

Valera Veryazov valera.veryazov@teokem.lu.se

March 2025



What is a Basis Set from a Mathematical Perspective?

- Basis set in 3D space
- Completeness?
- Uniqueness?
- Transformation of basis sets
- Ill-defined basis sets
- Orthonormality
- Generalization to a Hilbert space
- Partial Differential Equations vs Linear Algebra

- First try for an H basis set: e^{-r/a_0}
- What happens if we use it for H₂? too weak bond, electrons prefer not to leave atoms
- Possible solution: $e^{-\zeta r/a_0}$
- How to find ζ ? Will it be the same for all H-compounds?
- Can we use different types of functions, not only e^r ?
- Is it enough to use only 's' functions for H? Minimal basis (MB)
- How far we should go in 'spdfgh' ?
- Would it help to use several 's' functions?
- Is any limit here?
- Why everyone prefer small basis sets?

- $Ne^{-\zeta r/a_0}$
- Exhibits correct behaviour
- Hard to compute two-electron integrals
- Alternatives: Gaussian type orbitals $(e^{-\alpha r^2})$, plane waves
- A single GTO poorly represents an STO, but STO can be presented as a sum of several GTO
- GTO are very nice for computing integrals

Gaussian Basis Sets

•
$$\psi(\mathbf{r}) = \sum_{jim} C_{ij,lm} e^{-\alpha_i r^2} Y_{lm}(\mathbf{r})$$

Pros

- Compact, only small number is needed
- Integrals are analytic
- Very well studied and used
- Many existing basis sets
- Gaussian basis sets are used in 99% of molecular codes
- Cons
 - Over complete, linear dependence
 - Non-orthogonal, superposition error
 - Atom-centred

•
$$\psi(\mathbf{r}) = \sum_{lm} f_{lm}(r) Y_{lm}(\mathbf{r})$$

- Pros
 - Better completeness
- Cons
 - Integrals must be evaluated numerically
 - Harder to control accuracy
 - Used in few codes, e.g. SIESTA

Plane-waves basis sets

•
$$\psi_{i,k}(\mathbf{r}) = \sum_{\mathbf{G}}^{|\mathbf{G}| < G_{max}} c_{ik,\mathbf{G}} e^{i(\mathbf{k}+\mathbf{G})\mathbf{r}}$$

Pros

- Translational symmetry
- Complete and orthonormal
- Easy to evaluate integrals
- Not atom-centred
- Can be systematically improved by increasing cut-off parameter
- Can use efficient FFT algorithms.
- Cons
 - Huge number of basis coefficients
 - Not chemically intuitive (both in set up and in interpretation of results: even atomic charges are hard to compute!)
 - Very hard in description of core-electrons (has to be combined with pseudo-potentials)
 - Voids as expensive as atoms

- Muffin-Tin orbitals
- Augmented Plane Waves
- Splines
- Pseudo-potentials vs full electron basis sets

- Orbital shells. p_x , p_y , p_z differ only by angular part
- p_x , p_y , p_z vs p_{-1} , p_0 , p_{+1}
- $N_l(\alpha_k)r^l e^{-\alpha_k r^2} Y_{l,m}(\theta,\phi)$
- What about *d*-orbital? 6 Cartesian terms x^2 , xy, xz, y^2 , yz, and z^2 , but there are only five *d* components.
- Are you familiar with basis sets that use Cartesian terms?

- Can the basis set on one atom influence another?
- There are no orthogonality rules for basis functions on different centers.
- Basis sets which are too 'non orthogonal' will spread out to another centers.
- BSSE error $E_{AB} E_A E_B$ can not be used to compute the binding energy.
- BSSE correction compute A+B, A+X(only basis, no electrons), X(only basis, no electrons)+B
- BSSE quality is an attribute of the basis set.
- A similar effect not balanced basis sets.
- Q: Consider NaCl with a) very large and diffuse basis set on Na and very small on Cl. b) other way around. Which of these situations is worse?

- Optimize exponents. Note that you need to choose a proper Hamiltonian for optimization! Where will you use your basis set?
- DZ, DZP, TZ, TZP, QZP, 5ZP language.
- Test: how many functions are in the DZ basis set for water?
- These basis sets are 'coarse'. Expanding from DZ to TZ is a huge increase of the size of the basis set
- Uncontracted basis set: only exponents are optimized. Why it is not so popular?

- segmented contraction: each primitive function contributes to one contraction function
- 6-31G one contracted Gaussian composed of 6 primitives for inner shell (1s for O), and 3 contracted for valence shell (2s and 2p for O) and one extra diffuse primitive
- Example for O atom: exponents for 1s: 5484., 825., 188., 52., 16., 5. For 2s: 15., 3., 1. For 3s: 0.27.

All coefficients are separate.

• Dunning basis sets: CC- and AUG-CC-

- general contraction: all primitive functions contribute to all contraction functions, so exponents are the same for 1s, 2s, 3s etc.
- Coding of integrals with general contraction is more difficult
- It is possible to use a code. written for segmented contraction, but it will be slow..

