

Optimisation of Basis Sets and Pseudopotentials

Sanliang Ling

University College London

4th CP2K Tutorial, 31st August – 4th September 2015, Zurich

Electronic structure methods in CP2K

GPW: Gaussian and plane waves method

- Goedecker-Teter-Hutter pseudopotentials
- Gaussian basis sets for valence electrons

GAPW: Gaussian and augmented plane waves method

- all electron calculations

LCAO

LCAO: Linear Combination of Atomic Orbitals

$$\phi_i = \sum_{\alpha}^{M_{\text{basis}}} c_{\alpha i} \chi_{\alpha}$$

molecular orbital (MO) **MO coefficient (unknown)** **atomic orbital (basis function)**

The diagram shows the equation $\phi_i = \sum_{\alpha}^{M_{\text{basis}}} c_{\alpha i} \chi_{\alpha}$ with three red arrows pointing from the terms to their definitions below. The first arrow points from ϕ_i to 'molecular orbital (MO)'. The second arrow points from $c_{\alpha i}$ to 'MO coefficient (unknown)'. The third arrow points from χ_{α} to 'atomic orbital (basis function)'. All definitions are in blue text.

Gaussian type orbitals (GTOs)

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

normalisation constant

exponent:
width of orbital

sum of l_x, l_y, l_z determines type of orbital:
0 for s, 1 for p, 2 for d, 3 for f, etc

Contracted basis sets

$$\chi(\text{CGTO}) = \sum_i^k a_i \chi_i(\text{PGTO})$$

**contraction coefficient
(to be optimised)**

Polarisation function

- **basis functions with higher angular momentum (than the valence orbital)**
- **first shell of polarisation functions are most important**
- ***p*-function for H-Be, *d*-function for B-Ca, etc**
- **adds additional flexibility to the basis set, provides better descriptions to bonding**

Diffuse function

- **basis function with small exponent**
- **better representation of the “tail” of the wavefunction**
- **important for loosely bound electrons (anions or excited state) and molecules in the gas phase**

All-electron basis set for GAPW calculations

- Pople style basis sets (e.g. 6-31G*, etc)
- Correlation consistent basis sets (aug-cc-pVDZ, etc)
- and more

see \$CP2K/cp2k/data, 'ALL_BASIS_SETS' and 'EMSL_BASIS_SETS'
 additional all-electron basis sets can be found from EMSL Basis Set
 Exchange, see <https://bse.pnl.gov/bse/portal>

Potential needs to be defined in &KIND section for GAPW calculations, see
 \$CP2K/cp2k/data/POTENTIAL, choose "ALL" potential

Basis set for GPW calculations

- **MOLOPT basis sets: basis sets optimised from molecular calculations, see ‘BASIS_MOLOPT’**
- **‘DZVP-MOLOPT-SR-GTH’ for solids (‘SR’ denotes shorter range, i.e. less and thus less diffuse primitives)**
- **always check the basis set convergence (DZVP/TZVP/...)**
- **do not use SZV for production run**
- **more basis sets for GTH pseudos can be found in ‘BASIS_ZIJLSTRA’ and ‘GTH_BASIS_SETS’**
- **all basis set files can be found in \$CP2K/cp2k/data**

Basis set construction

- **trade-off between computational cost and accuracy**
- **route for systematic improvements**
(SZV/DZVP/TZVP/TZV2P/...)
- **same basis set should perform in various chemical environments, e.g. from isolated molecules to solids**
- **lead to well conditioned overlap matrices (suitable for linear scaling calculations)**
- **condition number: ratio of the largest to smallest eigenvalue of the overlap matrix**

MOLOPT basis set

		CP2K	All-electron (Gaussian/NWCHEM)
H-Rn	}	SZV	STO-3G
		DZVP	6-31G*
limited availability	}	TZVP	6-311G*
		TZV2P	6-311G(2df, 2pd)

SZV: single-zeta valence, i.e. one contracted function per orbital

DZVP: double-zeta valence, i.e. two contracted functions per orbital plus one set of polarisation functions with $l = l_{\max} + 1$

