Chapter 3

Basis functions in quantum chemistry

This chapter is adapted from Chapter 5 of Jensen's book: F. Jensen, 'Introduction to Computational Chemistry', Wiley.

In the derivation in the previous chapter, we have introduced the concept of basis function for the expansion of the one-electron molecular orbitals used for the generation of the many-electrons wave functions (Slater determinants or linear combination of Slater determinants).
There we derived the following expansion (eq. 2.17):

$$\phi_m(\mathbf{r}, s) = \sum_n D_{mn}\chi_n(\mathbf{r}) \tag{3.1}$$

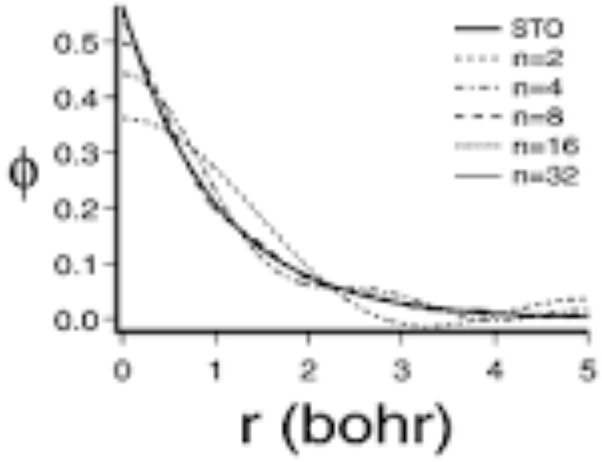
(where χ_n is an atom centered basis function and the spin dependent part of the wavefunctions is left out).
In this chapter, we introduce the different basis functions,χ_n commonly used in computational quantum chemistry.

Finiteness of Basis Sets: Approximations

One of the approximations inherent in essentially all *ab initio* methods is the introduction of a finite basis set. Expanding an unknown function, such as a molecular orbital, in a set of known functions is not an approximation, if the basis is complete. However, a complete basis means that an infinite number of functions must be used, which is impossible in actual calculations. An unknown MO can be thought of as a function in the infinite coordinate system spanned by the complete basis set. When a finite basis is used, only the components of the MO along those coordinate axes corresponding to the selected basis can be represented. The smaller the basis, the poorer the representation. The type of basis functions used also influences the accuracy. The better a single basis function is able to reproduce the unknown function, the fewer basis functions necessary for achieving a given level of accuracy. Knowing that the computational effort of ab initio methods scales formally at least as M^4, it is of course of prime importance to make the basis set as small as possible without compromising the accuracy.

3.1 Slater and Gaussian Type Orbitals

There are two types of basis functions (also called Atomic Orbitals, AO, although in general they are not solutions to an atomic Schrödinger equation) commonly used in electronic structure calculations: Slater Type Orbitals (STO) and Gaussian Type Orbitals (GTO).
A procedure that has come into wide use is to fit a Slater-type orbital (STO) to a linear combination of $n = 1, 2, 3, \ldots$ primitive Gaussian functions. This is the STO-nG procedure. In particular, STO-3G basis sets are often used in polyatomic calculations, in preference to evaluating integrals with Slater functions.



1

Figure 3.1: Comparison of Slater function with Gaussian function: least squares fits of a $1s$ Slater function ($\zeta = 1.0$) by a n GTOs

1. Slater type orbitals have the functional form

$$\chi_{\zeta,n,l,m}(r, \theta, \varphi) = N \, Y_{l,m}(\theta, \varphi) \, r^{n-1} \, e^{-\zeta r} \tag{3.2}$$

N is a normalization constant and $Y_{l,m}$ are the usual spherical harmonic functions. The exponential dependence on the distance between the nucleus and the electron mirrors the exact decay behavior of the orbitals for the hydrogen atom. However, since STOs do not have any radial nodes, nodes in the radial part are introduced by making linear combinations of STOs. The exponential dependence ensures a fairly rapid convergence with increasing number of functions, however, the calculation of three- and four-centre two electron integrals cannot be performed analytically. STOs are primarily used for atomic and diatomic systems where high accuracy is required, and in semi-empirical methods where all three- and four- center integrals are neglected.

2. Gaussian type orbitals can be written in terms of polar or cartesian coordinates

$$\chi_{\zeta,n,l,m}(r,\theta,\varphi) = N\,Y_{l,m}(\theta,\varphi)\,r^{2n-2-l}\,e^{-\zeta r^2} \tag{3.3}$$

$$\chi_{\zeta,l_x,l_y,l_z}(x,y,z) = N\,x^{l_x}\,y^{l_y}\,y^{l_z}\,e^{-\zeta r^2} \tag{3.4}$$

where the sum of l_x, l_y and l_z determines the type of orbital (for example $l_x + l_y + l_z = 1$ is a p-orbital)[1].

