17. Population Analysis

Population analysis is the study of charge distribution within molecules. The intention is to accurately model partial charge magnitude and location within a molecule. This can be thought of as a rigorous version of assigning partial charges on the atoms like chemists often do in Lewis dot structures. Partial atomic charges are not observable characteristics of molecules, and therefore the entire idea of modeling electron population is not unique. In order to assign charges to atoms, one must define the spatial region of those atoms, then add up all the charge in that region (integrate charge density over the volume). This may sound trivial, but where does one atom’s electron cloud end and the next begin? Which nucleus does a specific region of the electron cloud belong to? The process is arbitrary, but the results are useful.

17.1. Mulliken population analysis

Mulliken population analysis is by default always performed in Gaussian. It is based on the linear combination of atomic orbitals and therefore the wave function of the molecule. The electrons are partitioned to the atoms based on the nature of the atomic orbitals’ contribution to the molecular wave function. Generally, the total number of electrons in the molecule \( N \) can be expressed as:

\[
N = \sum_j \text{electrons} \int \psi_j(r_j)\psi_j(r_j)dr_j
= \sum_j \text{electrons} \sum_{r,s} \int c_{jr}\varphi_r(r_j)c_{js}\varphi_s(r_s)dr_j
= \sum_j \left( \sum_r c_{jr}^2 + \sum_{r\neq s} c_{jr}c_{js}S_{rs} \right)
\]

Where \( r \) and \( s \) index the AO basis functions \( \varphi \), \( c_{jr} \) are coefficients of the basis function \( r \) in the MO \( j \), and \( S \) is the overlap matrix defined before. This shows that the total number of electrons can be divided into two sums: the first one including only squares of single AO basis function (\( r \)), and the other one products of two different AO functions (\( r \) and \( s \)). Clearly the first term can be thought of as electrons belonging to the particular atom. It is the second term that causes problems – there is no single best way how to divide the shared electrons between the two atoms.

Mulliken suggested to split the shared density 50:50. Then the electrons associated with the atom \( k \) are given by:

\[
N_k = \sum_j \text{electrons} \left( \sum_{r\in k} c_{jr}^2 + \sum_{r,s\in k,r\neq s} c_{jr}c_{js}S_{rs} + \sum_{r\in k,s\notin k} c_{jr}c_{js}S_{rs} \right)
\]
where the first two terms come from the basis functions on the $k$th atom and the last term is the part shared with all other atoms. The partial charge on the atom $k$ is then:

\[ q_k = Z_k - N_k \]  

(127)

where $Z_k$ is its atomic number.

**Advantages:** This theory is available for use in nearly every software program for molecular modeling, and is computationally cheap. For minimal and small split valence basis sets, this method quickly gives chemically intuitive charge sign on atoms and usually reasonable charge magnitudes. It works well for comparing changes in partial charge assignment between two different geometries when the same size basis set is used.

**Disadvantages:** The partial charges assigned to atoms using Mulliken population analysis vary significantly for the same system when different size basis sets are used, so computations using different basis sets cannot be compared. If non-orthonormal basis sets are used, individual basis functions can have occupation numbers greater than 1 (or 2 in restricted theories), which is physically meaningless. This instability of charge with increasing basis set size is a major disadvantage to the theory. There is no way to account for differences in electronegativities of atoms within the molecule; the method always equally distributes shared electrons between two atoms.

17.2. Löwdin population analysis:

The Löwdin population analysis method sought to improve upon the Mulliken method, mostly to correct the instability of predicted charges with increasing basis set size. This is achieved by transforming the atomic orbital basis functions into an orthonormal set of basis functions prior to the population analysis. Orthonormal means that they do not overlap (the overlap - $S$ - is zero). Therefore, the transformations eliminates the overlap term and with it the need for deciding what to do with it. After proper transformation the formula for $N_k$ from Löwdin analysis would correspond to that for the Mulliken one above, but only with the first term: the other two would be zero due to orthonormalizaiton.

**Advantages:** Löwdin population analysis is more stable than Mulliken with changes in the basis set, although with very large basis sets it may also have problems.

**Disadvantages:** This method uses a symmetric orthogonalization scheme of the atomic orbitals, and still does not account for electronegativity of different atoms. It is more computationally expensive than Mulliken analysis. Unfortunately it appears there is no way to use Löwdin’s method with our software.
17.3. Natural Bond Orbitals (NBO):

To use NBO for electron population analysis in Gaussian, include Pop=NPA in the input file. NPA stands for Natural Population Analysis and is based on the Natural Bond Orbital (NBO) scheme (the complete NBO analysis can be done using Pop=NBO).

Natural bond analysis classifies and localizes orbitals into three distinct groups: non-bonding natural atomic orbitals (NAOs), orbitals involved in bonding and antibonding (NBOs), and Rydberg type orbitals. The Rydberg type orbitals and NAOs are made up of basis sets of single atoms and the NBOs are a combination of basis set atomic orbitals of two atoms. This is similar to our notion of core electrons, lone pairs of electrons, and valence electrons, and works under the assumption that only the bonding orbitals should be made by combinations of two atoms’ basis sets. Based on this model of electron partitioning, Natural Population Analysis then treats the NBOs as the Mulliken method treats all the orbitals.

**Advantages:** This method differentiates between the orbitals that will overlap to form a bond and those that are too near the core of an atom to be involved in bonding. This results in convergence of atomic partial charge to a stable value as the basis set size is increased.

