Hartree-Fock Stability

Applications to the Homogeneous Electron Gas And Implications in Electron Correlation

Evan Curtin

November 3, 2016

University of Illinois at Urbana-Champaign

Table of contents

Background Information

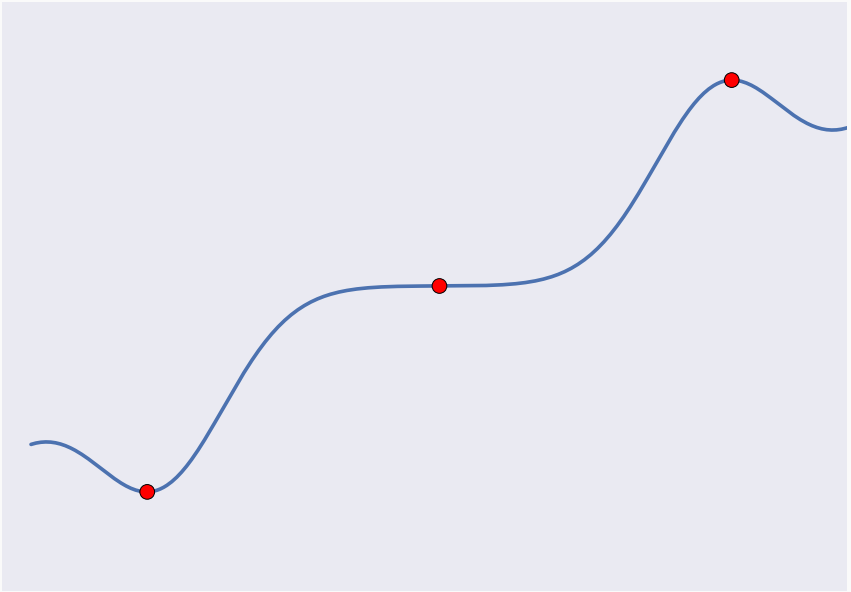
Levels of Hartree-Fock Theory

Method	Spinorbital	DoF	Eigenfunction of
Restricted	$\chi_j^{\alpha}(\vec{r},\sigma) = \sum_{i=1}^{N} c_{ij}\phi_i(\vec{r})\,\alpha(\sigma)$ $\chi_j^{\beta}(\vec{r},\sigma) = \sum_{i=1}^{N} c_{ij}\phi_i(\vec{r})\,\beta(\sigma)$	N/2	$\hat{S}^2$, $\hat{S}_z$
Unrestricted	$\chi_j^{\alpha}(\vec{r},\sigma) = \sum_{i=1}^{N} c_{ij}^{\alpha}\phi_i(\vec{r})\alpha(\sigma)$ $\chi_j^{\beta}(\vec{r},\sigma) = \sum_{i=1}^{N} c_{ij}^{\beta}\phi_i(\vec{r})\beta(\sigma)$	N	$\hat{S}_z$
General	$\chi_j(\vec{r},\sigma) = \sum_{i=1}^{N}[c_{ij}^{\alpha}\phi_i(\vec{r})\,\alpha(\sigma) + c_{ij}^{\beta}\phi_i(\vec{r})\beta(\sigma)]$	2N	Neither

Seeger, R.; Pople, J. A. J. Chem. Phys. 1977, 66 (7), 3045.