3. Comparison between STO and GTO

i. The r^2 dependence in the exponent makes the GTOs inferior to the STOs in two aspects. At the nucleus the GTO has zero slope, in contrast to the STO which has a "**cusp**" (discontinuous derivative), and GTOs have problems representing the proper behavior near the nucleus.

ii. The other problem is that the GTO falls off too rapidly far from the nucleus compared with an STO, and the "tail" of the wave function is consequently represented poorly.

iii. Both STOs and GTOs can be chosen to form a complete basis, but the above considerations indicate that more GTOs are necessary for achieving a certain accuracy compared with STOs. A rough guideline says that three times as many GTOs as STOs are required for reaching a given level of accuracy. The increase in number of basis functions, however, is more than compensated for by the ease by which the required

---

[1] Although a GTO appears similar in the two sets of coordinates, there is a subtle difference. A d-type GTO written in terms of the spherical functions has five components ($Y_{2,2}, Y_{2,1}, Y_{2,0}, Y_{2,-1}, Y_{2,-2}$), but there appear to be six components in the Cartesian coordinates ($x_2, y_2, z_2, xy, xz, yz$). The latter six functions, however, may be transformed to the five spherical d-functions and one additional s-function ($x^2 + y^2 + z^2$). Similarly, there are 10 Cartesian "f-functions" which may be transformed into seven spherical f-functions and one set of spherical p-functions. Modern programs for evaluating two-electron integrals are geared to Cartesian coordinates, and they generate pure spherical d-functions by transforming the six Cartesian components to the five spherical functions. When only one d-function is present per atom the saving by removing the extra s-function is small, but if many d-functions and/or higher angular moment functions (f-, g-, h- etc. functions) are present, the saving can be substantial. Furthermore, the use of only the spherical components reduces the problems of linear dependence for large basis sets, as discussed below.

integrals can be calculated. In terms of computational efficiency, GTOs are therefore preferred, and used almost universally as basis functions in electronic structure calculations.

3.2 Classification of Basis Sets

Having decided on the type of basis function (STO/GTO) and their location (nuclei), the most important factor is the number of functions to be used. The smallest number of functions possible is a **minimum basis set**. Only enough atomic orbital functions are employed to contain all the electrons of the neutral atom(s).

3.2.1 Minimum basis sets. Examples

For hydrogen (and helium) this means a single *s*-function. For the first row in the periodic table it means two *s*-functions ($1s$ and $2s$) and one set of *p*-functions ($2p_x$, $2p_y$ and $2p_z$). Lithium and beryllium formally only require two *s*-functions, but a set of p-functions is usually also added. For the second row elements, three *s*-functions ($1s$, $2s$ and $3s$) and two sets of *p*-functions ($2p$ and $3p$) are used.

3.2.2 Improvements

1. The first improvement in the basis sets is a doubling of all basis functions, producing a **Double Zeta** (DZ) type basis. The term zeta stems from the fact that the exponent of STO basis functions is often denoted by the greek letter ζ.
   A DZ basis thus employs two s-functions for hydrogen (1s and 1s'), four s-functions (1s, 1s', 2s and 2s') and two p-functions (2p and 2p') for first row elements, and six s-functions and four p-functions for second row elements. Doubling the number of basis functions allows for a much better description of the fact that the electron distribution in molecules can differ significantly from the one in the atoms and the chemical bond may introduce directionalities which can not be described by a minimal basis.
   The chemical bonding occurs between valence orbitals. Doubling the $1s$-functions in for example carbon allows for a better description of the $1s$-electrons. However, the 1s orbital is essentially independent of the chemical environment, being very close to the atomic case. A variation of the DZ type basis only doubles the number of valence orbitals, producing a **split valence basis**. [2].

---

[2] In actual calculations a doubling of the core orbitals would rarely be considered, and the term DZ basis is also used for split valence basis sets (or sometimes VDZ, for valence double zeta)

2. The next step up in basis set size is a **Triple Zeta** (TZ) basis. Such a basis contains three times as many functions as the minimum basis, i.e. six s-functions and three p-functions for the first row elements. Some of the core orbitals may again be saved by only splitting the valence, producing a triple zeta split valence basis set. The names **Quadruple Zeta** (QZ) and **Quintuple Zeta** (5Z, not QZ) for the next levels of basis sets are also used, but large sets are often given explicitly in terms of the number of basis functions of each type.