- Jan Almlöf and Peter R. Taylor, 1990
- use several states in the calculation of contraction coefficients
- ANO-S and ANO-L basis sets. Recently also ANO-Orca etc.
- O.ANO-L.Widmark.14s9p4d3f.7s7p4d3f.
- all combinations inside the limit 7s7p4d3f are possible
- but there are also a nice labels, like O.ANO-L-VDZ = 3s2p
- average of several ${}^{3P}O$, ${}^{3P}O$ (with field), ${}^{1D}O$, ${}^{4S}O^+$, ${}^{2P}O^-$
- Can this basis set describe O^{-2} ? or O with a core hole in 1s?

- More smooth steps in the increase of basis set size
- Wide class of tasks, including excited states
- very small BSSE error
- A simple way to make transformation from one basis set to another, just by adding zeros: EXPBAS code can be used to make a nice starting point for calculations with a larger basis set

- Why (and how) is relativity important for chemistry?
- Dirac equation and 4 components.
- Approximate Relativistic Hamiltonians: DKH, DKH2, DKH(∞), X2C.
- A practical observation (if your chemistry includes elements not heavier than U): it is enough to use 2 components Hamiltonians.
- Another practical observation: if your chemistry has elements heavier that Cl, you must use relativistic Hamiltonians.
- Why do you need to use special basis sets for relativistic calculations?
- Mixing relativistic Hamiltonians with non relativistic basis set can give an error in order of 1 a.u. per heavy atom.

- ANO-RCC family is well tested and widely used basis sets for DKH2 Hamiltonian
- since ANO-RCC is rather expensive for light atoms use ANO-XS for C, O, H etc.
- Never mix ANO-RCC and ANO-L (or 6-31G)
- ANO-R uses more mathematically formal criteria for basis set quality, thus ANO-R-1, ANO-R-2, ANO-R-3 labes are used instead of VDZ, VTZ etc.
- ANO-R works with X2C Hamiltonian
- ANO-R is slightly smaller than ANO-RCC

D '

Dissociation of water dimer (experimental value: -21 kJ/mol						
	MB	VDZ	VDZP	VTZP	VQZP	V5ZP
BLYP	-25.9	-38.7	-21.6	-18.9	-17.6	-17.6
PBE	-31.1	-44.8	-25.3	-22.4	-21.4	-21.4
B3LYP	-24.8	-42.3	-22.8	-20.6	-19.2	-19.1
PBE0	-26.4	-46.5	-24.5	-21.9	-20.9	-20.8
M06	-26.9	-43.6	-22.9	-20.6	-19.3	-19.7
M06L	-28.1	-43.3	-23.6	-20.9	-19.7	-19.3

01 k / ٦J

- ANO-R doesn't include anions in the training set. . .
- R-1, R-2 gives totally wrong Electron Affinity for O
- For most of the applications TZ-like quality (4s3p2d) (R-2) for Oxygen atom in molecules is sufficient
- But if EA of Oxygen is needed, one should use something like 8s8p4d3f, which is absolutely impractical for 'ordinary' calculations.
- Home exercise: compute IP and EA of Oxygen with 6-31G basis set. What do you expect?
- Once again there is no such thing as 'universal' basis set

- good old days: core valence separation
- the main computational cost is in the virtual orbitals!
- Can one do calculation with 2000 basis functions?
- Is the answer different for HF, DFT, CASSCF, CASPT2, NEWPT2?

- Population matrix PS
- Quantities constructed with a use of PS: atomic charges, bond orders, covalence, full valence
- Spacial density (and orbitals)

- The size of conventional integrals is enormous
- Recomputing the integrals is possible but still expensive
- Semi-direct schemes
- Packing of integrals
- indexes of integrals
- Game changing: Matrix transformations: Cholesky decomposition and Resultion of identity

CD, RI, RICD

- Cholesky decomposition
 - Idea: compact form of the matrix. More efforts to compute, but more simple way to reduce the small components
 - Implemented in several codes: Dalton, Molcas
 - Controllable decomposition stops by a threshold
 - Note, that it might take time
 - Note, it might break the symmetry
- Resolution of identity
 - Auxiliary basis
 - First implemented in Turbomole, now it is everywhere
 - Depends on the method
- RICD option in Molcas
 - Use Cholesky decomposition to generate on fly RI
 - Different flavours
 - Atomic RICD (default) is symmetric and fast

- A decision about the basis set is important
- You need a basis set, which reproduces the property you study!
- Never publish results computed with DZ quality of the basis set (or at least, make a warning)
- Never publish results for excited states computed with 6-31G family of the basis sets (or include a warning)
- Try to find a proper balance between quality and computational resources
- It might happen that you need re-optimize a basis set for your particular case
- Finally, note that the integrals, computed with the basis set, are usually transformed by Cholesky decomposition (CD) or Resolution of Identity (RI).