Parameter Specification for (generalized) Pariser-Parr-Pople Models

Recall that the generalized Pariser-Parr-Pople model is defined as:

$$\hat{H}_{\text{PPP+P}} = \sum_{pq} h_{pq} a_p^\dagger a_q + \sum_p U_p \hat{n}_{p\alpha} \hat{n}_{p\beta} + \frac{1}{2} \sum_{p \neq q} \gamma_{pq} (\hat{n}_{p\alpha} + \hat{n}_{p\beta} - Q_p)(\hat{n}_{q\alpha} + \hat{n}_{q\beta} - Q_q) + \sum_{p \neq q} g_{pq} a_{p\alpha}^\dagger a_{p\beta}^\dagger a_{q\beta} a_{q\alpha}$$

To use this Hamiltonian, one must somehow define the parameters therein.

Direct specification.

For each atom/site in the system, one can explicitly define the key integrals. These are: - `h[:,:]` a 2D symmetric array with extent given by the number of atoms/sites. - `u_onsite[:]` a 1D array with extent given by the number of atoms/sites. - `gamma[:,:]` a 2D symmetric array with extent given by the number of atoms/sites. The possibility that $\gamma_{pp} \neq 0$ is supported but if it is present, the constant term is discarded and the other terms are partitioned into h and U. - `charges[:]` a 1D array specifying the charge on each atoms/sites. - `g_pair[:,:]` a 2D symmetric array specifying the pairing interaction.

Indirect specification as constants

It is not required to directly specify the matrices but if they are not specified, one must define them or abide by the defaults. The simplest choice is to define the key parameters below. The default values are based on carbon $2p$ orbitals for a π-electron Hamiltonian. - `alpha`. $h_p p = \alpha$. Default value is $\alpha_{\text{default}} = -0.414$ a.u.. - `beta`. $h_{p \neq q} = \beta$ if p is connected to q and otherwise $h_{p \neq q} = 0$. Default value is $\beta_{\text{default}} = -0.0533$ a.u.. This value is relatively large compared to most accepted values. - `U_onsite`. The "Hubbard" U value is chosen based on the Pariser-Parr strategy as the chemical hardness of the Carbon atom. $U_{\text{default}} = 0.417$ a.u.. The reference data from T. G. Schmalz, "From the Hubbard to the PPP Model," *Croatica Chemica Acta* **86**, 419-423 (2013) is used. - `gamma`. The default value of the Pariser-Parr off-site repulsion is based on the Ohno model and the reference T. G. Schmalz, "From the Hubbard to the PPP Model," *Croatica Chemica Acta* **86**, 419-423 (2013). $\gamma_{\text{default}} = 0.0784$ a.u.. - `g_pair`. The default value of the pairing strength is set to zero, $g_{pq} = 0$.

Specification through atom-type and bond-type dictionaries.

For each atom/site in the system, one can explicitly define the atom types. If the atom types are specified, then the adjacency matrix is used (if needed) to

specify the bond types. The parameters are defined as below. - `atom_types[:]` a 1D array with extent given by the number of atoms/sites.

There are two(ish) dictionaries. One is a dictionary based on isolated atoms; the other is based on Rauk's compilation.

Once α and β are defined, the 1-electron integrals are also defined.

$$h_{pp} = \alpha_p h_{p \neq q} = \beta_{pq}$$

For the dictionaries for α_X and U_X, the dictionary has the form of a list of atom types (keys) with the values of the corresponding parameter. For the dictionaries for β_{XY} and γ_{XY}, the dictionaries have the form of a set of two atom-types (keys) and the values are an array; if connectivity information is provided, the one is subtracted from the integer value for the bond-type, and that entry of the array is used. (This allows different bond types between identical atom types.)

