Ab initio methods: Post-Hartree–Fock

Alston J. Misquitta

Centre for Condensed Matter and Materials Physics
Queen Mary, University of London

March 6, 2012

Hartree–Fock equations I

We have introduced basis sets in solving the Hartree–Fock equations, but what exactly are they?

Brief summary of Hartree–Fock theory:

- Born–Oppenheimer approximation: fixed nuclei.
- Ground-state N-electron wavefunction: single Slater determinant:

$$\Psi_0(x_1, x_2, \cdots, x_N) = |\chi_1 \chi_2 \cdots \chi_N\rangle \tag{1}$$

- Variational Principle:

$$E_0 \leq E^{\mathrm{HF}} = \min \langle \Psi_0 | \mathcal{H} | \Psi_0 \rangle \tag{2}$$

subject to the conditions that the spin-orbitals are orthonormal.

Hartree–Fock equations II

- Fock equations:

$$f(i)\chi(x_i) = \epsilon\chi(x_i) \tag{3}$$

  where $f(i)$ is an effective operator called the *Fock operator*

$$f(i) = -\frac{1}{2}\nabla_i^2 - \sum_\alpha \frac{Z_\alpha}{r_{i\alpha}} + v^{\mathrm{HF}}(i) \tag{4}$$

  where $v^{\mathrm{HF}}(i)$ is the Hartree–Fock *effective* potential.

- Self-consistent solution (iterations needed).
- Infinity of solutions to the Fock equations.
- Introduce a basis for the spatial part of spin-orbitals:

$$\chi_i(r) = \sum_m C_{im}\phi_m(r) \tag{5}$$

Hartree–Fock equations III

- Leads to linear equations:

$$\mathbf{FC} = \epsilon\mathbf{SC} \qquad (6)$$

Q: How do we choose the basis?

Slater-type orbitals I

A reasonable choice for basis sets for finite systems would be what are called Slater-type orbitals: these are very like solutions of the 1-electron Hamiltonian. They differ in two ways: (1) the radial part is simpler and (2) the exponent is not integral but can be varied to account for *screening* effects.

$$\phi = R_{nl}(r)\mathcal{Y}_{lm}(\theta, \phi) \tag{7}$$

where $\mathcal{Y}_{lm}$ is a (real) spherical harmonic and the radial part is given by

$$R_{nl}(r) = \frac{(2\zeta)^{n+1/2}}{[(2n)!]^{1/2}} r^{n-1} e^{-\zeta r} \tag{8}$$

Slater-type orbitals II
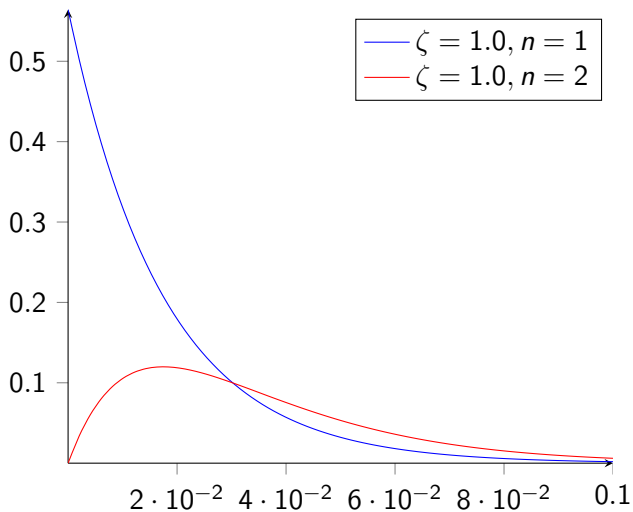
Slater-type orbitals III

Comments on Slater-type orbitals:

- GOOD Nuclear cusp condition satisfied.

$$\frac{\partial}{\partial r}\langle\rho(r)\rangle_{\mathrm{sph}}\bigg|_{r=0} = -2Z\langle\rho(0)\rangle_{\mathrm{sph}}$$

- GOOD Exact wavefunction has the long-range form of a Slater orbital.
  If we pull one electron out of an N-electron molecule the wavefunction behaves like

$$\Psi(N) \rightarrow \Psi(N-1) \times e^{-\sqrt{2*I}r}$$

  where I is the first (vertical) ionization energy.

