**RexPoN Optimizer Package**

A python package that calls LAMMPS as a library and uses scipy optimizer modules.

1. **Setting up the Environment**

Include the following lines into your .cshrc file to make use of RexPoN optimizer package.

# begin vasek setup

# note: overriding the default python breaks SCREAM

# platform dependet:

setenv biosoft /project/Biogroup/Software

set ARCH = `uname -m`

if ($ARCH == x86\_64) then

setenv VASEKINSTALL $biosoft/lammps/64

set redhatversion = `cat /etc/redhat-release | cut -d" " -f3 | cut -d"." -f1`

if ( $redhatversion =~ [0-9] ) then

if ( 6 <= $redhatversion ) then

set path = ( /exec/python/pythonEPD-7.0-2-rh5-x86\_64/bin $path )

else

set path = ( /exec/python/pythonEPD-7.0-2-rh3-x86\_64/bin $path )

endif

else

set path = ( /exec/python/pythonEPD-7.0-2-rh3-x86\_64/bin $path )

endif

else

setenv VASEKINSTALL $biosoft/lammps/32

set path = ( /exec/python/pythonEPD-7.0-2-rh3-x86/bin $path )

endif

setenv VASEK $biosoft/vasek

setenv VASEKDATA $biosoft/lammps

setenv LD\_LIBRARY\_PATH ${LD\_LIBRARY\_PATH}:$VASEKINSTALL/lib

setenv PYTHONPATH ${VASEK}:${VASEKINSTALL}/lib:${PYTHONPATH}

set path = ( $VASEK/bin $VASEKINSTALL/bin $path )

# end vasek setup

setenv PYTHONPATH ${PYTHONPATH}:/net/hulk/home2/naserifa/qeq/python/python/sharedlib\_rexpon:/net/hulk/home2/naserifa/qeq/LAMMPS-Optimizer-RexPoN

setenv LD\_LIBRARY\_PATH ${LD\_LIBRARY\_PATH}:/net/hulk/home2/naserifa/qeq/python/python/sharedlib\_rexpon

# python with new scipy

alias pythons '/project/exec/schrodinger\_2015/utilities/python'

# RexPoN optimizer

alias lmpopt 'pythons /net/hulk/home2/naserifa/qeq/LAMMPS-Optimizer/bin/main.py -f ffield -p params -t trainset'

1. **Input Files**

* geo
* ffield
* trainset
* params
  1. *geo*

Convert your .xyz files to lammps data files using the following script:

> python /ul/naserifa/bin/lammps/rexpon/xyz2rexpon.py input.xyz ffield data\_lammps\_output

where ffield is the force field that you are going to use during optimization (should contain correct atomic types). Provide a reasonable “output” name appended to “data\_lammps\_”. The output name will be used by trainset to perform optimization.

* 1. *ffield*

The ffield file includes the initial set of parameters plus some information for the optimizer. It is better to start from a sample ffield and modify it for your system. I have included an example for N and Fe system.

* 1. *trainset*

This is your training set which includes the reference values for the “data\_lammps\_names” files. For example:

ENERGY

1.0 + N2Fe\_2.0Ang/1 - N2Fe\_10.Ang/1 -10.0

1.0 + N2Fe\_3.0Ang/1 - N2Fe\_10.Ang/1 -8.0

ENDENERGY

Here, we are doing energy optimization. The first line means take geometry “data\_lammps\_N2Fe\_2.0Ang” and subtract its energy from “data\_lammps\_ N2Fe\_10.Ang” and compare the result with the reference energy -10.0 kcal/mol. First number “1.0” is the weight for optimization. For better accuracy use smaller value for weight. Error = [(Eref-Erexpon)/weight]^2

* 1. *params*

This tells the program which parameters of ffield file to optimize. For example:

BOUNDS # use parameters range

1 BONDS 1 3 0.00 200.0

1 BONDS 1 4 0.00 200.0

1 BONDS 1 5 0.00 200.0

2 VDW 1 6 0.0 10.0

3 VDW 1 10 0.0 10.0

Each line starts with an arbitrary number, name of the section, line number in that section, parameter number in the line, lower, and upper bounds for the parameter range

1. **Structure of the ffield file**

We are still developing RexPoN and that is why the force file has many parameters which lots of them are not used by the program or have been added for other systems. Below, I show you the parameters for bond order, bond energy, and vdW energy terms. Contact me if you want to know the location of parameters for other energy terms in the force field file.

