Lecture 8

Gaussian basis sets

C.-K. Skylaris
Solving the Kohn-Sham equations on a computer

• The SCF procedure involves solving the Kohn-Sham single-electron equations for the molecular orbitals

\[
\left( -\frac{1}{2} \nabla^2 + V_{KS}(r) \right) \psi_j(r) = \varepsilon_j \psi_j(r)
\]

• Where the Kohn-Sham potential of the non-interacting electrons is given by

\[
V_{KS}(r) = V_{\text{ext}}(r) + V_{\text{Coul}}(r) + V_{xc}(r)
\]

\[
= V_{\text{ext}}(r) + \int \frac{n(r')}{|r - r'|} \, dr' + \frac{\delta E_{xc}}{\delta n}
\]

• We all have some experience in solving equations on paper but how we do this with a computer?
Linear Combination of Atomic Orbitals (LCAO)

- We will **express the MOs** as a linear combination of atomic orbitals (LCAO).

- The strength of the LCAO approach is its general applicability: it can work on any molecule with any number of atoms.

Example:

$$\psi(\mathbf{r}) = c_1\chi_1(\mathbf{r}) + c_2\chi_2(\mathbf{r}) + c_3\chi_3(\mathbf{r}) + c_4\chi_4(\mathbf{r}) + c_5\chi_5(\mathbf{r}) + c_6\chi_6(\mathbf{r}) + c_7\chi_7(\mathbf{r}) + \ldots$$

AOs on atom A  AOs on atom B  AOs on atom C
Example: find the AOs from which the MOs of the following molecules will be built
Basis functions

• We can take the LCAO concept one step further:

  • Use a larger number of AOs (e.g. a hydrogen atom can have more than one s AO, and some p and d AOs, etc.). This will achieve a more flexible representation of our MOs and therefore more accurate calculated properties according to the variation principle

  • Use AOs of a particular mathematical form that simplifies the computations (but are not necessarily equal to the exact AOs of the isolated atoms)

• We call such sets of more general AOs **basis functions**

• Instead of having to calculate the mathematical form of the MOs (impossible on a computer) the problem is reduced to determining the MO expansion coefficients in terms of the basis functions

\[
\psi_i (\mathbf{r}) = \sum_{\alpha=1}^{N_{BF}} G_\alpha (\mathbf{r}) C_{\alpha i}
\]

- molecular (spatial) orbital
- basis function (fixed)
- coefficient (a number, to be determined by the SCF calculation)
Gaussian basis functions

Functions that resemble hydrogen AOs (Slater functions) are very suitable for expanding MOs because they have the correct shape:

- Near the nucleus (cusp)
- Far from the nucleus (decay like $e^{-ar}$)

However, Gaussian functions are preferred in practice because they allow for efficient computation of molecular integrals (simpler formulas)

\[
s_{\nu}(\mathbf{r}) = e^{-\zeta_{\nu}|\mathbf{r}|} = e^{-\zeta_{\nu}\sqrt{x^2 + y^2 + z^2}}
\]

\[
g_{\nu}(\mathbf{r}) = e^{-\zeta_{\nu}r^2} = e^{-\zeta_{\nu}(x^2 + y^2 + z^2)}
\]
Primitive Gaussian functions

• In the jargon of quantum chemistry a single Gaussian function is called a **primitive Gaussian function**, or **primitive GTO** (Gaussian Type Orbital)

• Some programs use Cartesian primitive GTOs:

\[ g_\nu(r) = x^k y^m z^n e^{-\zeta \nu r^2} = x^k y^m z^n e^{-\zeta \nu (x^2+y^2+z^2)} \]

• Other programs use Spherical primitive GTOs

\[ g_\nu(r) = Y_{l,m}(\theta, \phi) r^l e^{-\zeta \nu r^2} = Y_{l,m}(\theta, \phi) r^l e^{-\zeta \nu (x^2+y^2+z^2)} \]

• Spherical and Cartesian functions are the same for up to \( l=1 \) (p functions) but differ slightly for \( l=2 \) or higher
Cartesian primitive Gaussian functions

Similar to atomic orbitals, we define **Cartesian Gaussian** atomic functions by an angular momentum "quantum number" $l$ and the numbers $k$, $m$ and $n$ where $l = k + m + n$.

