Chapter 5

Non-orthonormal basis sets

We consider here the extension of the variational method to the case of non-orthonormal basis sets. We consider in particular the cause of Gaussian functions as basis set. This kind of basis set is especially used and useful in molecular calculations using Quantum Chemistry approaches.

5.1 Non-orthonormal basis set

Linear-algebra methods allow to treat with no special difficulty also the case in which the basis is formed by functions that are not orthonormal, i.e. for which

\[ S_{ij} = \langle b_i | b_j \rangle = \int b_i^* b_j \, dv \] (5.1)

is not simply equal to \( \delta_{ij} \). The quantities \( S_{ij} \) are known as overlap integrals. It is sometimes practical to work with basis of this kind, rather than with an orthonormal basis.

In principle, one could always obtain an orthonormal basis set from a non-orthonormal one using the Gram-Schmidt orthogonalization procedure. An orthonormal basis set \( \tilde{b}_i \) is obtained as follows:

\[
\begin{align*}
\tilde{b}_1 &= b_1 \\
\tilde{b}_2 &= b_2 - \frac{\tilde{b}_1 \langle \tilde{b}_1 | b_2 \rangle}{\langle \tilde{b}_1 | \tilde{b}_1 \rangle} \\
\tilde{b}_3 &= b_3 - \frac{\tilde{b}_2 \langle \tilde{b}_2 | b_3 \rangle}{\langle \tilde{b}_2 | \tilde{b}_2 \rangle} - \frac{\tilde{b}_1 \langle \tilde{b}_1 | b_3 \rangle}{\langle \tilde{b}_1 | \tilde{b}_1 \rangle}
\end{align*}
\] (5.5)

and so on. In practice, this procedure is seldom used and it is more convenient to follow a similar approach to that of Sect. 4.2. In this case, Eq. (4.27) is generalized as

\[ G(c_1, \ldots, c_N) = \sum_{ij} c_i^* c_j (H_{ij} - \epsilon S_{ij}) \] (5.5)

and the minimum condition (4.31) becomes

\[ \sum_j (H_{ij} - \epsilon S_{ij}) c_j = 0 \] (5.6)

or, in matrix form,

\[ Hc = \epsilon Sc \] (5.7)
known as generalized eigenvalue problem.

The solution of a generalized eigenvalue problem is in practice equivalent to the solution of two simple eigenvalue problems. Let us first solve the auxiliary problem:

\[ S d = \sigma d \] (5.8)

completely analogous to the problem (4.33). We can thus find a unitary matrix \( D \) (obtained by putting eigenvectors as columns side by side), such that \( D^{-1} S D \) is diagonal \( (D^{-1} = D^\dagger) \), and whose non-zero elements are the eigenvalues \( \sigma \). We find an equation similar to Eq.(4.39):

\[
\sum_i D_{ik}^* \sum_j S_{ij} D_{jn} = \sigma_n \delta_{kn}.
\] (5.9)

Note that all \( \sigma_n > 0 \): an overlap matrix is positive definite. In fact,

\[
\sigma_n = \langle \tilde{b}_n | \tilde{b}_n \rangle, \quad | \tilde{b}_n \rangle = \sum_j D_{jn} | b_j \rangle
\] (5.10)

and \( | \tilde{b} \rangle \) is the rotated basis set in which \( S \) is diagonal. Note that a zero eigenvalue \( \sigma \) means that the corresponding \( | \tilde{b} \rangle \) has zero norm, i.e. one of the \( b \) functions is a linear combination of the other functions. In that case, the matrix is called singular and some matrix operations (e.g. inversion) are not well defined.

Let us define now a second transformation matrix

\[
A_{ij} = \frac{D_{ij}}{\sqrt{\sigma_j}}.
\] (5.11)

We can write

\[
\sum_i A_{ik}^* \sum_j S_{ij} A_{jn} = \delta_{kn}
\] (5.12)

(note that \( A \) is not unitary) or, in matrix form, \( A^\dagger S A = I \). Let us now define

\[
c = A v
\] (5.13)

With this definition, Eq.(5.7) becomes

\[
H A v = \epsilon S A v
\] (5.14)

We multiply to the left by \( A^\dagger \):

\[
A^\dagger H A v = \epsilon A^\dagger S A v = \epsilon v
\] (5.15)

Thus, by solving the secular problem for operator \( A^\dagger H A \), we find the desired eigenvalues for the energy. In order to obtain the eigenvectors in the starting base, it is sufficient, following Eq.(5.13), to apply operator \( A \) to each eigenvector.
5.1.1 Gaussian functions