Restricted Minimization

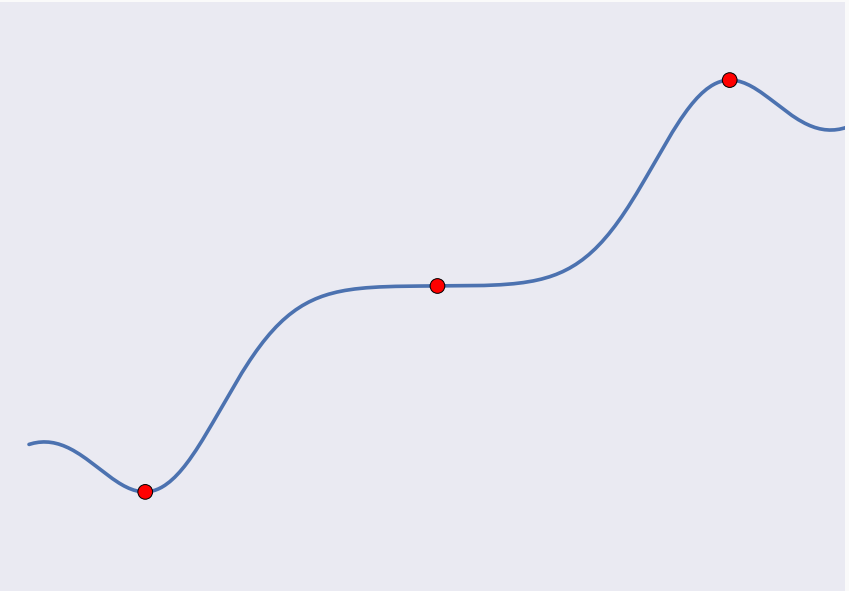
- Hartee-Fock SCF guarantees only stationary energy w.r.t. change in orbitals
- The solution may be a maximum, minimum or saddle point



Restricted Minimization

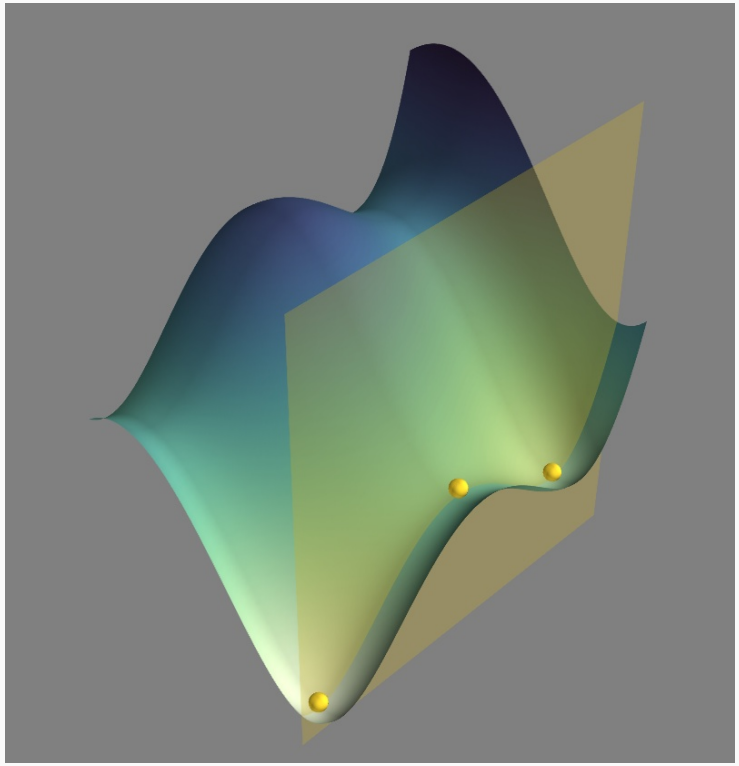
- Hartee-Fock SCF guarantees only stationary energy w.r.t. change in orbitals
- The solution may be a maximum, minimum or saddle point