s functions

\[ l = 0 \quad \left\{ \begin{array}{c} e^{-\zeta r^2} \end{array} \right. \]

p functions

\[ l = 1 \quad \left\{ \begin{array}{c} xe^{-\zeta r^2} \\ ye^{-\zeta r^2} \\ ze^{-\zeta r^2} \end{array} \right. \]

d functions

\[ l = 2 \quad \left\{ \begin{array}{c} x^2 e^{-\zeta r^2} \\ y^2 e^{-\zeta r^2} \\ z^2 e^{-\zeta r^2} \\ xy e^{-\zeta r^2} \\ xz e^{-\zeta r^2} \\ yz e^{-\zeta r^2} \end{array} \right. \]

etc.
Spherical primitive Gaussian functions

Similar to atomic orbitals, we define **Spherical Gaussian** atomic functions by an angular momentum “quantum number” \( l \) and its components \( m=-l,...,l \)

s functions

\[ l = 0 \quad \left\{ \begin{array}{l}
Y_{1,-1}(\theta, \phi) \, r \, e^{-\zeta r^2}
\end{array} \right. \]

p functions

\[ l = 1 \quad \left\{ \begin{array}{l}
Y_{1,-1}(\theta, \phi) \, r \, e^{-\zeta r^2} \\
Y_{1,0}(\theta, \phi) \, r \, e^{-\zeta r^2} \\
Y_{1,1}(\theta, \phi) \, r \, e^{-\zeta r^2}
\end{array} \right. \]

d functions

\[ l = 2 \quad \left\{ \begin{array}{l}
Y_{2,-2}(\theta, \phi) \, r^2 \, e^{-\zeta r^2} \\
Y_{2,-1}(\theta, \phi) \, r^2 \, e^{-\zeta r^2} \\
Y_{2,0}(\theta, \phi) \, r^2 \, e^{-\zeta r^2} \\
Y_{2,1}(\theta, \phi) \, r^2 \, e^{-\zeta r^2} \\
Y_{2,2}(\theta, \phi) \, r^2 \, e^{-\zeta r^2}
\end{array} \right. \quad \text{etc.} \]
In practice, fixed linear combinations of “prIMITIVE” Gaussian functions are used.

These are called “Contracted Gaussians” (CGs):

\[ g_{\alpha}(r) = \sum_{\nu=1}^{N_{\alpha}} c_{\nu} g_{\nu}(r) \]

The simplest kind of CGs are the STO-nG basis sets.

These basis sets attempt to approximate Slater-type orbitals (STOs) by n primitive Gaussians.

E.g. STO-nG functions for the 1s orbital of a hydrogen atom:

- **STO-1G**: \( e^{-0.27r^2} \)
- **STO-2G**: \( 0.68e^{-0.15r^2} + 0.43e^{-0.86r^2} \)
- **STO-3G**: \( 0.44e^{-0.11r^2} + 0.53e^{-0.41r^2} + 0.15e^{-2.23r^2} \)
Gaussian basis sets

The STO-nG basis sets are rather unsatisfactory as they include only one contracted Gaussian (CG) per atomic orbital. Improved basis sets are obtained by including:

• More than one CG per atomic orbital, e.g.: **DZ** (“double zeta”), **TZ** (“triple zeta”), **QZ** (“quadruple zeta”)

• One CG per “core” atomic orbital and more than one for the valence atomic orbitals, e.g.: **SV**, **3-21G**, **4-31G**, **6-31G**, **6-311G**

Write down how many CGs some of the above basis sets will include for the following atoms: H, C, S

And for the following molecules: H₂O, CH₄
Polarisation and diffuse functions

Increasing the number of CGs per atomic orbital will never result in a good quality basis set. Other types of CGs need to be included, such as:

• CGs of angular momentum higher than in the valence orbitals of each atom. These “polarisation functions” enhance the “flexibility” of atoms to form chemical bonds in any direction and hence improve calculated molecular structures. Examples: 3-21G*, 6-31G*, 6-31G**, DVP, TZP, cc-pVDZ, cc-pVTZ

• CGs which extend further from the nucleus than the atomic orbitals. Such “diffuse functions” improve the predicted properties of species with extended electronic densities such as anions or molecules forming hydrogen bonds. Examples: 4-31+G, 6-31+G

• Basis sets are considered “balanced” when they include both polarisation and diffuse functions. Examples: 6-31+G*, 6-311++G**, aug-cc-pVDZ

Write down how many CGs some of the above basis sets will include for the following atoms: H, F, S and molecules: H₂O, CH₄
The complete basis set limit

• Basis sets are an approximation introduced in order to solve the KS equations for the MOs on a computer

• The MOs obtained are solutions of the Kohn-Sham equations only within the “function space” of the basis set used (so solutions within the STO-3G set of functions, or the 6-31G set, etc.)

• Improving the quality of the basis set requires increasing the number of CGs

• A complete basis set can represent exactly any molecular orbital

• Unfortunately, complete basis sets tend to have an infinite number of functions and are therefore not practical for calculations
Extrapolation to the complete basis set limit

• We can estimate the complete basis set result by systematically increasing the number of basis functions and extrapolating to an infinite-size basis set.