**Disadvantages:** More computationally expensive than both Mulliken and Löwdin methodologies. NPA also tends to predict larger charges than several other population analysis methods, so like Mulliken charges NPA is best used for comparing differences rather than determining absolute atomic charges.

17.4. Atoms in Molecules (AIM):

The population analysis using Atoms in Molecules theory is requested by keyword AIM in a Gaussian input file. Note that you don’t find this in the manual (or Gaussian website).

The notion of an atom in the molecule, and therefore the location of the partial charges seems trivial, but how do we define the atom within a molecule? It is not a point, nor is it a clearly defined sphere. The electron cloud is diffuse, and the exact solution to the Schrödinger equation requires us to consider all space, so how to define the spatial existence of an atom is not so trivial. Atoms in Molecules theory creates the spatial partition of atoms with zero-flux surfaces of the electron density. Electron density is measureable, though AIM bases its calculations on the calculated electron density. The zero flux surfaces define the spatial region over which the electron density is integrated to determine total charge in an atom’s “space”. The existence of critical points (where the derivative of the flux surface is zero) defines the existence of a bond between two nuclei in AIM.

As stated by Cramer, “We may define the zero-flux surface mathematically as the union of all points for which \( \nabla \rho \cdot \mathbf{n} = 0 \) where \( \rho \) is the electron density function (from the wave function) and \( \mathbf{n} \) is the unit vector normal to the surface”. Once the zero-flux surfaces (and
therefore the spatial region $\Omega_k$ for an atom) are calculated, the partial charge on the atom is computed as:

$$q_k = Z_k - \int_{\Omega_k} \rho(r) \, dr$$ \hspace{1cm} (128)

The image below is from the Lewars text book. It shows the trajectories of the gradient vector field of electron density (the arrows), A and B are two nuclei, the dark line labeled S is a slice of the zero-flux surface that defines the spatial region of each atom, and the point C is the bond critical point.

**Advantages:** The partitioning of an atom’s volume makes sense, and this method does take electronegativity into account. This method is independent of basis set as well as the calculation method (level of theory).

**Disadvantages:** This method produces partial charges that are often seemingly odd. Given as an example in the Lewars text, a saturated hydrocarbon is assigned weakly positive carbons and weakly negative hydrogens, not what every other method for computing partial charges predicts. The odd behavior is rooted in the non-uniform distribution of charge within the AIM definition of an atoms space. This method is also relatively computationally expensive. As a result, this method is not often used.
The above image (also from Lewars) provides a good visualization for the meaning of a bond critical point. It resembles the transition state structure of a PES, but the “surface” is the inverse of electron density.

17.5. Population analyses base on the Molecular Electrostatic Potential

The molecular electrostatic potential (MEP) is observable and is used for both Merz-Kollman and ChelpG (CHarges from ELectrostatic Potentials using a Grid based method) population analysis methods. These methods are requested through Pop=MK and Pop=ChelpG, respectively.

Though the MEP is observable (can be measured) in the population analysis it is computed from the wave function of the system. The most notable disadvantage to both MK and ChelpG is that they do not work well for large systems, especially those with “inner” atoms because the MEP is a surface characteristic.

The ChelpG and MK population analysis assign atomic partial charges such that the MEP is replicated. The quality that sets the two methods apart is that with ChelpG the MEP is calculated along points that are spaced 3.0picometers apart and lie along a cube that encompasses the molecule. Any points lying on this cube but within the van der Waals surface are discarded. The dimensions of the cube are such that the closest the cube is to the molecule is 28.0picometers. Once the cube is formed, and the MEP has been calculated at points on the cube, the atomic charges are assigned such that the MEP is reproduced. (Ramachandran, p. 254)

The MK population analysis method calculates the MEP along 4 layers encompassing the molecule, each layer is scaling factor larger than the van der Waals surface (layer one is at 1.4 the van der Waals radii, layer two at 1.6, layer three at 1.8, and layer four at 2.0) (Ramachandran, 254). The molecular electrostatic potential is computed at many points along the layers from the
wave function, and the electron distribution is then made to replicate the molecular electrostatic potential. The number of layers can be increased by calling for it in the input file.

17.6. Comparing methods:

The following table compares the calculated charge on the hydrogen atom of HF using several sized basis sets and three different population analysis methods. The basis set size dependence of the Mulliken method is apparent, while for the electrostatic (MK and ChelpG) and NBO methods there is significantly less variation in predicted charges as basis set size is changed.

<table>
<thead>
<tr>
<th>Level</th>
<th>Charge on H (= −charge on F)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mulliken</td>
</tr>
<tr>
<td>HF/STO-3G</td>
<td>0.19</td>
</tr>
<tr>
<td>HF/3–21G(*)</td>
<td>0.45</td>
</tr>
<tr>
<td>HF/6–31G*</td>
<td>0.52</td>
</tr>
<tr>
<td>HF/6–31G**</td>
<td>0.39</td>
</tr>
<tr>
<td>HF/6–311G**</td>
<td>0.32</td>
</tr>
<tr>
<td>6–31+G*</td>
<td>0.57</td>
</tr>
<tr>
<td>6–311++G**</td>
<td>0.3</td>
</tr>
<tr>
<td>MP2/6–31G*</td>
<td>0.52</td>
</tr>
</tbody>
</table>

References