TZVP/TZV2P: triple-zeta valence, i.e. three contracted functions per orbital plus one/two set of polarisation functions with $l = l_{\max} + 1$

MOLOPT basis set format

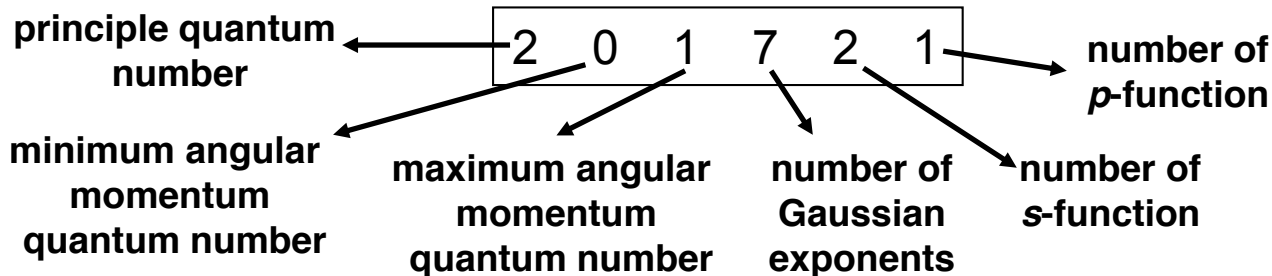
element basis set name number of valence electrons in pseudo

H	DZVP -MOLOPT-GTH DZVP -MOLOPT-GTH-q1		
1	number of CGTO	contraction coefficients	
2 0 1 7 2 1			
11.478000339908	0.0249162432	-0.0125124214	0.0245109182
3.700758562763	0.0798254900	-0.0564490711	0.0581407941
1.446884268432	0.1288626753	0.0112426847	0.4447094985
0.716814589696	0.3794488946	-0.4185875483	0.6462079731
0.247918564176	0.3245524326	0.5903632167	0.8033850182
0.066918004004	0.0371481214	0.4387031330	0.8929712087
0.021708243634	-0.0011251955	-0.0596931713	0.1201013165

Gaussian exponents

s-function

p-function



Basis set optimisation

- **number of Gaussian exponents (to be determined before optimisation)**
- **number of basis functions per angular momentum**
- **choice of training molecules (transferability)**
- **strategy of basis set optimisation, e.g. whether or not to optimise different basis sets concurrently**
- **weight of condition number in optimisation**

Choice of training molecules

- **small molecules formed with different elements and with different coordination environments**
- **preferably with only two elements (including the target element) in each molecule**
- **a good source of small molecules (with optimised geometries) can be found in the Supporting Information of “Ahlrachs et al., Phys. Chem. Chem. Phys., 7, 3297 (2005)”**

Basis set optimisation: MOLOPT

$$\Omega(\{\alpha_i, c_i\}) = \sum_B \sum_M (E_{\text{tot}}^{B,M}(\{\alpha_i, c_i\}) + \gamma \ln \kappa^{B,M}(\{\alpha_i, c_i\}))$$

The diagram illustrates the components of the MOLOPT objective function. The equation is annotated with arrows pointing to labels:

- exponents** (red text) points to α_i .
- contraction coefficients** (red text) points to c_i .
- objective function** (blue text) points to $\Omega(\{\alpha_i, c_i\})$.
- basis sets to be optimised** (blue text) points to the summation index B .
- training molecules** (blue text) points to the summation index M .
- total energy** (blue text) points to $E_{\text{tot}}^{B,M}(\{\alpha_i, c_i\})$.
- weight** (blue text) points to γ .
- condition number** (blue text) points to $\kappa^{B,M}(\{\alpha_i, c_i\})$.