Within the Constrained Space



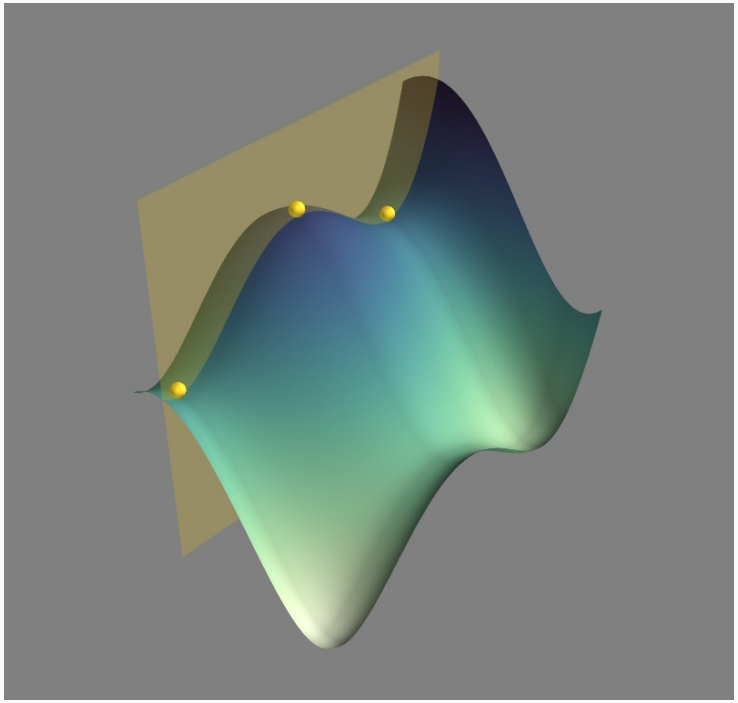
Restricted Minimization

- Restricted minima may correspond to absolute minima



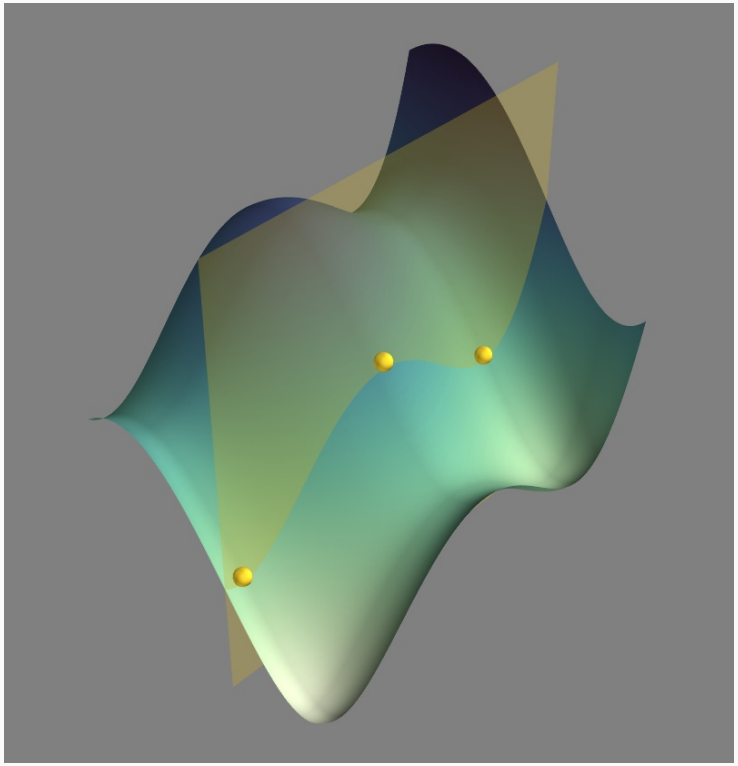
Restricted Minimization

- Restricted minima may correspond to absolute minima
- Restricted minima may correspond to absolute maxima



Restricted Minimization

- Restricted minima may correspond to absolute minima
- Restricted minima may correspond to absolute maxima
- Restricted minima may be nonstationary



Hartree-Fock Stability

Outline

- Inspired by biological evolution
- Evolve a population over generations
- Survival of the fittest
- Requirements:
  - Represent individuals as vector
  - Fitness function
- $v = \left(x_1 \ y_1 \ z_1 \ x_2 \ y_2 \ z_2 \ \dots \ x_N \ y_N \ z_N \right)$
- $F = \frac{E_{max} - E}{E_{max} - E_{min}}$

Thouless, D. J. Nucl. Phys. 1960, 21, 225232.