Rauk's dictionary

Rauk's dictionary is taken from: Arvi Rauk, The Orbital Interaction Theory of Organic Chemistry , Second Edition, Wiley-Interscience, New York, 2001..

and based on F. A. Van-Catledge, "A Pariser-Parr-Pople-based set of Hueckel molecular orbital parameters." *J. Org. Chem.* **45**, 4801-4802 (1980)

which chose parameters based on D. L. Beveridge and J. Hinze, "Parameterization of Semiempirical π-electron Molecular Orbital Calculations. π-systems containing Carbon, Nitrogen, Oxygen, and Fluorine." *J. Am. Chem. Soc.* **93**, 3107-3114 (1971).

where Hubbard-U parameters are not available, the Pariser approximation is used with ionization potentials and electron affinities taken from: C. Cardenas, F. Heidar-Zadeh, and P. W. Ayers, "Benchmark values of chemical potential and chemical hardness for atoms and atomic ions (including unstable anions) from the energies of isoelectronic series" *PCCP* **18**, 25721-25734 (2016).

The Rauk dictionary uses *only* bond connectivity data; all bonds between atoms of the same type are assumed to be the same. Therefore: - The Rauk dictionary uses only data from the adjacency matrix, not from the distance matrix. - The Rauk dictionary does not allow different bond types between atoms of the same types. - The only atom types supported are `C, B, N2, N3, O1, O2, F, Si,`

`P2`, `P3`, `S1`, `S2`, `Cl`. The numbers after `N`, `O`, `P`, and `S` refer to the coordination number.

Traditionally, the values of α and β are defined for the 2p orbital in an sp^2 hybridized carbon atom, with reasonable values being:

$$\alpha_C = -11.26 \text{ eV} = 0.414 \text{ Hartree}$$

(2)

$$\beta_{CC} = -1.45 \text{ eV} = 0.0533 \text{ Hartree}$$

(3) The α_C value is chosen from the binding energy (minus one times the ionization potential) of the Carbon atom; the β_{CC} value is defined by Rauk based on the carbon-carbon bond strength in ethylene, and is on the high end of the "accepted" range of values. For heteroatoms, these values are shifted by:

$$\alpha_X = \alpha + h_X \cdot |\beta|$$

(4)

$$\beta_{XY} = k_{XY}|\beta|$$

(5) A table of these values is given by Rauk.

Rauk's parameters come from Van Catledge, who in turn defines them based on a procedure developed by Beveridge and Hinze. This means that the U and γ values should be deduced using that procedure. Not all values are available, but as Beveridge and Hinze base their parameters on the Pariser approximation, $U_X = I_X - A_X$, that approximation can be used with additional parameters are needed. The predefined values are:

$$U_C = 0.409 U_{N2} = 0.453 U_{N3} = 0.616 U_{O1} = 0.560 U_{O2} = 0.692 U_F = 0.815$$

Based on the Pariser approximation, we would deduce the additional values. These values tend to be systematically too small (because they include relaxation of the core and the sigma framework that are ignored in PPP methods). One could argue for shifting these values, but it is very *ad hoc*

$$U_B = 0.295 U_{Si} = 0.293 U_{P2} = 0.358 U_{P3} = 0.358 U_{S2} = 0.304 U_{S3} = 0.304 U_{Cl} = 0.344$$

The Parr-Pariser integrals then have values given by:

$$\gamma_{XY} = \frac{\bar{U}_{XY}}{\bar{U}_{XY} R_{XY} + e^{-\frac{1}{2}\bar{U}_{XY}^2 R_{XY}^2}}$$

where

$$\bar{U}_{XY} = \tfrac{1}{2}(U_X + U_Y)$$

and bond lengths and energies are in atomic units.

The site background charges are all one by default $Q_X = 1$. However, this doesn't mean that the sites would not be charged: typically the number of

electrons on the Boron site is zero, so the "effective charge" on a Boron site is +1. Conversely, in a neutral molecule, `N3, O2, F, P3, S2, Cl` all contribute 2 electrons (so the effective charges on those sites is negative).

In Rauk's dictionary, $g_{XY} = 0$.