Basis optimisation with OPTIMIZE_BASIS

Choosing a reference (complete) basis

Performing accurate molecular calculations with ref. basis

Choosing a form of the basis to be fitted

Minimizing the objective function

$$\Omega(\{\alpha_i, c_j\}) = \sum_B \sum_M (\Delta\rho^{B,M}(\{\alpha_i, c_j\}) + \gamma \ln \kappa^{B,M}(\{\alpha_i, c_j\}))$$

Basis optimisation with OPTIMIZE_BASIS

- Reference (Complete) basis set
 - check **GTH-def2-QZVP** and **aug-GTH-def2-QZVP** included in `$CP2K/cp2k/data/BASIS_ADMM`
 - generate uncontracted basis sets with the **ATOM** code (see Marcella's slides and examples in `$CP2K/cp2k/tests/ATOM`)
- Molecular calculations
 - use reference basis sets for all elements
 - avoid homonuclear diatomic molecules
 - use equilibrium geometry (i.e. `GEO_OPT`)

Generate uncontracted basis set with ATOM

```

&GLOBAL
  PROJECT Na
  PROGRAM_NAME ATOM
&END GLOBAL
&ATOM
  ELEMENT Na
  RUN_TYPE BASIS_OPTIMIZATION
  ELECTRON_CONFIGURATION CORE 2s2 2p6 3s1
  CORE 1s2
  MAX_ANGULAR_MOMENTUM 1
&METHOD
  METHOD_TYPE KOHN-SHAM
  &XC
    &XC_FUNCTIONAL PBE
    &END XC_FUNCTIONAL
  &END XC
&END METHOD
&OPTIMIZATION
  EPS_SCF 1.e-8
&END OPTIMIZATION
&PP_BASIS
  NUM_GTO 6 6
  S_EXPONENTS 7.92602574 5.92602574 1.59655262 0.71279902 0.28969807 4.00675308
  P_EXPONENTS 7.92602574 5.92602574 1.59655262 0.71279902 0.28969807 4.00675308
&END PP_BASIS
&POTENTIAL
  PSEUDO_TYPE GTH
  POTENTIAL_FILE_NAME POTENTIAL
  POTENTIAL_NAME GTH-PBE-q9
&END POTENTIAL
&POWELL
  ACCURACY 1.e-8
  STEP_SIZE 1.0
&END POWELL
&END ATOM

```

Generate uncontracted basis set with ATOM

```

Orbital energies  State      L      Occupation  Energy[a.u.]      Energy[eV]
                1         0         2.000      -2.092687         -56.944917
                2         0         1.000      -0.098547          -2.681599
                1         1         6.000      -1.047513         -28.504279

POWELL| Number of function evaluations                273
POWELL| Final value of function                      -47.1609800227

Optimized Basis

***** Uncontracted Gaussian Type Orbitals *****

s Exponents:
                1         3.37371675
                2         1.19458113
                3        21.63024049
                4         0.39968861
                5         8.84460076
                6         0.04267359

p Exponents:
                1         3.37371675
                2         1.19458113
                3        21.63024049
                4         0.39968861
                5         8.84460076
                6         0.04267359

*****

```

```

[sling@klb231 Na]$ grep "Final" Na.out.*
Na.out.4: POWELL| Final value of function          -47.0387118701
Na.out.5: POWELL| Final value of function          -47.1463995925
Na.out.6: POWELL| Final value of function          -47.1609800227
Na.out.7: POWELL| Final value of function          -47.1637208082
Na.out.8: POWELL| Final value of function          -47.1647939416
Na.out.9: POWELL| Final value of function          -47.1649267320

```

Generate uncontracted basis set with ATOM

```

Na CBS
8
2 0 2 1 1 1 1
    23.51400109 1.0 1.0 1.0
2 0 2 1 1 1 1
    11.54276369 1.0 1.0 1.0
2 0 2 1 1 1 1
    4.98513381 1.0 1.0 1.0
2 0 2 1 1 1 1
    2.06401264 1.0 1.0 1.0
2 0 2 1 1 1 1
    0.83224580 1.0 1.0 1.0
2 0 2 1 1 1 1
    0.31900235 1.0 1.0 1.0
2 0 2 1 1 1 1
    0.06577574 1.0 1.0 1.0
2 0 2 1 1 1 1
    0.02386738 1.0 1.0 1.0
    
```

GTH-def2-QZVP basis set

Go to <https://bse.pnl.gov/bse/portal>,
select 'H' element and 'Def2-QZVP' basis
set, use 'Gaussian 94' format:

H GTH-def2-QZVP

```

12
2 0 0 7 1
  11.478000339908 0.024916243200
  3.700758562763 0.079825490000
  1.446884268432 0.128862675300
  0.716814589696 0.379448894600
  0.247918564176 0.324552432600
  0.066918004004 0.037148121400
  0.021708243634 -0.001125195500
1 0 0 1 1
  6.50959430 1.00000000
1 0 0 1 1
  1.84124550 1.00000000
1 0 0 1 1
  0.59853725 1.00000000
1 0 0 1 1
  0.21397624 1.00000000
1 0 0 1 1
  0.08031629 1.00000000
1 1 1 1 1
  2.29200000 1.00000000
1 1 1 1 1
  0.83800000 1.00000000
1 1 1 1 1
  0.29200000 1.00000000
1 2 2 1 1
  2.06200000 1.00000000
1 2 2 1 1
  0.66200000 1.00000000
1 3 3 1 1
  1.39700000 1.00000000
  
```

SVZ

```

H 0
S 4 1.00
190.6916900 0.70815167E-03
28.6055320 0.54678827E-02
  6.5095943 0.27966605E-01
  1.8412455 0.10764538
S 1 1.00
  0.59853725 1.00000000
S 1 1.00
  0.21397624 1.00000000
S 1 1.00
  0.80316286E-01 1.00000000
P 1 1.00
  2.29200000 1.00000000
P 1 1.00
  0.83800000 1.00000000
P 1 1.00
  0.29200000 1.00000000
D 1 1.00
  2.06200000 1.00000000
D 1 1.00
  0.66200000 1.00000000
F 1 1.00
  1.39700000 1.00000000
  
```

(use exponents between 0.05~20 only)

Input Structure: OPTIMIZE_BASIS

```

&GLOBAL
  PROJECT optbas
  PROGRAM_NAME OPTIMIZE_BASIS
  PRINT_LEVEL HIGH
&END GLOBAL
&OPTIMIZE_BASIS
  BASIS_TEMPLATE_FILE BASIS_SET_TEMPLATE
  BASIS_WORK_FILE WORK_BASIS_STRUCTURE
  BASIS_OUTPUT_FILE Ti_FIT10
#  USE_CONDITION_NUMBER Y
#  CONDITION_WEIGHT 0.0005
  WRITE_FREQUENCY 10
&OPTIMIZATION
  MAX_FUN 50000
&END OPTIMIZATION
...
&TRAINING_FILES
  DIRECTORY ../ticl4
  INPUT_FILE_NAME ticl4.inp
&END TRAINING_FILES
...
&FIT_KIND Ti
  BASIS_SET FIT10
  INITIAL_DEGREES_OF_FREEDOM EXPONENTS
  &CONSTRAIN_EXPONENTS
    BOUNDARIES 0.1 20
    USE_EXP -1 -1
  &END CONSTRAIN_EXPONENTS
&END FIT_KIND
&END OPTIMIZE_BASIS
  
```



Ti FIT10			
10			
1 0 0 1 1	0.10001966	1.00000000	} s-functions
1 0 0 1 1	1.06186104	1.00000000	
1 0 0 1 1	0.40963197	1.00000000	
1 0 0 1 1	4.39901876	1.00000000	
1 1 1 1 1	0.52985233	1.00000000	} p-functions
1 1 1 1 1	1.57394040	1.00000000	
1 1 1 1 1	11.83843422	1.00000000	
1 2 2 1 1	0.25675246	1.00000000	} d-functions
1 2 2 1 1	1.02358115	1.00000000	
1 2 2 1 1	4.21355677	1.00000000	

(Ti electron configuration: [Ne] 3s² 3p⁶ 4s² 3d²)