From Structure to Vector

- Several Ways to Define Structure
  - Cartesian
  - Internal Coordinates (bond length, angle ...) [1]
  - SMILES, InChI [2]

1. Supady, A.; Blum, V.; Baldauf, C. J. Chem. Inf. Model. 2015, 55 (11), 23382348.
2. http://www.chemspider.com/Chemical-Structure.2298795.html

From Structure to Vector

- Several Ways to Define Structure
  - Cartesian
  - Internal Coordinates (bond length, angle ...)
  - SMILES, InChI

1

- InChI[2]= 1S/C8H16/c1-5-7(3)8(4)6-2/h5-6H2,1-4H3/b8-7-

1. Supady, A.; Blum, V.; Baldauf, C. J. Chem. Inf. Model. 2015, 55 (11), 23382348.
2. http://www.chemspider.com/Chemical-Structure.2298795.html

From Structure to Vector

- Several Ways to Define Structure
  - Cartesian
  - Internal Coordinates (bond length, angle ...)
  - SMILES, InChI

1

- InChI[2]= 1S/C8H16/c1-5-7(3)8(4)6-2/h5-6H2,1-4H3/b8-7-
- Equivalent in theory

1. Supady, A.; Blum, V.; Baldauf, C. J. Chem. Inf. Model. 2015, 55 (11), 23382348.
2. http://www.chemspider.com/Chemical-Structure.2298795.html

From Structure to Vector

- Several Ways to Define Structure
  - Cartesian
  - Internal Coordinates (bond length, angle ...)
  - SMILES, InChI

1

- InChI[2]= 1S/C8H16/c1-5-7(3)8(4)6-2/h5-6H2,1-4H3/b8-7-
- Equivalent **in theory**

1. Supady, A.; Blum, V.; Baldauf, C. J. Chem. Inf. Model. 2015, 55 (11), 23382348.
2. http://www.chemspider.com/Chemical-Structure.2298795.html

Selecting Parents

- Several methods are common

Selecting Parents

- Several methods are common
- Reinforce good characteristics

Selecting Parents

- Several methods are common
- Reinforce good characteristics
- Still give losers a chance

Selecting Parents

- Several methods are common
- Reinforce good characteristics
- Still give losers a chance
- 'Breed' pairs of winners

Selecting Parents

- Several methods are common
- Reinforce good characteristics
- Still give losers a chance
- 'Breed' pairs of winners
- 'Roulette Wheel' Method

The Next Generation

Crossover distinguishes this from Monte Carlo

http://www.turingfinance.com/computational-finance-updates-ieee-world-congress-computational-intelligence

The Whole Algorithm

1. Generate N random, sensible geometries

The Whole Algorithm

1. Generate N random, sensible geometries
2. Add each to blacklist

The Whole Algorithm

1. Generate N random, sensible geometries
2. Add each to blacklist
3. Optimize each geometry

The Whole Algorithm

1. Generate N random, sensible geometries
2. Add each to blacklist
3. Optimize each geometry
4. Select Parents

The Whole Algorithm

1. Generate N random, sensible geometries
2. Add each to blacklist
3. Optimize each geometry
4. Select Parents
5. Crossover & Mutate into sensible geometry

The Whole Algorithm

1. Generate N random, sensible geometries
2. Add each to blacklist
3. Optimize each geometry
4. Select Parents
5. Crossover & Mutate into sensible geometry
6. Add Children to population

The Whole Algorithm

1. Generate N random, sensible geometries
2. Add each to blacklist
3. Optimize each geometry
4. Select Parents
5. Crossover & Mutate into sensible geometry
6. Add Children to population
7. Remove High energy individuals

The Whole Algorithm

1. Generate N random, sensible geometries
2. Add each to blacklist
3. Optimize each geometry
4. Select Parents
5. Crossover & Mutate into sensible geometry
6. Add Children to population
7. Remove High energy individuals
8. If converged:

The Whole Algorithm

1. Generate N random, sensible geometries
2. Add each to blacklist
3. Optimize each geometry
4. Select Parents
5. Crossover & Mutate into sensible geometry
6. Add Children to population
7. Remove High energy individuals
8. If converged:
    - Done!