3. In most cases higher angular momentum functions are also important, these are denoted **polarization functions**. Consider for example a C-H bond which is primarily described by the hydrogen s-orbital(s) and the carbon s- and p_z-orbitals. It is clear that the electron distribution along the bond will be different than that perpendicular to the bond. If only s-functions are present on the hydrogen, this cannot be described. However, if a set of p-orbitals is added to the hydrogen, the *p* component can be used for improving the description of the H-C bond. The p-orbital introduces a polarization of the s-orbital(s). Similarly, d-orbitals can be used for polarizing p-orbitals, f-orbitals for polarizing d-orbitals etc. Once a p-orbital has been added to a hydrogen s-orbital, it may be argued that the p-orbital now should be polarized by adding a d-orbital, which should be polarized by an f-orbital, etc. For single determinant wave functions, where electron correlation is not considered, the first set of polarization functions (i.e. p-functions for hydrogen and d-functions for heavy atoms) is by far the most important, and will in general describe all the important charge polarization effects.
   Adding a single set of polarization functions (p-functions on hydrogens and d-functions on heavy atoms) to the DZ basis forms a Double Zeta plus Polarization (DZP) type basis [3]. Similarly to the sp-basis sets, multiple sets of polarization functions with different exponents may be added. If two sets of polarization functions are added to a TZ sp-basis, a Triple Zeta plus Double Polarization (TZ2P) type basis is obtained. For larger basis sets with many polarization functions the explicit composition in terms of number and types of functions is usually given. At the HF level there is usually little gained by expanding the basis set beyond TZ2P, and even a DZP type basis set usually gives "good" results (compared to the HF limit).

---

[3] There is a variation where polarization functions are only added to non-hydrogen atoms. This does not mean that polarization functions are not important on hydrogens. However, hydrogens often have a "passive" role, sitting at the end of bonds which does not take an active part in the property of interest. The errors introduced by not including hydrogen polarization functions are often rather constant and, as the interest is usually in energy differences, they tend to cancel out. As hydrogens often account for a large number of atoms in the system, a saving of three basis functions for each hydrogen is significant. If hydrogens play an important role in the property of interest, it is of course not a good idea to neglect polarization functions on hydrogens.

3.3 Basis set balance

In principle many sets of polarization functions may be added to a small sp-basis. This is not a good idea. If an insufficient number of sp-functions bas been chosen for describing the fundamental electron distribution, the optimization procedure used in obtaining the wave function (and possibly also the geometry) may try to compensate for inadequacies in the sp-basis by using higher angular momentum functions, producing artefacts. A rule of thumb says that the number of functions of a given type should at most be one less than the type with one lower angular momentum. A 3s2p1d basis is balanced, but a 3s2p2d2f1g basis is too heavily polarized.

Another aspect of basis set balance is the occasional use of **mixed basis sets**, for example a DZP quality on the atoms in the "interesting" part of the molecule and a minimum basis for the "spectator" atoms. Another example would be addition of polarization functions for only a few hydrogens which are located "near" the reactive part of the system. For a large molecule this may lead to a substantial saving in the number of basis functions. It should be noted that this may bias the results and can create artefacts. For example, a calculation on the H_2 molecule with a minimum basis at one end and a DZ basis at the other end will predict that H_2 has a dipole moment, since the variational principle will preferentially place the electrons near the center with the most basis functions. The majority of calculations are therefore performed with basis sets of the same quality (minimum, DZP, TZ2P, . . .) on all atoms, possibly cutting polarization and/or diffuse (small exponent) functions on hydrogens.

Except for very small systems it is impractical to saturate the basis set so that the absolute error in the energy is reduced below chemical accuracy, for example 1 kcal/ mol. The important point in choosing a balanced basis set is to keep the error as constant as possible. The use of mixed basis sets should therefore only be done after careful consideration. Furthermore, the use of small basis sets for systems containing elements with substantially different numbers of valence electrons (like LiF) may produce artefacts.