Basis optimisation with OPTIMIZE_BASIS

```

BASOPT| Information at iteration number:      390
BASOPT| Training set | Combination | Rho difference | Condition num. | Time
BASOPT| -----
BASOPT|      1      |      1      | 0.36863360E-02 | 0.46570176E+02 | 2.804
BASOPT|      2      |      1      | 0.12483294E-01 | 0.16639694E+03 | 1.371
BASOPT|      3      |      1      | 0.39469184E-02 | 0.97623037E+02 | 1.586
BASOPT|      4      |      1      | 0.45506749E-02 | 0.14458737E+03 | 2.067
BASOPT| -----
BASOPT| Total residuum value: -.14866668E+02

BASOPT| Information at iteration number:      400
BASOPT| Training set | Combination | Rho difference | Condition num. | Time
BASOPT| -----
BASOPT|      1      |      1      | 0.36863752E-02 | 0.46570547E+02 | 2.876
BASOPT|      2      |      1      | 0.12483327E-01 | 0.16639794E+03 | 1.406
BASOPT|      3      |      1      | 0.39468986E-02 | 0.97623212E+02 | 1.627
BASOPT|      4      |      1      | 0.45506441E-02 | 0.14458778E+03 | 2.120
BASOPT| -----
BASOPT| Total residuum value: -.14866668E+02

-----
-
-
-
-
-----
COUNTER                                CPU                                GPU                                GPU%

```

Basis set superposition error

- **MOLOPT basis sets are incomplete**
- **BSSE correction using the Boys and Bernardi counterpoise correction scheme**

Boys & Bernardi, Mol. Phys., 19, 553 (1970)

- **useful for binding energy calculations, etc**
- **use larger basis sets to reduce BSSE**

```

&GLOBAL
  PROJECT_NAME project
  RUN_TYPE BSSE
&END GLOBAL
...
&FORCE_EVAL
...
  &BSSE
    &FRAGMENT
      LIST 1..272
    &END FRAGMENT
    &FRAGMENT
      LIST 273..368
    &END FRAGMENT
  &END BSSE
...
  SCF_GUESS ATOMIC
...
  &KIND H_ghost
    BASIS_SET DZVP-MOLOPT-SR-GTH
    GHOST
  &END KIND
...
  
```


Pseudopotentials

Goedecker-Teter-Hutter (GTH) pseudopotentials

$$\begin{aligned}
 V_{\text{loc}}^{\text{PP}}(r) = & -\frac{Z_{\text{eff}}}{r} \text{erf}(\alpha^{\text{PP}} r) \quad \text{long-ranged term} \\
 & + \sum_{i=1}^4 C_i^{\text{PP}} \left(\sqrt{2}\alpha^{\text{PP}} r\right)^{2i-2} \exp\left[-(\alpha^{\text{PP}} r)^2\right] \quad \text{short-ranged term}
 \end{aligned}$$

ionic charge error function $\alpha^{\text{PP}} = \frac{1}{\sqrt{2}r_{\text{loc}}^{\text{PP}}}$
 Local part coefficients

r_{loc} : range of Gaussian ionic charge distribution

Pseudopotentials

$$V_{\text{nl}}^{\text{PP}}(\mathbf{r}, \mathbf{r}') = \sum_{lm} \sum_{ij} \langle \mathbf{r} | p_i^{lm} \rangle h_{ij}^l \langle p_j^{lm} | \mathbf{r}' \rangle$$

Non-local part \downarrow **coefficients**

Gaussian-type projectors

$$\langle \mathbf{r} | p_i^{lm} \rangle = N_i^l Y^{lm}(\hat{r}) r^{l+2i-2} \exp \left[-\frac{1}{2} \left(\frac{r}{r_l} \right)^2 \right]$$

normalisation constant **spherical harmonics** **radius**

GTH pseudopotential format

Element Name Number of valence electrons

	Ti	GTH-PBE-q12	GTH-PBE	
	4	6	2	→ Number of valence electrons in each shell (s/p/d)
r_{loc}	0.38000000	2	8.71144218	-0.70028677
	3	→ Number of non-local projectors		
r_s	0.33777078	2	2.57526386	3.69297065
				-4.76760461
r_p	0.24253135	2	-4.63054123	8.87087502
				-10.49616087
r_d	0.24331694	1	-9.40665268	

number of potential functions

coefficients

GTH pseudopotential

- **LDA (PADE): H-Rn (including lanthanides)**
- **PBE: H-Rn (excluding lanthanides)**
- **PBEsol: H-Kr (plus a few selected)**
- **BP: H-Kr (plus a few selected)**
- **HCTH: a few selected elements**
- **Non-linear core corrected (NLCC) pseudopotentials:
a few selected elements**

