96334597e-05j  -3.00284734e-01 -1.59333923e-03j
   -4.99119159e-02 +3.12481906e-04j  -1.53973302e-01 -4.91939976e-04j
   -1.11641209e-01 +5.86759237e-03j  -6.35029034e-02 +4.91954254e-04j
    4.05844934e-02 -1.20638403e-03j  -1.06700823e-01 +2.60450515e-03j
   -2.89936834e-02 +1.76411754e-03j  -8.49668701e-02 +7.28669350e-04j
    1.35008408e-01 +1.53224410e-04j  -1.14829175e+00 -7.28647203e-04j
    3.31059841e-02 +1.76824014e-03j  -9.86966220e-02 +2.60445134e-03j]
 [  1.35417080e-02 -2.36687097e-04j   1.30966129e-02 +2.36686877e-04j
   -4.71377246e-02 +1.74330101e-06j   3.28248051e-02 -1.74330641e-06j
   -2.06269460e-01 -1.74327545e-06j  -3.53867076e-02 +3.39619939e-04j
    1.32916513e-02 +1.74331484e-06j   3.49803178e-02 +3.39619912e-04j
    3.43283731e-01 +2.47522716e-04j   1.86644680e-01 -1.44570929e-05j
    2.98242030e+00 +1.44375352e-05j   6.97986581e-02 -1.00631390e-06j
```

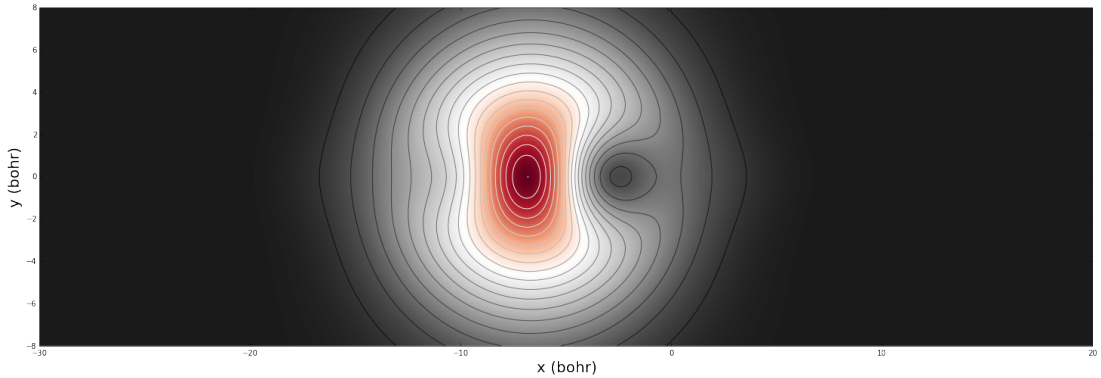
Summarizing so far

What's been presented so far:

- Periodic systems may be expressed using infinite block Toeplitz matrices.
- The conventional framework of linear algebra is easily extended to these matrices (matrix-matrix product, inversion, transpose and so on).
- Optimization of orbitals may be seen as a unitary rotation of the initial guess:

$$C^*_{optimized} = C_{original} U \tag{18}$$

Locality measures



- A prerequisite for calculating local correlation in extended systems is *local* orbitals (Wannier Functions).
- Several measures for locality are common, most notably perhaps the *orbital spread* σ:

$$\sigma_p^2 = \langle p|(\hat{r} - \langle p|\hat{r}|p\rangle)^2|p\rangle \tag{19}$$

- The optimization of objective functions based on measures of locality may be used to make the orbitals more suited for electronic correlation calculations such as MP2 and Coupled Cluster.

Localization

A range of objective functions is available for doing localization in molecules. The Foster-Boys objective function [1] minimize the sum over orbital spreads:

$$\xi_{Boys} = \sum_p \sigma_p^2, \tag{20}$$

where, as on the previous slide:

$$\sigma_p^2 = \langle p|(\hat{r} - \langle p|\hat{r}|p\rangle)^2|p\rangle. \tag{21}$$

This method has been extended to periodic systems in the Wannier-Boys scheme, as implemented by Claudio Zicovich-Wilson in Crystal. [5] [2].