* 1. **Covalent Bond Energy (Ebond)**

**The Bond order (BO) expression.** To describe covalent bond, we used the bond order (BO) concept. Here we use a sigmoid function as in ReaxFF to define the bond distance to bond order relation,

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

where, *pbo1* and *pbo2* are parameters, *Re* is equilibrium bond distance, and *r* is the bond distance between atom *i* and atom *j*. We choose the BO=1.0 at *Re* which results in *pbo1*=2.3026. This allows the formal bond order to approach 1.1 for very small values of *r*. Therefore, the only free parameter that needs to be determined is *pbo2*. To make an initial good guess for *pbo2* we choose the BO=0.5 at the inflection point distance (RH/2) of the bond energy curve. The QM bond energy curves (see below) are used to obtain *pbo2* parameter.

We define the bond energy (*Ebond*) to match the bond dissociation curves. Thus, given *Evdw* and *Eelect*, we define *EQ2* and *Ebond* as

|  |  |  |
| --- | --- | --- |
|  |  | (2) |
|  |  | (3) |

where *EQM* is the total QM energy due to stretching bond between atom *i* and *j*. *k* and *l* are atom indices (*k,l*≠*i,j*). *EQ2* is the net two-body QM energy between atom *i* and *j*, which is obtained bysubtracting from *EQM* the *EvdW* between all atom pairs except for *i* and *j* and also subtracting the total *Eelect* of the system. Therefore, based on *EvdW* and *Eelectro* which are already established, the *Ebond* is defined automatically to match the QM bonding energy curves.

The *EQ2* curve is fitted to an extended-Rydberg type function of the form

|  |  |  |
| --- | --- | --- |
|  |  | (4) |

where *D0* and a are the parameters to be determined. b0=4 and c0=1

* 1. **Van der Waals (Evdw)**

The vdW energy is described by

where DvdW and RvdW have in kcal/mol and **Å . L is the curvature related parameter. The only parameters that you need to specify are *Dvdw, RvdW*, and *L*. The rest of the parameters are defined based on the universal curve. It is not recommended but you have the option to optimize the following parameters too.**

|  |  |
| --- | --- |
| **β** | **1.00348500** |
| **α0** | **1.00000000** |
| **α1** | **1.02009000** |
| **α2** | **0.01678480** |
| **α3** | **0.00327294** |
| **α4** | **0.00365706** |
| **α5** | **0.00106613** |

Table 1. The location of parameters in the ffield file.

|  |  |  |  |
| --- | --- | --- | --- |
| **Parameter** | ***Section*** | ***number*** |  |
| *Re* | VDW | 6 |  |
| *pbo1* | BONDS | 15 |  |
| *pbo2* | BONDS | 16 |  |
| *D0* | BONDS | 3 |  |
| *a0* | BONDS | 6 |  |
| *c0* | BONDS | 14 |  |
| *b0* | BONDS | 11 |  |
| ***Dvdw*** | VDW | 3 |  |
| ***RvdW*** | VDW | 4 |  |
| ***L*** | VDW | 5 |  |