TODO Make a table of Rauk's values.

TODO Using characteristic molecules, find a reasonable choice for detail bond-length values, R_{XY}.

Pariser-Parr Atomic Dictionary

The Pariser-Parr approach allows parameters to be defined for every (neutral) atom, albeit with the relatively severe assumption that there is only one orbital per site. The basic approximation is:

$$\alpha_X = -I_X$$

$$U_X = I_X - A_X$$

$$Q_X = 1$$

where I_X and A_X are the ionization potential and electron affinity of the neutral atom X in its ground state. If only connectivity information is provided, the atomic-distance matrix is populated by using the sum of the covalent radii of the atoms. (This tends to be too long for *pi*-electron systems, as the covalent radii are more indicative of the length of single bonds.)

The values of the bond parameters are determined by the Wolfsberg-Helmholz approximation,

$$\beta_{XY} = 1.75 S_{XY} \frac{\alpha_X + \alpha_Y}{2}$$

The user can pass in an overlap matrix but, failing this, we provide generic overlap matrices appropriate for $2p$-π orbitals and $1s$-σ orbitals. To evaluate these, we first define:

$$p = -\frac{\alpha_X + \alpha_Y}{2} R_{XY}$$

$$t = \left| \frac{\alpha_X - \alpha_Y}{\alpha_X + \alpha_Y} \right|$$

$$a_n(x) = n! \sum_{k=1}^{n+1} \frac{1}{x^k (n-k+1)!}$$

$$b_n(x) = n! \sum_{k=1}^{n+1} \frac{(-1)^{n-k}}{x^k (n-k+1)!}$$

where R_{XY} is the distance between atoms X and Y. The overlap between two $1s$ orbitals binding in the π orientation can be deduced from the reference of Mulliken, Rieke, Orloff, and Orloff (MROO) as:

$$S_{XY}^{1s-\sigma} = \begin{cases} e^{-p}(1 + p + \frac{1}{3}p^2) & t = 0 \text{ (Eq. 51 in MROO)} \\ \dfrac{(1-t^2)^{3/2}p^3}{4}\left[A_2 B_0 - A_0 B_2\right] & t \neq 0 \text{ (Eq.18 in MROO)} \end{cases}$$

The overlap between two $2p$ orbitals binding in the π orientation can be deduced from the reference of Mulliken, Rieke, Orloff, and Orloff (MROO) as:

$$S_{XY}^{2p-\pi} = \begin{cases} e^{-p}(1 + p + \frac{2}{5}p^2 + \frac{1}{15}p^3) & t = 0 \text{ (Eq. 61 in MROO)} \\ \dfrac{(1-t^2)^{5/2}p^5}{120}\left[A_4(B_0 - B_2) - A_2(B_0 - B_4) + A_0(B_2 - B_4)\right] & t \neq 0 \text{ (Eq. 46 in MROO)} \end{cases}$$

where
$$A_n = e^{-p}a_n(p)$$

$$B_n = \begin{cases} \frac{2}{k+1} & t = 0 \\ -e^{-pt}a_n(pt) - e^{pt}b_n(pt) & t > 0 \end{cases}$$

These formulas, which I transcribed from MROO, should be double-checked. One unit test is to affirm that the $t = 0$ limit is recovered when $t \to 0^+$. Another is to check that $R \to 0^+$, $S_{XY} = (1-t^2)^{5/2}$ (Eq. 68 in MROO).

The Pariser-Parr integrals can be approximated with the Ohno approximation,

$$\gamma_{XY} = \frac{\bar{U}_{XY}}{\sqrt{1 + \bar{U}_{XY}^2 R_{XY}^2}}$$

In this dictionary, $g_{XY} = 0$.

Overlaps of Slater-type orbitals from R. S. Mulliken, C. A. Rieke, D. Orloff, and H. Orloff, "Formulas and Numerical Tables for Overlap Integrals", *J. Chem. Phys.* **17**, 1248-1267 (1949).