Localization II

There is no consensus with regards to what constitutes the *optimal* localization function. A slight modification of the Foster-Boys objective function will for example give a penalty to "outliers":

$$\xi_{PSM} = \sum_{p} \left(\langle p | (\hat{r} - \langle p | \hat{r} | p \rangle)^2 | p \rangle \right)^m \tag{22}$$

In general, such a modification is named the PSM-function (powers of the second central moment). One advantage is that it mainly reduce the maximal orbital spread, which could potentially be further exploited when doing electron correlation.

Optimization

The optimization of objective functions represents a practical challenge, conventionally solved by iteratively descending towards a minimum. Such a descent may be achieved for example by

- Rotation of every pair of orbitals in the descending direction. (Jacobi sweeps[4])
- Computing first and higher order derivatives of the function, estimating the best route towards the minimum.

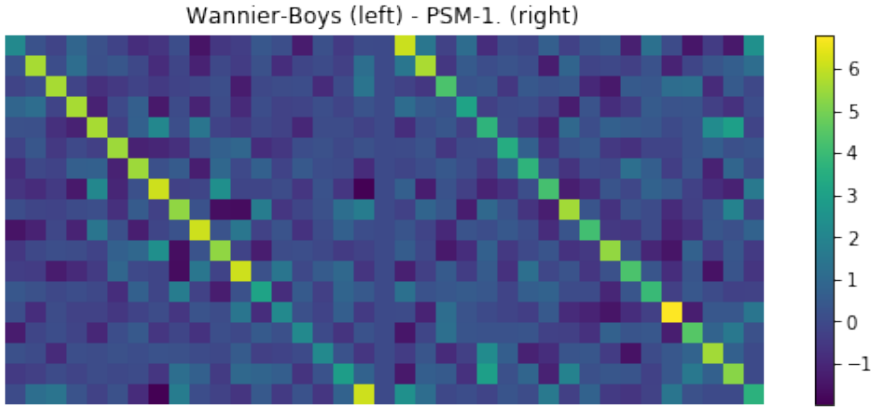
We have been investigating an alternative way of reaching the *global* minimum - optimization through *simulated annealing*.

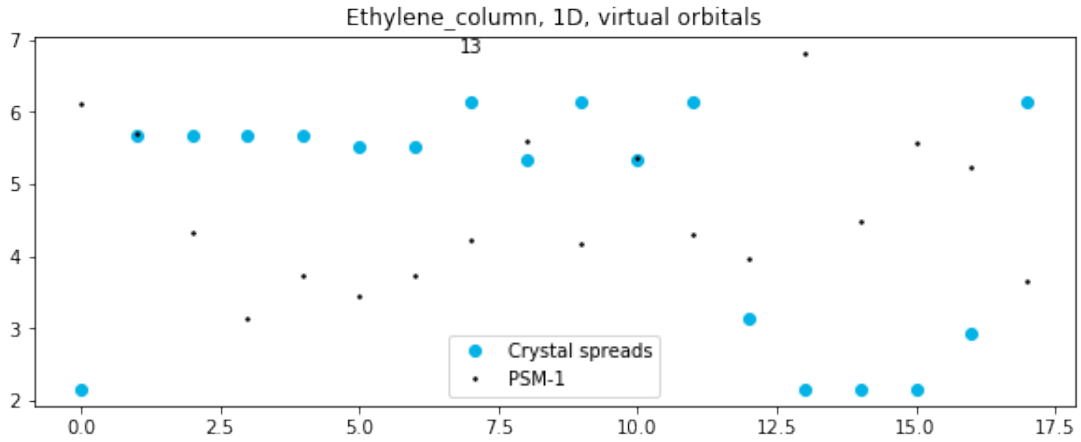
Simulated annealing - Algorithm

Simulated annealing for localization.