- BAD Integrals very difficult for multi-atom systems.

Gaussian-type orbitals (GTOs) I

In 1950 S. F. Boys pointed out that the problem of computing integrals could be resolved by using not Slater-type orbitals, but rather Gaussian-type orbitals (GTOs):
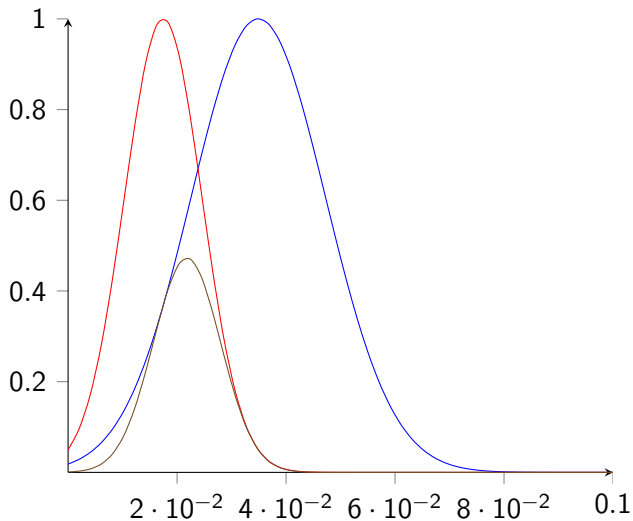
$$R_{nl} \sim r^l e^{-\alpha(r-\mathbf{A})^2} \tag{9}$$

where $\mathbf{A}$ is the centre of the GTO. The main reason for the efficacy of GTOs is that the product of two GTOs is a third GTO, centred at a point in between:

$$\exp(-\alpha(r-\mathbf{A})^2)\exp(-\beta(r-\mathbf{B})^2) = \exp(-\gamma(\mathbf{A}-\mathbf{B})^2)\exp(-\mu(r-\mathbf{P})^2)$$

where $\mu = \alpha + \beta$, $\gamma = \alpha\beta/\mu$ and $\mathbf{P} = (\alpha\mathbf{A} + \beta\mathbf{B})/\mu$.
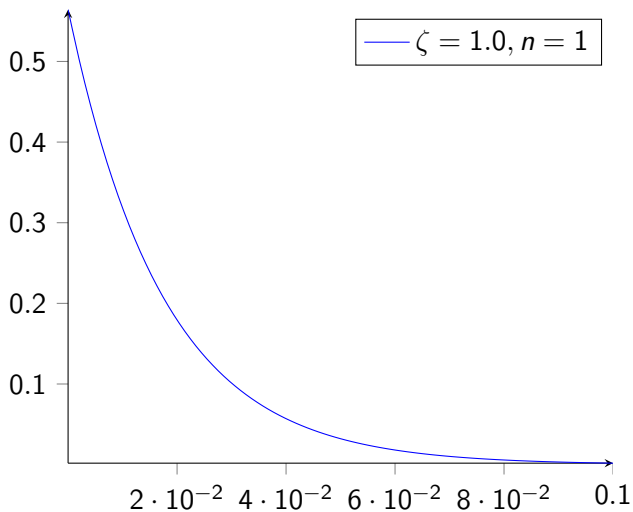
Gaussian-type orbitals (GTOs) II

Gaussian-type orbitals (GTOs) III

- GOOD GTOs makes the integrals that appear in the HF energy expression much simpler.
- BAD Nuclear cusp condition violated: zero derivative at origin.
- BAD Wrong long-range form: dies off too fast with distance.

The two negative points can, to some extent, be remedied by using not single GTOs, but *linear combinations* of GTOs. These groups of GTOs are termed *contractions*.