The Whole Algorithm

1. Generate N random, sensible geometries
2. Add each to blacklist
3. Optimize each geometry
4. Select Parents
5. Crossover & Mutate into sensible geometry
6. Add Children to population
7. Remove High energy individuals
8. If converged:
   - Done!

   Otherwise:

The Whole Algorithm

1. Generate N random, sensible geometries
2. Add each to blacklist
3. Optimize each geometry
4. Select Parents
5. Crossover & Mutate into sensible geometry
6. Add Children to population
7. Remove High energy individuals
8. If converged:
   - Done!

   Otherwise:
   - Go to 2

Finding Low Energy Conformers of Dipeptides

Dipeptide Structures

Combinatorics

- GA beats other methods
  if space is large
- Space gets large **fast**

Coverage

- Smaller systems are reliably sampled
- As # of conformers increases, miss more and more
- Is there a pattern to what is missed?

Coverage

- Most misses are very high energy

Coverage

- Most misses are very high energy
- Algorithm favors low energy areas of the space

Coverage

- Most misses are very high energy
- Algorithm favors low energy areas of the space
- Features low in energy are favored and recombined

Energy Cutoff

Energy Cutoff

- GA is more sensitive to energy cutoff

Energy Cutoff

- GA is more sensitive to energy cutoff
- For finding low energy ensemble, GA outperforms purely stochastic/deterministic method

Concluding Remarks

Concluding Remarks

- Finding all the low energy conformers for a molecule is hard

Concluding Remarks

- Finding all the low energy conformers for a molecule is hard
- In order to get accurate energies/structures, computationally expensive methods should be employed

Concluding Remarks

- Finding all the low energy conformers for a molecule is hard
- In order to get accurate energies/structures, computationally expensive methods should be employed
- These take time, so we want to minimize how many of these we do

Concluding Remarks

- Finding all the low energy conformers for a molecule is hard
- In order to get accurate energies/structures, computationally expensive methods should be employed
- These take time, so we want to minimize how many of these we do
- The Genetic Algorithm provides a framework for a refined global search

Concluding Remarks

- Finding all the low energy conformers for a molecule is hard
- In order to get accurate energies/structures, computationally expensive methods should be employed
- These take time, so we want to minimize how many of these we do
- The Genetic Algorithm provides a framework for a refined global search
- It shines when asked to find a host of low energy solutions

Concluding Remarks

- Finding all the low energy conformers for a molecule is hard
- In order to get accurate energies/structures, computationally expensive methods should be employed
- These take time, so we want to minimize how many of these we do
- The Genetic Algorithm provides a framework for a refined global search
- It shines when asked to find a host of low energy solutions
- GA wrapper can be interfaced with a variety of electronic structure packages(NWChem, ORCA) and is available under the GNU Lesser General Public License at `https://github.com/adrianasupady/fafoom`

Questions?

Backup slide

- Geometry optimization step makes the algorithm more Lamarckian (Jean Baptiste Larmarck, [1744-1829])

Genetic Algorithm Parameters

Geometry Optimization: DFT PBE + VdW, *tier1* basis in FHI-aims[1].
Convergence at 0.005 eV / Å

(1) Blum, V. et. al., M. Comput. Phys. Commun. 2009, 180 (11), 21752196.
(2) Supady, A.; Blum, V.; Baldauf, C. J. Chem. Inf. Model. 2015, 55 (11), 23382348.