Basis Set Superposition Error (BSSE). A short intro

Posted by joaquinbarroso

Molecular Orbitals (MOs) are linear combinations of Atomic Orbitals (AOs), which in turn are linear combinations of other functions called 'basis functions'. A basis, or more accurately a basis set, is a collection of functions which obey a set of rules (such as being orthogonal to each other and possibly being normalized) with which all AOs are constructed, and although these are centered on each atomic nucleus, the canonical way in which they are combined yield delocalized MOs; in other words, an MO can occupy a large space spanning several atoms at once. We don't mind this expansion across a molecule, but what about between two molecules? Calculating the interaction energy between two or more molecular fragments leads to an artificial extra–stabilization term that stems from the fact that electrons in molecule 1 can occupy AO's (or the basis functions which form them) centered on atoms from molecule 2.

Fundamentally, the interaction energy of any A—B dimer, E_{int}, is calculated as the energy difference between the dimer and the separately calculated energies for each component (Equation 1).

$$E_{int} = E_{AB} - E_A - E_B \quad (\mathbf{1})$$

However the calculation of E_{int} by this method is highly sensitive to the choice of basis set due to the Basis Set Superposition Error (BSSE) described in the first paragraph. The BSSE is particularly troublesome when small basis sets are used, due to the poor description of dispersion interactions but treating this error by just choosing a larger basis set is seldom useful for systems of considerable sizes. The Counterpoise method is a nifty correction to equation 1, in which EA and EB are calculated with the basis set of A and B respectively, i.e., only in EAB a larger basis set (that of A and B simultaneously) is used. The Counterpoise method calculates each component with the AB basis set (Equation 2)

$$E_{int}^{CP} = E_{AB}^{AB} - E_A^{AB} - E_B^{AB} \quad (\mathbf{2})$$

where the superscript AB means the whole basis set is used. This is accomplished by using '*ghost*' atoms with no nuclei and no electrons but empty basis set functions centered on them.

In Gaussian, BSSE is calculated with the Counterpoise method developed by Boys and Simon. It requires the keyword Counterpoise=N where N is the number of fragments to be considered (for an A—B system, N=2). Each atom in the coordinates list must be specified to which fragment pertains; additionally, the charge and multiplicity for each fragment and the whole supermolecular ensemble must be specified. Follow the example of this hydrogen fluoride dimer.

```
%chk=HF2.chk
#P opt wB97XD/6-31G(d,p) Counterpoise=2

HF dimer

0,1 0,1 0,1
```

```
H(Fragment=1) 0.00 0.00 0.00
F(Fragment=1) 0.00 0.00 0.70
H(Fragment=2) 0.00 0.00 1.00
F(Fragment=2) 0.00 0.00 1.70
```

For closed shell fragments the first line is straightforward but one must pay attention that the first pair of numbers in the charge multiplicity line correspond to the whole ensemble, whereas the folowing pairs correspond to each fragment in consecutive order. Fragments do not need to be specified contiguously, i.e., you don't need to define all atoms for fragment 1 and after those the atoms for fragment 2, etc. They could be mixed and the program still assigns them correctly. Just as an example I typed wB97XD but any other method, DFT or ab initio, may be used; only semiempirical methods do not admit a BSSE calculation because they don't make use of a basis set in the first place!

The output provides the corrected energy (in atomic units) for the whole system, as well as the BSSE correction (which added to the previous term yields the un-corrected energy of the system). Gaussian16 also provides these values in kcal/mol as 'Complexation energies' first raw (uncorrected) and then the corrected energy.

BSSE is always present and cannot be entirely eliminated because of the use of finite basis sets but it can be correctly dealt with if the Counterpoise method is included.