Basis sets consist of groups of contractions together with some un-contracted GTOs. The better the basis, the more of these there will be and the more GTOs in a contraction.
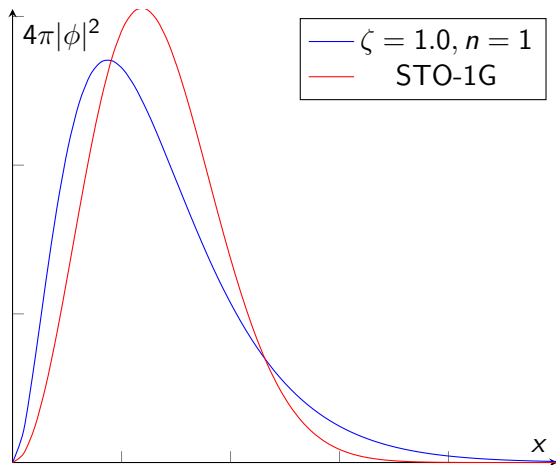
Gaussian-type orbitals (GTOs) IV

Gaussian-type orbitals (GTOs) V
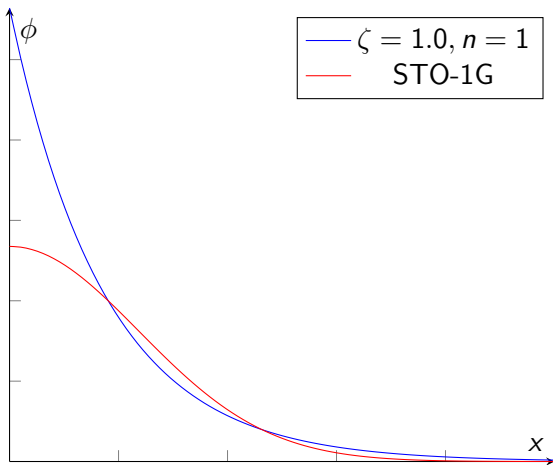
Gaussian-type orbitals (GTOs) VI

Gaussian-type orbitals (GTOs) VII

Gaussian-type orbitals (GTOs) VIII

Plot legend: $\zeta = 1.0, n = 1$ (blue); STO-3G (red)
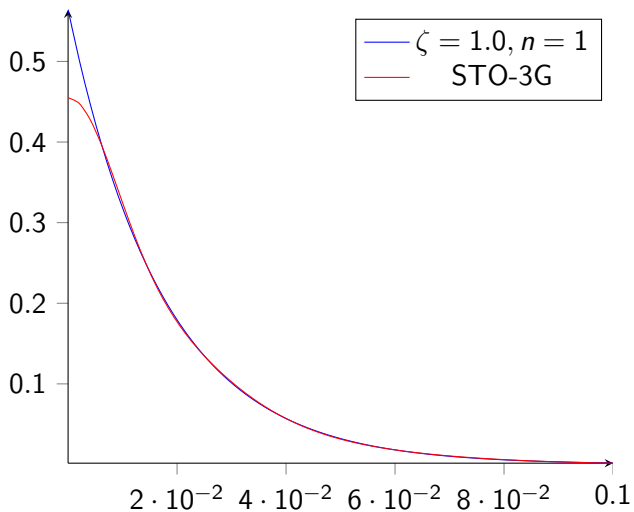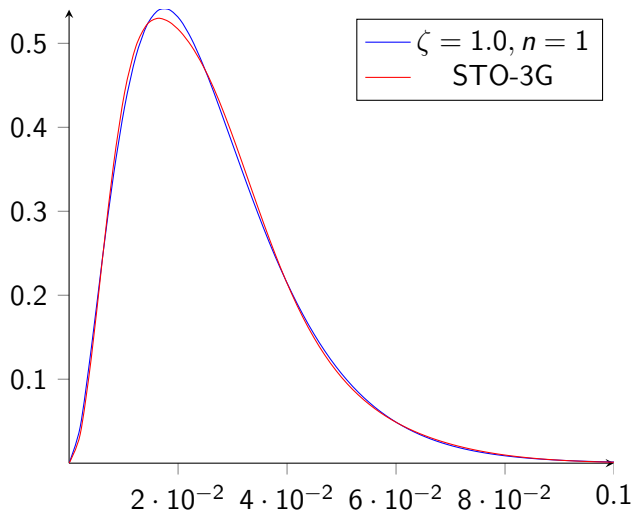
Gaussian-type orbitals (GTOs) IX

Basis set recommendations:

- GOOD Complete basis set (CBS) limit
- Geometry optimization: moderate size basis sets. Double-ζ.
- Energies: At least triple-ζ quality.
- Properties: Triple-ζ or more.

We will have another look at basis sets after discussion correlated methods.

Correlation I

Q: How to we improve on the complete basis set HF results? I.e.,
how do we get beyond the HF limit?