3.4 How do we choose the exponents in the basis functions?

The values for s- and p-functions are typically determined by performing variational HF calculations for atoms, using the exponents as variational parameters. The exponent values which give the lowest energy are the "best", at least for the atom. In some cases the optimum exponents are chosen on the basis of minimizing the energy of a wave function which includes electron correlation. The HF procedure cannot be used for determining exponents of polarization functions for atoms. By definition these functions are unoccupied in atoms, and therefore make no contribution to the energy. Suitable polarization exponents may be chosen by performing variational calculations on molecular systems (where the HF energy does depend on polarization functions) or on atoms with correlated wave functions. Since the main function of higher angular momentum functions is to recover electron correlation, the latter approach is usually preferred. Often only the optimum exponent is determined for a single polarization function, and multiple polarization functions are generated by splitting the exponents symmetrically around the optimum value for a single function. The splitting factor is typically taken in the range 2-4. For example if a single d-function for carbon has an exponent value of 0.8, two polarization functions may be assigned with exponents of 0.4 and 1.6 (splitting factor of 4).

3.5 Contracted Basis functions

One disadvantage of all **energy optimized basis sets** is the fact that they primarily depend on the wave function in the region of the inner shell electrons. The 1s-electrons account for a large part of the total energy, and minimizing the energy will tend to make the basis set optimal for the core electrons, and less than optimal for the valence electrons. However, chemistry is mainly dependent on the valence electrons. Furthermore, many properties (for example polarizability) depend mainly on the wave function "tail" (far from the nucleus), which energetically is unimportant. An energy optimized basis set which gives a good description of the outer part of the wave function needs therefore to be very large, with the majority of the functions being used to describe the 1s-electrons with an accuracy comparable to that for the outer electrons in an energetic sense. This is not the most efficient way of designing basis sets for describing the outer part of the wave function. Instead energy optimized basis sets are usually augmented explicitly with **diffuse functions** (basis functions with small exponents). Diffuse functions are needed whenever loosely bound electrons are present (for example in anions or excited states) or when the property of interest is dependent on the wave function tail (for example polarizability).
The fact that many basis functions go into describing the energetically important, but chemically unimportant, core electrons is the foundation for **contracted basis sets**.

1

An example. The carbon atom Consider for example a basis set consisting of 10 s-functions (and some p-functions) for carbon. Having optimized these 10 exponents by variational calculations on the carbon atom, maybe six of the 10 functions are found primarily to be used for describing the 1s orbital, and two of the four remaining describe the "inner" part of the 2s-orbital. The important chemical region is the outer valence. Out of the 10 functions, only two are actually used for describing the chemically interesting phenomena. Considering that the computational cost increases as the fourth power (or higher) of the number of basis functions, this is very inefficient. As the core orbitals change very little depending on the chemical bonding situation, the MO expansion coefficients in front of these inner basis functions also change very little. The majority of the computational effort is therefore spent describing the chemically uninteresting part of the wave function, which furthermore is almost constant. Consider now making the variational coefficients in front of the inner basis functions constant, i.e. they are no longer parameters to be determined by the variational principle. The 1s-orbital is thus described by a fixed linear combination of say six basis functions. Similarly the remaining four basis functions may be contracted into only two functions, for example by fixing the coefficient in front of the inner three functions. In doing this the number of basis functions to be handled by the variational procedure has been reduced from 10 to three.

Combining the full set of basis functions, known as the **primitive GTOs** (**PGTOs**), into a smaller set of functions by forming fixed linear combinations is known as **basis set contraction**, and the resulting functions are called **contracted GTOs** (**CGTOs**)

$$\chi(\mathrm{CGTO}) = \sum_{i}^{k} a_i \, \chi_i(\mathrm{PGTO}) \tag{3.5}$$

The previously introduced acronyms DZP, TZ2P etc., refer to the number of contracted basis functions. Contraction is especially useful for orbitals describing the inner (core) electrons, since they require a relatively large number of functions for representing the wave function cusp near the nucleus, and furthermore are largely independent of the environment. Contracting a basis set will always increase the energy, since it is a restriction of the number of variational parameters, and makes the basis set less flexible, but will also reduce the computational cost significantly. The decision is thus how much loss in accuracy is acceptable compared to the gain in computational efficiency.

3.5.1 The degree of contraction

The degree of contraction is the number of PGTOs entering the CGTO, typically varying between 1 and 10. The specification of a basis set in terms of primitive and contracted functions is given by the notation

$$(10\mathrm{s}4\mathrm{p}1\mathrm{d}/4\mathrm{s}1\mathrm{p}) \longrightarrow [3\mathrm{s}2\mathrm{p}1\mathrm{d}/2\mathrm{s}1\mathrm{p}] \quad . \tag{3.6}$$

The basis in parentheses is the number of primitives with heavy atoms (first row elements) before the slash and hydrogen after. The basis in the square brackets is the number of contracted functions. Note that this does not tell how the contraction is done, it only indicates the size of the final basis (and thereby the size of the variational problem in HF calculations).