• The cc-pVDZ, cc-pVTZ, cc-pVQZ, etc, basis sets are an example of a systematic series of basis sets that can be extrapolated to the complete basis set limit.
Example calculations: Protonation energy of a water molecule

![Diagram showing protonation reaction]

<table>
<thead>
<tr>
<th>Basis set</th>
<th>$\text{H}_2\text{O}$ energy ($E_h$)</th>
<th>$\text{H}_3\text{O}^+$ energy ($E_h$)</th>
<th>Protonation energy ($E_h$)</th>
<th>Protonation energy (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>STO-3G</td>
<td>-75.3133</td>
<td>-75.6817</td>
<td>-0.3684</td>
<td>-231.2</td>
</tr>
<tr>
<td>STO-6G</td>
<td>-76.0366</td>
<td>-76.4015</td>
<td>-0.3649</td>
<td>-229.0</td>
</tr>
<tr>
<td>6-31G</td>
<td>-76.3852</td>
<td>-76.6721</td>
<td>-0.2869</td>
<td>-180.1</td>
</tr>
<tr>
<td>6-31++G</td>
<td>-76.4000</td>
<td>-76.6753</td>
<td>-0.2753</td>
<td>-172.7</td>
</tr>
<tr>
<td>6-31G**</td>
<td>-76.4197</td>
<td>-76.7056</td>
<td>-0.2859</td>
<td>-179.4</td>
</tr>
<tr>
<td>6-31++G**</td>
<td>-76.4341</td>
<td>-76.7078</td>
<td>-0.2738</td>
<td>-171.8</td>
</tr>
</tbody>
</table>
Example: the available choices of basis set in a quantum chemistry program

CHEM6085  Density Functional Theory
Basis sets on the Web

- Many kinds of basis sets have been developed over the years
- Most are available for download from websites, such as
  - [https://bse.pnl.gov/bse/portal](https://bse.pnl.gov/bse/portal)
  - [http://www.emsl.pnl.gov/forms/basisform.html](http://www.emsl.pnl.gov/forms/basisform.html)
Exponents and contraction coefficients

Basis sets are essential input data for calculations. Gaussian bases are represented by two kinds of numbers:

1) Exponents
2) Contraction coefficients

\[ G_\alpha(\mathbf{r}) = \sum_{\nu=1}^{N_\alpha} c_\nu e^{-\zeta_\nu \mathbf{r}} \]

Example: STO-3G basis for hydrogen

![Graph showing exponents and contraction coefficients for STO-3G basis](image)
Downloading basis sets

Example: Data needed for calculation on water with 6-31G** basis
Homework

1) Describe how many and what type (s, p, d) primitive and contracted Gaussians you will have in the STO-2G, DZP and 3-21G basis sets for the O atom.

2) The 3-21G* basis set for a carbon atom can be input into a quantum chemistry program using the following parameters (exponents and contraction coefficients):

\[
\begin{array}{c|c|c}
C & s & \\
172.256000000 & 0.06176690 \\
25.910900000 & 0.35879400 \\
5.533350000 & 0.70071300 \\
\end{array}
\]

\[
\begin{array}{c|c|c}
C & sp & \\
3.664980000 & -0.39589700 & 0.23646000 \\
0.770545000 & 1.21584000 & 0.86061900 \\
\end{array}
\]

\[
\begin{array}{c|c|c}
C & sp & \\
0.195857000 & 1.00000000 & 1.00000000 \\
\end{array}
\]

Which of the above parameters describe functions for the core electrons, valence electrons and for polarisation? Describe how you can “uncontract” the basis set and what effect this would have on your calculations.

3) Substitute the expression for the basis set expansion of a molecular orbital into the Schrödinger equation for the Kohn-Sham orbitals and derive a matrix representation of the Schrödinger equation, involving the “matrix elements” of the Kohn-Sham Hamiltonian \( H_{\alpha\beta} = \int G_{\alpha}(r)\hat{h}_{KS}G_{\beta}(r)\,dx \) and the overlap matrix of the basis functions (which are not orthogonal). This matrix equation can be solved on a computer to obtain the orbital expansion coefficients (diagonalisation of the Hamiltonian matrix) and is part of the traditional SCF procedure.
5-minute quiz

Name:
Surname:

1) Why are Slater functions more suitable than Gaussians as basis functions?

2) What do we mean when we say that a basis set contains “polarisation functions” and what do we mean by “diffuse functions”?

3) What is a “double-zeta basis set”? Would you expect a triple-zeta basis set to give better results than STO-3G and why?