- Variational Principle: More flexibility. Leads to Configuration
  Interaction (CI).
- Perturbation Theory
- Coupled-cluster methods
- Density functional theory

Configuration Interaction I

CI: Increase the flexibility in the wavefunction by including in addition the the HF ground state, excited states.
Q: What are excited states and how to we form them?

Configuration Interaction II

$$
\begin{array}{cc}
n+m \text{ ------} & \text{------} \\
\dotsc & \dotsc \\
\dotsc & \dotsc \\
n+1 \text{ ------} & \text{---}\downarrow\text{--} \\
n \text{ ---}\uparrow\downarrow\text{---} & \text{---}\uparrow\downarrow\text{---} \\
\dotsc & \dotsc \\
2 \text{ ---}\uparrow\downarrow\text{---} & \text{---}\uparrow\text{---} \\
1 \text{ ---}\uparrow\downarrow\text{---} & \text{---}\uparrow\downarrow\text{---}
\end{array}
$$

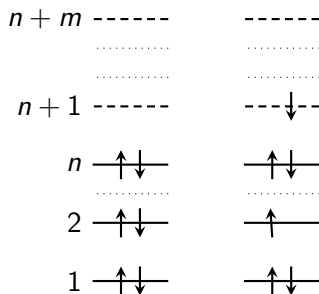Figure: Left: HF ground state configuration. Right: An example of an excited state configuration. If there are n occupied levels (2 electrons each, so $N = 2n$) and m virtual (un-occupied) levels, in how many ways can we form excited states? Each of these states will correspond to a Slater determinant.

Configuration Interaction III

The Full CI (FCI) wavefunction:

$$|\Psi\rangle = |\Psi_0\rangle + \sum_{ar} c_a^r |\Psi_a^r\rangle + \sum_{abrs} c_{ab}^{rs} |\Psi_{ab}^{rs}\rangle + \cdots$$

$$= |\Psi_0\rangle + c_S |S\rangle + c_D |D\rangle + \cdots$$

where electrons are excited from the *occupied* orbitals $a, b, c, \cdots$ to the *virtual* orbitals $r, s, t, \cdots$.

- GOOD This expansion will lead to the *exact* energy *within the basis set used*.
- BAD There are too many determinants!

$$\frac{(2(n+m))!}{(2n)!(2m)!}$$

Configuration Interaction IV

One solution to the problem is to use only some of the many
determinants. For example we could use only double excitations.
This leads to the CID method.

$$\left|\Psi^{\mathrm{CID}}\right\rangle = \left|\Psi_0\right\rangle + \sum_{abrs} c_{ab}^{rs} \left|\Psi_{ab}^{rs}\right\rangle$$

$$= \left|\Psi_0\right\rangle + c_D \left|D\right\rangle$$

BAD This theory, like all truncated CI methods, is not *size
extensive*.
Size-extensivity: If $E(N)$ is the energy of N *non-interacting*
identical systems then a method is size-extensive if
$E(N) = N \times E(1)$.

Configuration Interaction V

Q: Is CID size-extensive?

If $\hat{T}_2$ is an operator that creates all double excitations, then we can write the CID wavefunction as

$$
\begin{aligned}
\left|\Psi^{\mathrm{CID}}\right\rangle &= |\Psi_0\rangle + \sum_{abrs} c_{ab}^{rs} |\Psi_{ab}^{rs}\rangle \\
&= |\Psi_0\rangle + c_D |D\rangle \\
&= (1 + \hat{T}_2) |\Psi_0\rangle
\end{aligned}
$$

Configuration Interaction VI

The CID wavefunction for each of two identical non-interacting systems will be of that form, so the combined wavefunction will be

$$\left| \Psi_A^{\mathrm{CID}} \right\rangle \left| \Psi_B^{\mathrm{CID}} \right\rangle = (1 + \hat{T}_2(A)) \left| \Psi_0^A \right\rangle (1 + \hat{T}_2(B)) \left| \Psi_0^B \right\rangle$$

$$= (1 + \hat{T}_2(A) + \hat{T}_2(B) + \hat{T}_2(A)\hat{T}_2(B)) \left| \Psi_0^A \right\rangle \left| \Psi_0^B \right\rangle$$