1. **Running the program**

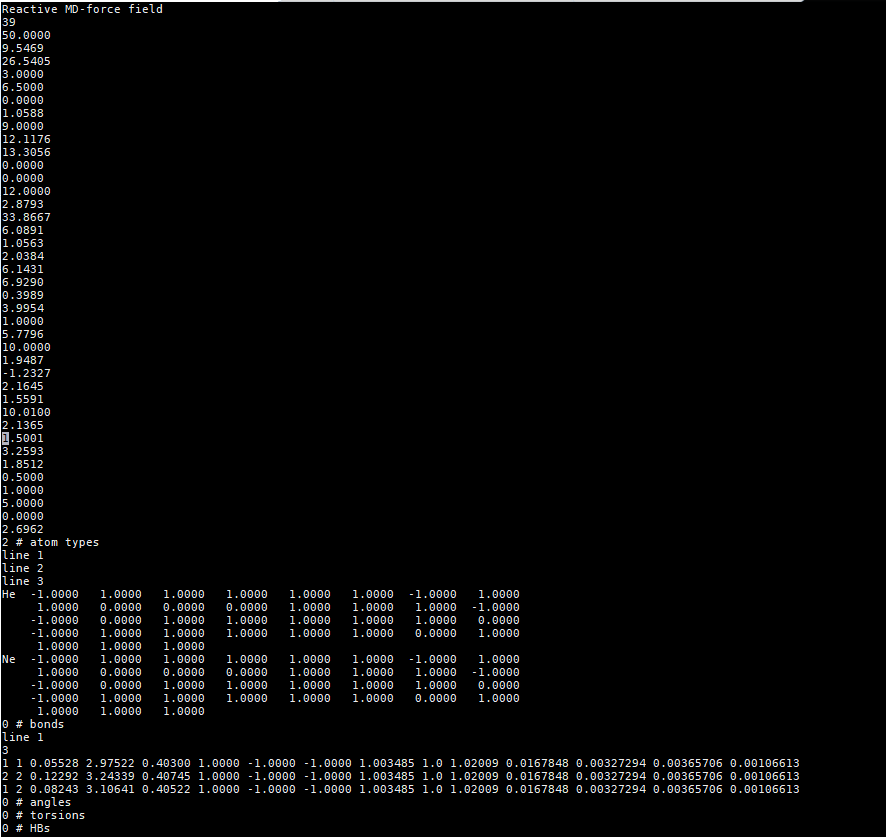
**Submit the job using**

**> lmpopt**

Or

> /project/exec/schrodinger\_2015/utilities/python /net/hulk/home2/naserifa/qeq/LAMMPS-Optimizer/bin/main.py -f ffield -p params -t trainset

Example 1) HeNe ffield



Both He and Ne have the 1st parameter < 0, so no need to define bonds, angles, or torsions parameters!

See example of water for the format of bond params in bond section

Not Used!

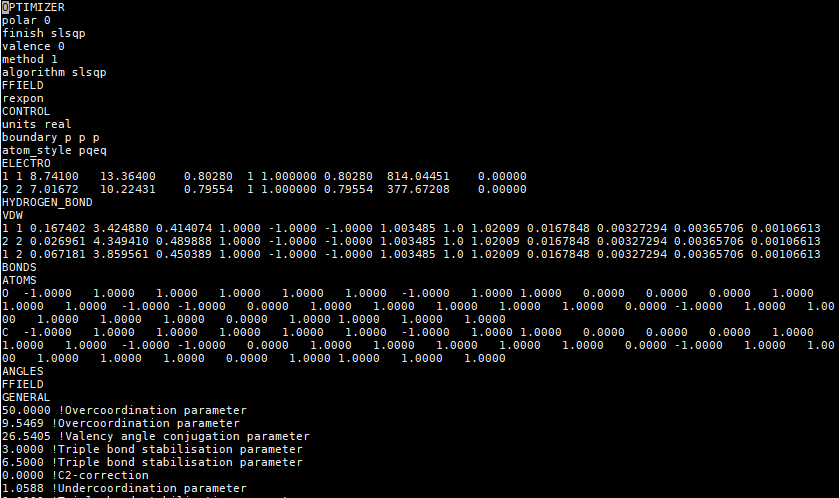
Setting this parameter to < 0.0 means BO=0.0 and all valence energy and force terms are ignored (i.e. Ebond=Eangle=Etorsion=0.0 and alse forces=0.0).

General Section

Other parameters of universal nonbond curve. These parameters define the shape of the universal curve and are not supposed to be changed even though you have the option too!

De, Re, and L parameters of universal nonbond curve.

Example 2) Sample of input ffield file for the optimizer:



Copy the initial parameters from your ffield.RexPoN to each section but put all parameters in one single line

Use this format in the params file to refere to each parameter during optimization:

*Integer, Section Name, Row Number in the Section, Parameter Number in the Row, lower bound, upper bound*

For example, if you want to optimize these 3 parameters:

BOUNDS

1 VDW 3 3 0.0 10.0

2 VDW 3 4 0.0 10.0

3 VDW 3 5 0.0 10.0