3.6 Example of Contracted Basis Sets; Pople Style Basis Sets

There are many different contracted basis sets available in the literature or built into programs, and the average user usually only needs to select a suitable quality basis for the calculation. For short description of some basis sets which often are used in routine calculations (see for instance the book of Frank Jensen, *Introduction to Computational Chemistry*, Wiley, 2002. Chapter 5).

STO-*n*G basis sets n PGTOs fitted to a 1 STO. This is a minimum type basis where the exponents of the PGTO are determined by fitting to the STO, rather than optimizing them by a variational procedure. Although basis sets with $n = 2 - 6$ have been derived, it has been found that using more than three PGTOs to represent the STO gives little improvement, and the STO-3G basis is a widely used minimum basis. This type of basis set has been determined for many elements of the periodic table. The designation of the carbon/hydrogen STO-3G basis is (6s3p/3s) $\longrightarrow$ [2s1p/1s].

***k*-*nlm*G basis sets** These basis sets have been designed by Pople and coworkers, and are of the split valence type, with the k in front of the dash indicating how many PGTOs are used for representing the core orbitals. The nlm after the dash indicate both how many functions the valence orbitals are split into, and how many PGTOs are used for their representation. Two values (e.g. nl) indicate a split valence, while three values (e.g. nlm) indicate a triple split valence. The values before the G (for Gaussian) indicate the s- and p-functions in the basis; the polarization functions are placed after the G. This type of basis sets has the further restriction that the same exponent is used for both the s- and p-functions in the valence. This increases the computational efficiency, but of course decreases the flexibility of the basis set. The exponents in the PGTO have been optimized by variational procedures.

3-21G This is a split valence basis, where the core orbitals are a contraction of three PGTOs, the inner part of the valence orbitals is a contraction of two PGTOs and the outer part of the valence is represented by one PGTO. The designation of the carbon/hydrogen 3-21G basis is (6s3p/3s) $\longrightarrow$ [3s2p/2s]. Note that the 3-21G basis contains the same number of primitive GTOs as the STO-3G, however, it is much more flexible as there are twice as many valence functions which can combine freely to make MOs.

6-31G This is also a split valence basis, where the core orbitals are a contraction of six PGTOs, the inner part of the valence orbitals is a contraction of three PGTOs and the outer part of the valence represented by one PGTO. The designation of the carbon/hydrogen 6-31G basis is (10s4p/4s) $\longrightarrow$ [3s2p/2s]. In terms of contracted basis functions it contains the same number as 3-21G, but the representation of each functions is better since more PGTOs are used.

6-311G This is a triple zeta split valence basis, where the core orbitals are a contraction of six PGTOs and the valence split into three functions, represented by three, one, and one PGTOs, respectively.

To each of these basis sets one can add **diffuse** and/or **polarization functions**.

- **Diffuse functions** are normally s- and p-functions and consequently go before the G. They are denoted by + or ++, with the first + indicating one set of diffuse s- and p-functions on heavy atoms, and the second + indicating that a diffuse s-function is also added to hydrogens. The arguments for adding only diffuse functions on non-hydrogen atoms is the same as that for adding only polarization functions on non-hydrogens.

- **Polarization functions** are indicated after the G, with a separate designation for heavy atoms and hydrogens. The *6-31+G(d)* is a split valence basis with one set of diffuse sp-functions on heavy atoms only and a single d-type polarization function on heavy atoms.
  A *6-311++G(2df,2pd)* is similarly a triple zeta split valence with additional diffuse sp-functions, and two d- and one f-functions on heavy atoms and diffuse s- and two p- and one d-functions on hydrogens. The largest standard Pople style basis set is *6-311 ++G(3df, 3pd).* These types of basis sets have been derived for hydrogen and the first row elements, and same of the basis sets have also been derived for second and higher row elements.
  If only one set of polarization functions is used, an alternative notation in terms of * is also widely used. The *6-31G** basis is identical to *6-31G(d)*, and **6-31G**** is identical to *6-31G(d,p).* A special note should be made for the *3-21G** basis. The 3-21G basis is basicly too small to support polarization functions (it becomes unbalanced). However, the *3-21G* basis by itself performs poorly for hypervalent molecules, such as sulfoxides and sulfones. This can be substantially improved by adding a set of d-functions. The *3-21G** basis has only d-functions on second row elements (it is sometimes denoted *3-21G(*)* to indicate this), and should not be considered a polarized basis. Rather, the addition of a set of d-functions should be considered an ad hoc repair of a known flaw.