All pseudopotentials can be found in `$CP2K/cp2k/data`, see 'POTENTIAL', 'GTH_POTENTIALS' and 'NLCC_POTENTIALS'

Pseudopotentials optimisation

All-electron calculation using a chosen DFT functional



Choosing a form of the pseudopotential to be fitted



Minimising the differences between eigenvalues and charges within an atomic sphere of the all-electron atom and the pseudo atom



Quality check

Pseudopotentials optimisation with ATOM

Example: optimise pseudo-potential of O using PBE0

```
&GLOBAL
  PROGRAM_NAME ATOM
&END GLOBAL
&ATOM
  ELEMENT O
  RUN_TYPE PSEUDOPOTENTIAL_OPTIMIZATION
  ELECTRON_CONFIGURATION [He] 2s2 2p4
  CORE [He]
  MAX_ANGULAR_MOMENTUM 2
  COULOMB_INTEGRALS ANALYTIC
  EXCHANGE_INTEGRALS ANALYTIC
&METHOD
  METHOD_TYPE KOHN-SHAM
  RELATIVISTIC DKH(2)
&XC
  &XC_FUNCTIONAL PBE0
  &END XC_FUNCTIONAL
&END XC
&END METHOD
```

```
&OPTIMIZATION
  EPS_SCF 1.e-10
&END OPTIMIZATION
&PRINT
  &BASIS_SET
  &END
&END
&AE_BASIS
  BASIS_TYPE GEOMETRICAL_GTO
&END AE_BASIS
&PP_BASIS
  BASIS_TYPE GEOMETRICAL_GTO
&END PP_BASIS
&POTENTIAL
  PSEUDO_TYPE GTH
  POTENTIAL_FILE_NAME POTENTIAL
  POTENTIAL_NAME GTH-PBE-q6
&END POTENTIAL
&POWELL
  ACCURACY 1.e-10
  STEP_SIZE 0.5
  WEIGHT_PSIRO 0.1
&END POWELL
&END ATOM
```

Pseudopotentials optimisation with ATOM

```

POWELL| Final errors of target values

Reference configuration          1          Method number          1
L   N   Occupation             Eigenvalue [eV]          dE [eV]          dCharge
0   1   2.00                   -26.0500594476          0.000072[ 0]     -0.000034[ 3]
0   2   0.00                    5.7901112679           -0.001447[ 0]     0.000149[ 0]
0   3   0.00                   15.4862366222          -0.004456[ 0]     0.000417[ 0]
1   1   4.00                   -8.7389403636           0.000114[ 1]     0.000195[96]
1   2   0.00                    8.4179458225            0.000184[ 0]     -0.000050[ 0]
1   3   0.00                   17.3085598983           0.001142[ 0]     -0.000133[ 0]
2   1   0.00                   10.9106631896           0.000775[ 0]     -0.000018[ 0]
s-states N= 1                   Wavefunction at r=0:     -0.000000[ 0]
s-states N= 2                   Wavefunction at r=0:     -0.024849[ 0]
s-states N= 3                   Wavefunction at r=0:     -0.045189[ 0]

```

Pseudopotentials optimisation with the standalone ATOM code

- check 'README_quick_GTH' in \$CP2K/potentials,
more details in Dr Matthias Krack's slides

Matthias Krack, 1st CP2K Tutorial, Zurich, 2009

- more options for quality check

A few final remarks ...

- **read the original references on basis sets and pseudopotentials before making your own basis sets and pseudopotentials**
- **do plenty of tests on optimised basis sets and pseudopotentials and compare the results with reference values before production run**
- **All data files used by CP2K can be downloaded directly from <http://sourceforge.net/p/cp2k/code/HEAD/tree/trunk/cp2k/data/>**

Further Reading

Accuracy and Efficiency (by Dr Matthias Krack)

http://www.cecim.org/upload/talk/presentation_3002.pdf

Basis Sets and Pseudo-Potentials (by Dr Fawzi Mohamed)

http://www.cecim.org/upload/talk/presentation_2994.pdf