Gaussian functions are frequently used as basis functions, especially for atomic and molecular calculations. They are known as GTO: Gaussian-Type Orbitals. An important feature of Gaussians is that the product of two Gaussian functions, centered at different centers, can be written as a single Gaussian:

\[ e^{-\alpha(r-r_1)^2} e^{-\beta(r-r_2)^2} = e^{-(\alpha+\beta)(r-r_0)^2} e^{-\frac{\alpha\beta}{\alpha+\beta}(r_1-r_2)^2}, \quad r_0 = \frac{\alpha r_1 + \beta r_2}{\alpha + \beta}. \quad (5.16) \]

Some useful integrals involving Gaussian functions:

\[ \int_0^\infty e^{-\alpha x^2} \, dx = \frac{1}{\sqrt{\alpha}} \left( \frac{\pi}{2} \right)^{1/2}, \quad \int_0^\infty xe^{-\alpha x^2} \, dx = \left[ -\frac{e^{-\alpha x^2}}{2\alpha} \right]_0^\infty = \frac{1}{2\alpha}, \quad (5.17) \]

from which one derives

\[ \int_0^\infty e^{-\alpha x^2} x^{2n} \, dx = (-1)^n \frac{n!}{2^n} \int_0^\infty e^{-\alpha x^2} x^n \, dx = \frac{(2n-1)!!}{2^{n+1} \alpha^{n+1/2}} \quad (5.18) \]

\[ \int_0^\infty e^{-\alpha x^2} x^{2n+1} \, dx = (-1)^n \frac{n!}{2^n} \int_0^\infty e^{-\alpha x^2} x^n \, dx = \frac{n!}{2\alpha^{n+1}} \quad (5.19) \]

5.1.2 Exponentials

Basis functions composed of Hydrogen-like wave functions (i.e. exponentials) are also used in Quantum Chemistry as alternatives to Gaussian functions. They are known as STO: Slater-Type Orbitals. Some useful orbitals involving STOs:

\[ \int e^{-2Zr} \, d^3r = 4\pi \int_0^\infty re^{-2Zr} \, dr = 4\pi \left[ e^{-2Zr} \left( -\frac{r}{2Z} - \frac{1}{4Z^2} \right) \right]_0^\infty = \frac{\pi}{2Z^2} \quad (5.20) \]

\[ \int e^{-2Z(r_1+r_2)} \frac{d^3r_1 d^3r_2}{|r_1 - r_2|} = \frac{5\pi^2}{8Z^5}. \quad (5.21) \]

5.2 Code: hydrogen_gauss

Code `hydrogen_gauss.f90` (or `hydrogen_gauss.c`) solves the secular problem for the hydrogen atom using two different non-orthonormal basis sets:

1. a Gaussian, “S-wave” basis set:

\[ b_i(r) = e^{-\alpha_i r^2}; \quad (5.22) \]

2. a Gaussian “P-wave” basis set, existing in three different choices, corresponding to the different values \( m \) of the projection of the angular momentum \( L_z \):

\[ b_i(r) = xe^{-\alpha_i r^2}, \quad b_i(r) = ye^{-\alpha_i r^2}, \quad b_i(r) = ze^{-\alpha_i r^2}. \quad (5.23) \]

(actually only the third choice corresponds to a well-defined value \( m = 0 \))
The Hamiltonian operator for this problem is obviously
\[ H = -\frac{\hbar^2 \nabla^2}{2m_e} - \frac{Ze^2}{r} \quad (5.24) \]

For the hydrogen atom, \( Z = 1. \)

Calculations for S- and P-wave Gaussians are completely independent. In fact, the two sets of basis functions are mutually orthogonal: \( S_{ij} = 0 \) if \( i \) is a S-wave, \( j \) is a P-wave Gaussian, as evident from the different parity of the two sets of functions. Moreover the matrix elements \( H_{ij} \) of the Hamiltonian are also zero between states of different angular momentum, for obvious symmetry reasons. The \( S \) and \( H \) matrices are thus block matrices and the eigenvalue problem can be solved separately for each block. The P-wave basis is clearly unfit to describe the ground state, since it doesn’t have the correct symmetry, and it is included mainly as an example.