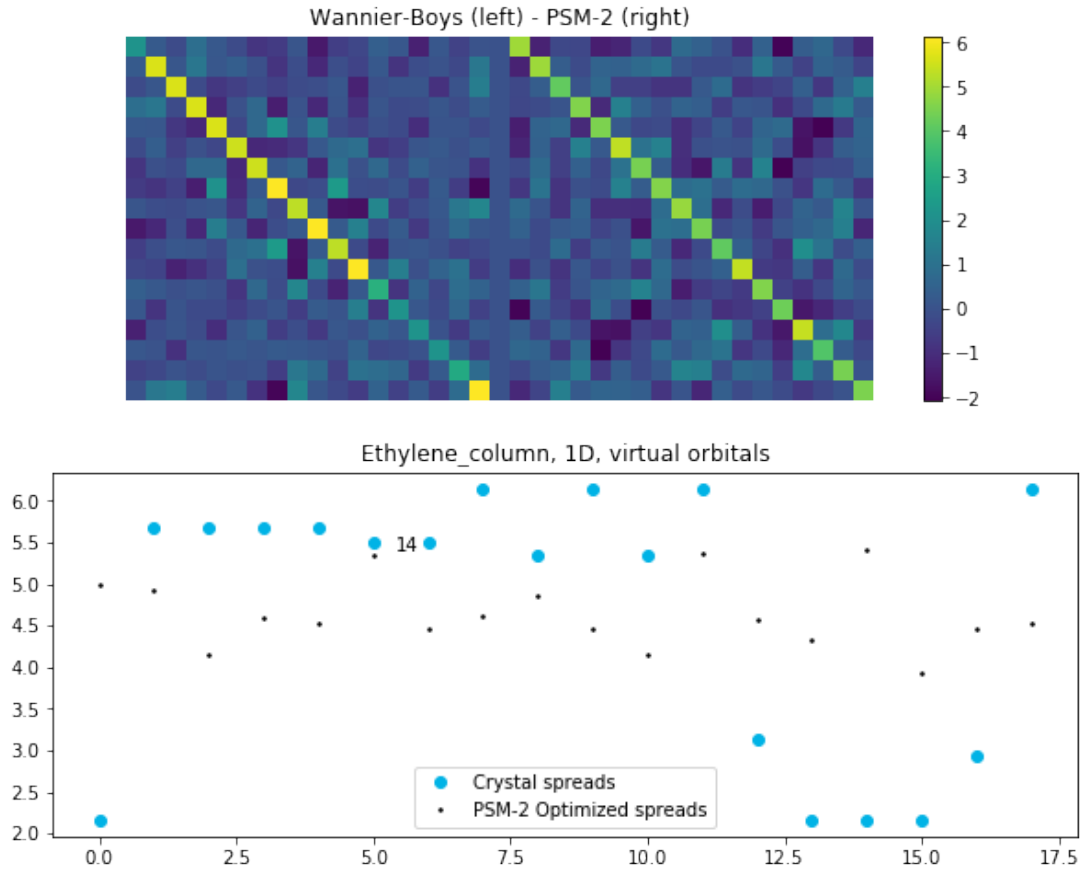
for input orbitals C, iteratively decreasing t, and initial unitary infinite block-Toeplitz matrix $U = 1$ **do**
    Generate a random unitary infinite block-Toeplitz matrix $\tilde{U}$
    Compute objective function $F(C\tilde{U})$
    if $C\tilde{U}$ minimize the function **then**
        Let $U \leftarrow \tilde{U}U$
    else Compute acceptance probability $P = e^{-\frac{\Delta F(C\tilde{U})}{t}}$:
        if $P > thresh$ **then**
            let $U \leftarrow \tilde{U}U$
        else pass

Preliminary tests - PSM-1

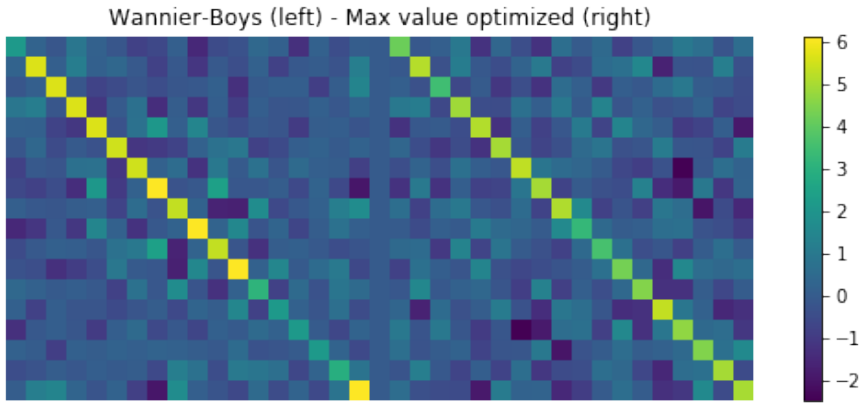


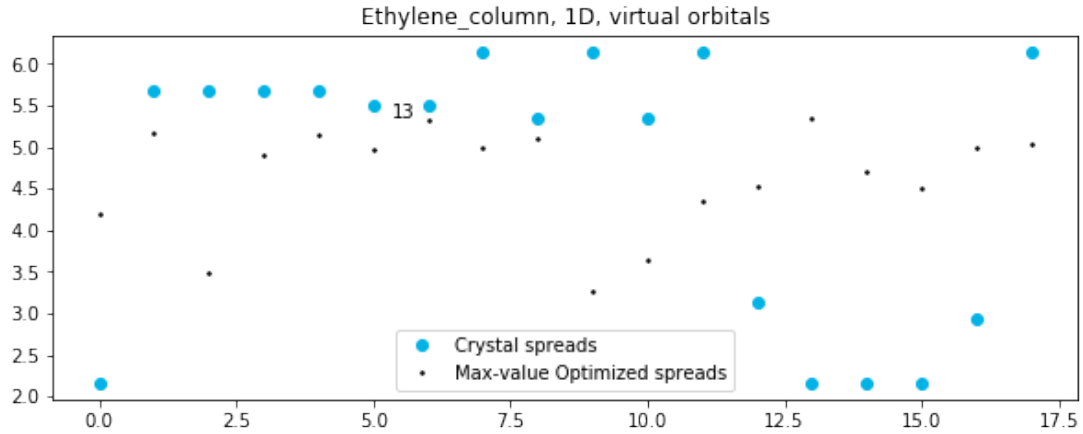


Preliminary tests - PSM-2

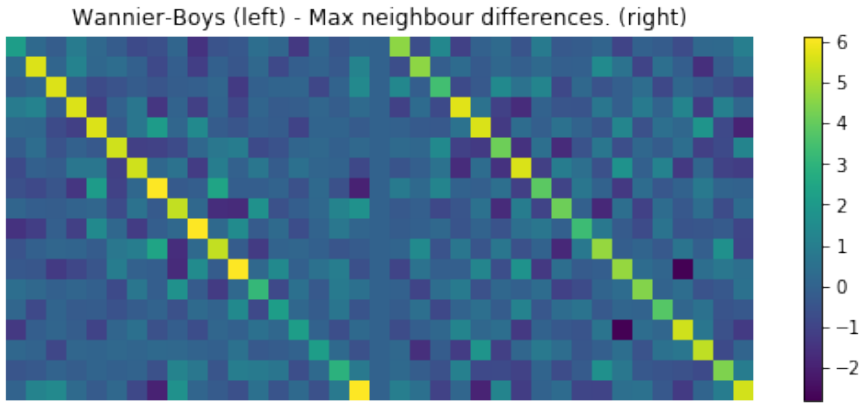