The last excitation term is a quadruple excitation so it will not be present in the CID wavefunction for the combined A and B systems. Therefore

$$E^{\mathrm{CID}}(AB) \neq E^{\mathrm{CID}}(A) + E^{\mathrm{CID}}(B).$$

Configuration Interaction VII
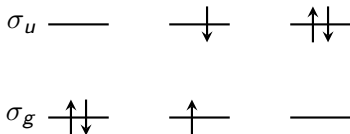
CI is not size-extensive: H_2

Figure: An energy level diagram for H_2 with a minimal basis. The two MOs are as shown. We can form three determinants from them. Left is the ground state. Middle is a *singly excited* determinant. Right is a *doubly excited* determinant. For reasons of symmetry the middle configuration does not contribute to the CI expansion.

Coupled-cluster Theory I

The problem of truncated CI methods is severe enough that using them is very problematic. A resolution to the problem is the class of coupled-cluster theories. In these the wavefunction is defined as:

$$\left|\Psi^{\mathrm{CC}}\right\rangle = \exp(\hat{T})\left|\Psi_0\right\rangle$$

where $\hat{T}$ is an appropriate excitation operator.

Coupled-cluster Theory II

For example, in CCD theory we use $\hat{T}_2$ as the excitation operator.
This gives:

$$\begin{aligned}
\left|\Psi^{\mathrm{CCD}}\right\rangle &= \exp(\hat{T}_2)\left|\Psi_0\right\rangle \\
&= (1 + \hat{T}_2 + \frac{1}{2!}\hat{T}_2\hat{T}_2 + \cdots)\left|\Psi_0\right\rangle
\end{aligned}$$

The first two terms give us CID theory. The rest are needed to
make CCD size-extensive:

$$\begin{aligned}
\exp(\hat{T}_2(A))\left|\Psi_0^A\right\rangle \times \exp(\hat{T}_2(B))\left|\Psi_0^B\right\rangle &= \exp(\hat{T}_2(A) + \hat{T}_2(B))\left|\Psi_0^A\right\rangle\left|\Psi_0^B\right\rangle \\
&\equiv \exp(\hat{T}_2(AB))\left|\Psi_0^A\right\rangle\left|\Psi_0^B\right\rangle
\end{aligned}$$

Coupled-cluster Theory III

- GOOD CC theories can be systematically improved.
- GOOD CCSD(T) is a very accurate theory. Here single and double excitations are included as described above and triple excitations are included through a perturbative treatment.
- GOOD Size-extensive.
- BAD Computationally very expensive: CCSD(T) scales as $\mathcal{O}(N^7)$. So double the system size and the calculation costs 128 times more.
- BAD (kind of!) These are single-determinant theories as described. If the system is multi-configurational (more than one state contributing dominantly) the standard CC methods are not appropriate.

Møller–Plesset Perturbation Theory I

Brief recap of Raleigh–Schrödinger perturbation theory:

- Split the Hamiltonian into two parts:

$$\mathcal{H} = \mathcal{H}_0 + \lambda \mathcal{V}$$

where $\mathcal{H}_0$ is a Hamiltonian which we know how to solve and $\mathcal{V}$ contains that troublesome parts. We expect $\mathcal{V}$ to be a perturbation so it must be small in some sense.

λ is a complex number that will be 1 for the physical solution.

Møller–Plesset Perturbation Theory II

- Let the solutions of $\mathcal{H}_0$ be:

$$\mathcal{H}_0 \Psi_i^{(0)} = E_i^{(0)} \Psi_i^{(0)}$$

  Here the '0' indicates that these eigenvalues and eigenfunctions are of *zeroth*-order in the perturbation $\mathcal{V}$. We will use the short-form:

$$|\Psi_i^{(0)}\rangle \equiv |i\rangle$$

- Express the solutions of $\mathcal{H}$ in a power-series:

$$\Psi_i = \Psi_i^{(0)} + \lambda \Psi_i^{(1)} + \lambda^2 \Psi_i^{(2)} + \cdots = \sum_n \lambda^n \Psi_i^{(n)}$$

$$E_i = E_i^{(0)} + \lambda E_i^{(1)} + \lambda^2 E_i^{(2)} + \cdots = \sum_n \lambda^n E_i^{(n)}$$