The code reads from file a list of exponents, \( \alpha_i \), and proceeds to evaluate all matrix elements \( H_{ij} \) and \( S_{ij} \). The calculation is based upon analytical results for integrals of Gaussian functions (Sec.5.1.1). In particular, for S-wave one has
\[ S_{ij} = \int e^{-(\alpha_i+\alpha_j)r^2} d^3r = \left( \frac{\pi}{\alpha_i + \alpha_j} \right)^{3/2} \quad (5.25) \]

while the kinetic and Coulomb terms in \( H_{ij} \) are respectively
\[ H^K_{ij} = \int e^{-\alpha_i r^2} \left[ -\frac{\hbar^2 \nabla^2}{2m_e} \right] e^{-\alpha_j r^2} d^3r = \frac{\hbar^2}{2m_e} \frac{6\alpha_i\alpha_j}{\alpha_i + \alpha_j} \left( \frac{\pi}{\alpha_i + \alpha_j} \right)^{3/2} \quad (5.26) \]
\[ H^V_{ij} = \int e^{-\alpha_i r^2} \left[ -\frac{Ze^2}{r} \right] e^{-\alpha_j r^2} d^3r = -\frac{2\pi Ze^2}{\alpha_i + \alpha_j} \quad (5.27) \]

For the P-wave basis the procedure is analogous, using the corresponding (and more complex) analytical expressions for integrals.

The code then calls subroutine diag that solves the generalized secular problem (i.e. it applies the variational principle). Subroutine diag returns a vector \( \mathbf{e} \) containing eigenvalues (in order of increasing energy) and a matrix \( \mathbf{v} \) containing the eigenvectors, i.e. the expansion coefficients of wave functions.

Internally, diag performs the calculation described in the preceding section in two stages. The solution of the simple eigenvalue problem is performed by the subroutine dsyev we have already seen in Sect.4.4.

In principle, one could use a single LAPACK routine, dsygv, that solves the generalized secular problem, \( H\psi = \epsilon S\psi \), with a single call. In practice, one has to be careful to avoid numerical instabilities related to the problem of linear dependencies among basis functions (see Eq.(5.10) and the following discussion). Inside routine diag, all eigenvectors of matrix \( S \) corresponding to very small eigenvectors, i.e. smaller than a pre-fixed threshold are thrown away, before proceeding with the second diagonalization. The number of linearly independent eigenvectors is reprinted in the output.

The reason for such procedure is that it is not uncommon to discover that some basis functions can almost exactly be written as sums of some other basis
functions. This does not happen if the basis set is well chosen, but it can happen if the basis set functions are too many or not well chosen (e.g. with too close exponents). A wise choice of the $\alpha_j$ coefficients is needed in order to have a high accuracy without numerical instabilities.

The code then proceeds and writes to files `s-coeff.out` (or `p-coeff.out`) the coefficients of the expansion of the wave function into Gaussians. The ground state wave function is written into file `s-wfc.out` (or `p-wfc.out`).

Notice the usage of `dgemm` calls to perform matrix-matrix multiplication. The header of `dgemm.f` contains a detailed documentation on how to call it. Its usage may seem awkward at a first (and also at a second) sight. This is a consequence in part of the Fortran way to store matrices (requiring the knowledge of the first, or “leading”, dimension of matrices); in part, of the old-style Fortran way to pass variables to subroutines only under the form of pointers (also for vectors and arrays). Since the Fortran-90 syntax and `MATMUL` intrinsic are so much easier to use: why bother with `dgemm` and its obscure syntax? The reason is efficiency: very efficient implementations of `dgemm` exist for modern architectures.

For the C version of the code, and how matrices are introduced and passed to Fortran routines, see Sec.0.1.4.

### 5.2.1 Laboratory

- Verify the accuracy of the energy eigenvalues, starting with 1 Gaussian, then 2, then 3. Try to find the best values for the coefficients for the 1$s$ state (i.e. the values that yield the lowest energy).

- Compare with the the solutions obtained using code `hydrogen_radial`. Plot the 1$s$ numerical solution (calculated with high accuracy) and the “best” 1$s$ solution for 1, 2, 3, Gaussians (you will need to multiply the latter by a factor $\sqrt{4\pi}$: why? where does it come from?). What do you observe? where is the most significant error concentrated?

- Compare with the results for the following optimized basis sets (a.u.):
  - three Gaussians: $\alpha_1 = 0.109818$, $\alpha_2 = 0.405771$, $\alpha_3 = 2.22776$ (known as “STO-3G” in Quantum-Chemistry jargon)
  - four Gaussians: $\alpha_1 = 0.121949$, $\alpha_2 = 0.444529$, $\alpha_3 = 1.962079$, $\alpha_4 = 13.00773$

- Observe and discuss the ground state obtained using the P-wave basis set

- Observe the effects related to the number of basis functions, and to the choice of the parameters $\alpha$. Try for instance to choose the characteristic Gaussian widths, $\lambda = 1/\sqrt{\alpha}$, as uniformly distributed between suitably chosen $\lambda_{\text{min}}$ and $\lambda_{\text{max}}$.

- For $Z > 1$, how would you re-scale the coefficients of the optimized Gaussians above?