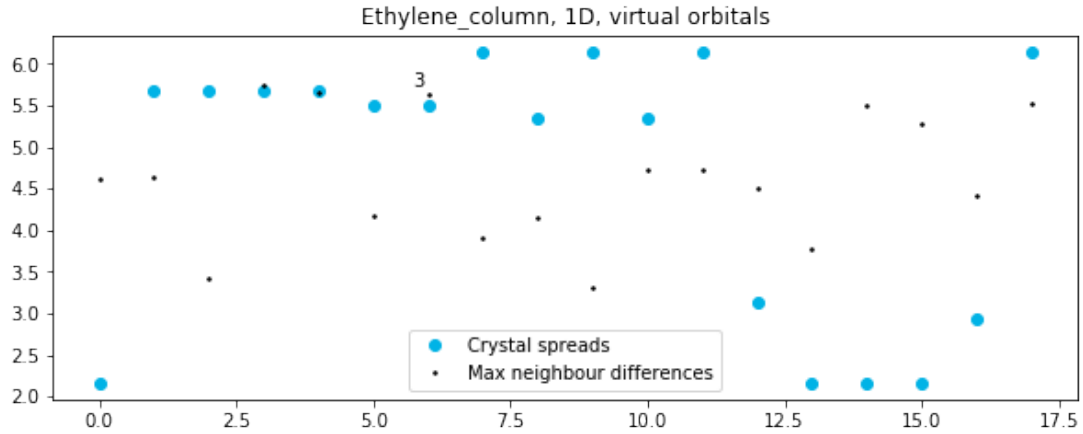
Preliminary tests - Maxval



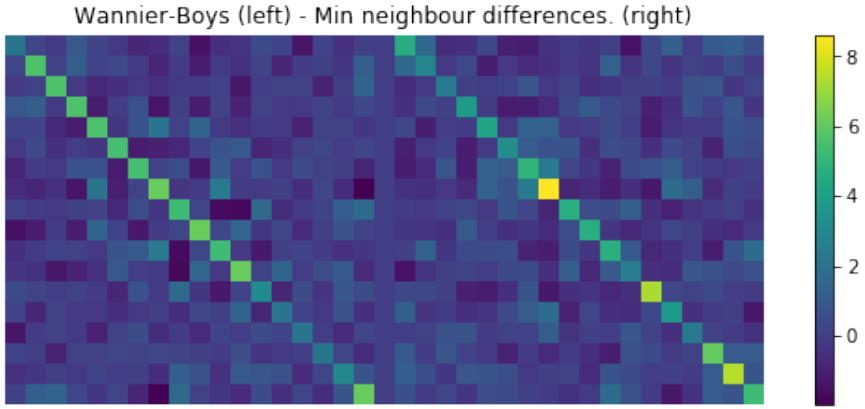


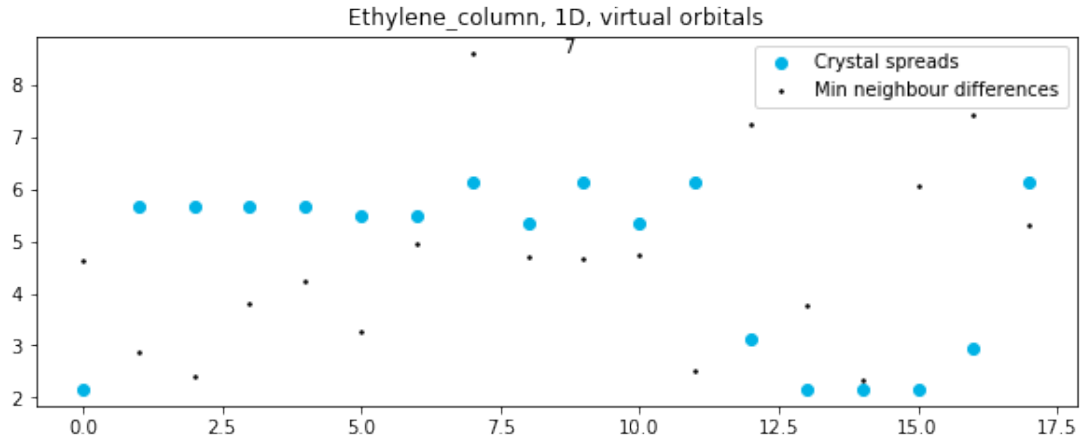
Preliminary tests - Max neighbour



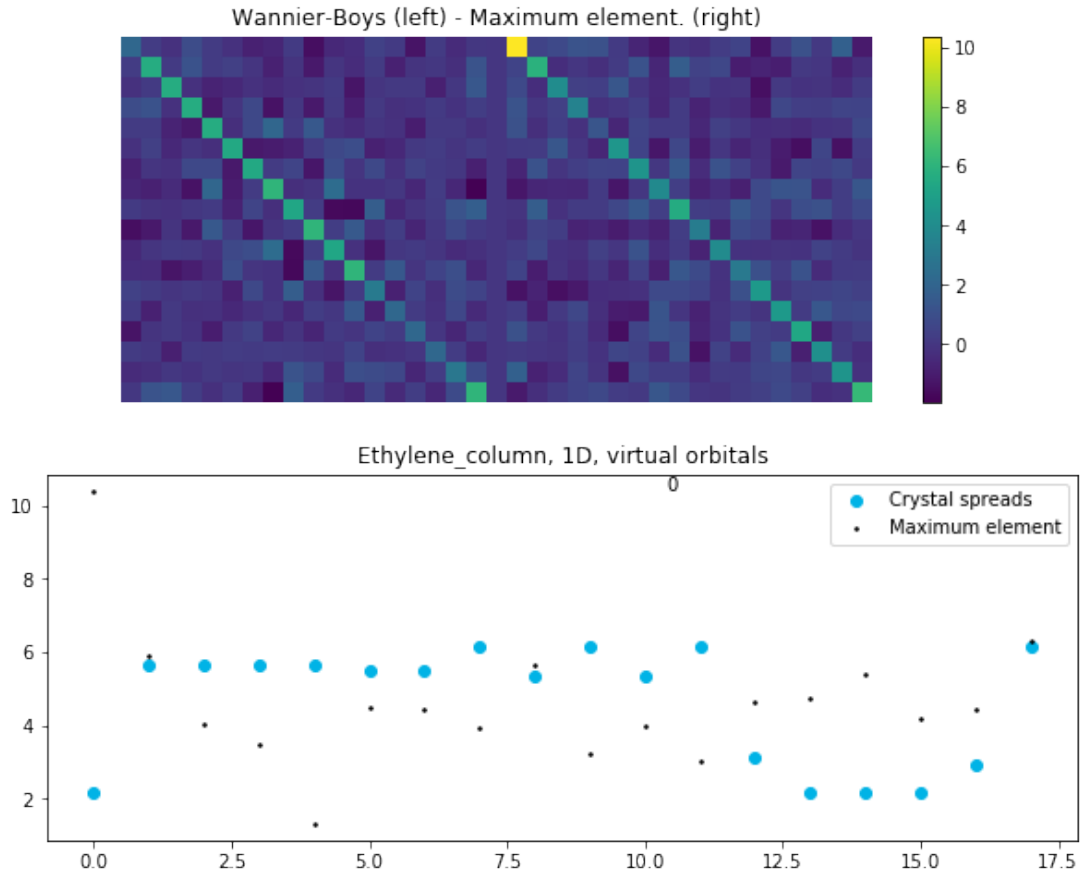


Preliminary tests - Min neighbour





Preliminary tests - Max element



Some notes on this approach:

- The procedure depends performance-wise on
  - The size of the matrix you want to transform.
  - The time spent on evaluating the objective function.
- From what we understand, this approach is likely to reach a global minimum.
- However, the algorithm is non-deterministic.
- The procedure is highly flexible with regards to objective functions, as it only requires the evaluation of the function at every step (no gradients).