Møller–Plesset Perturbation Theory III

- Energies can be calculated by collecting terms at various orders:

$$E_i^{(0)} = \langle \Psi_i^{(0)} | \mathcal{H}_0 | \Psi_i^{(0)} \rangle = \langle i | \mathcal{H}_0 | i \rangle$$
$$E_i^{(1)} = \langle \Psi_i^{(0)} | \mathcal{V} | \Psi_i^{(0)} \rangle = \langle i | \mathcal{V} | i \rangle$$
$$E_i^{(2)} = \langle \Psi_i^{(0)} | \mathcal{V} | \Psi_i^{(1)} \rangle$$

etc. for higher order terms.

- The first-order correction to the wavefunction is given by

$$|\Psi_i^{(1)}\rangle = \sum_{n \neq 0} \frac{|n\rangle \langle n | \mathcal{V} | i \rangle}{E_i^{(0)} - E_n^{(0)}}$$

Møller–Plesset Perturbation Theory IV

- So we get

$$E_i^{(2)} = \sum_{n \neq 0} \frac{|\langle n|\mathcal{V}|i\rangle|^2}{E_i^{(0)} - E_n^{(0)}}$$

Møller–Plesset Perturbation Theory V

Many-body perturbation theory (MBPT) starts from Hartree–Fock theory:

$$\mathcal{H}_0 = \sum_{i=1}^{N} f(i) = \sum_{i=1}^{n} \left(h(i) + v^{\mathrm{HF}}(i) \right) \tag{10}$$

where $h(i) = -\frac{1}{2}\nabla_i^2 - \sum_\alpha \frac{Z_\alpha}{r_{i\alpha}}$

We can now define the perturbation as

$$\mathcal{V} = \sum_{i=1}^{N} \sum_{j>i}^{N} \frac{1}{r_{ij}} - \sum_{i=1}^{N} v^{\mathrm{HF}}(i) \tag{11}$$

Unlike v^{HF}, the perturbation $\mathcal{V}$ is a 2-electron operator.

Møller–Plesset Perturbation Theory VI

MBPT energy at λ^0:

$$E_0^{(0)} = \langle 0|\mathcal{F}|0\rangle = \sum_{a \in \text{occ}} \epsilon_a$$

At first-order we get (no proof):

$$E_0^{(1)} = -\frac{1}{2} \sum_{a=1,b=1}^{N} \left[\langle ab|r_{12}^{-1}|ab\rangle - \langle aa|r_{12}^{-1}|bb\rangle \right]$$

We have not seen this before, but the sum of $E_0^{(0)}$ and $E_0^{(1)}$ is just the Hartree–Fock ground state energy:

$$E^{\text{HF}} = E_0^{(0)} + E_0^{(1)} \tag{12}$$

This means that we need to get to at least second-order in perturbation theory to go beyond the Hartree–Fock description.

Møller–Plesset Perturbation Theory VII

Here is what the second-order MBPT energy expression looks like:

$$E_0^{(2)} = \sum_{a,b>a}^{\text{occ}} \sum_{r,s>r}^{\text{vir}} \frac{\left[\langle ab|r_{12}^{-1}|rs\rangle - \langle ab|r_{12}^{-1}|sr\rangle\right]^2}{\epsilon_r + \epsilon_s - \epsilon_a - \epsilon_b} \tag{13}$$

This expression is termed as MBPT2 or MP2. The latter name comes from the other name for this kind of perturbation theory: Møller–Plesset perturbation theory.

Møller–Plesset Perturbation Theory VIII

- **BAD** A problem with Møller–Plesset perturbation theory: it diverges! See Olsen *et al.* J. Chem. Phys. **112**, 9736 (2000) for details. We now rarely go beyond MP2 in practical calculations.

- **GOOD** MP2 contains correlation.

- **BAD** But not enough correlation. Problems with systems with small HOMO-LUMO gaps (band gap — HOMO is highest occupied MO and LUMO is lowest unoccupied MO).

- **GOOD** (kind of!) It has a computational cost of $\mathcal{O}(N^5)$. I.e., double the system in size and it will cost 32 times more computational power.

- **GOOD** MBPT is *size-consistent*

Expectation Values and Forces I

Once we've got the HF solution, how do we calculate expectation values of operators and forces?