- The procedure could potentially be used also to solve equations, for example by using the Hartree-Fock condition as objective function.

Our group

- Karl R. Leikanger (PhD student)
- Audun Skau Hansen (PhD student)
- Dr. Elisa Rebolini (Postdoc)
- Dr. Gustav Baardsen (Postdoc)
- Prof. Thomas Bondo Pedersen

Some collaborations

- Dr. Lorenzo Maschio from the Crystal[3] group at the University of Turin.
- Dr. Ida-Marie Høyvik at NTNU


S. F. Boys.
Construction of some molecular orbitals to be approximately invariant for changes from one molecule to another.
Rev. Mod. Phys., 32:296–299, Apr 1960.


Nicola Marzari, Arash A. Mostofi, Jonathan R. Yates, Ivo Souza, and David Vanderbilt.
Maximally localized wannier functions: Theory and applications.
Rev. Mod. Phys., 84:1419–1475, Oct 2012.


A. Erba C. M. Zicovich-Wilson B. Civalleri S. Casassa L. Maschio M. Ferrabone M. De La Pierre P. D?Arco Y. Noel M. Causa M. Rerat B. Kirtman. R. Dovesi, R. Orlando.
Crystal14.
Int. J. Quantum Chem., 114:1287, 2014.


Edmiston, Clyde and Ruedenberg, Klaus
Localized Atomic and Molecular Orbitals
Rev. Mod. Phys. 35


Claudio M. Zicovich-Wilson, Roberto Dovesi, and Victor R. Saunders.
A general method to obtain well localized wannier functions for composite energy bands in linear combination of atomic orbital periodic calculations.
The Journal of Chemical Physics, 115(21):9708–9719, 2001.