Genetic Algorithm Potential Fitting Manual

OUTLINE

I. Compiling The Code

1) Build LAMMPS as a shared library. Go into your LAMMPS /src directory and do:

```
make mode=shlib ompi_g++
```

2) Put the following files into the directory that you wish to run the code:
   a) liblammps.so – Shared object file obtained from step 1
   b) liblammps_ompi_g++.so – Shared object file obtained from step 1
   c) ga.cpp – This is the genetic algorithm code
   d) lammps.h, input.h, atom.h, pointers.h, lmptype.h, and library.h – These are found in the LAMMPS /src directory

3) Set your library path with:

```
export LD_LIBRARY_PATH=/path/to/directory:$LD_LIBRARY_PATH
```

This tells your system where to find the shared libraries at runtime of the code.

4) Compile with:

```
mpiCC -std=c++11 -Wno-write-strings ga.cpp -o ga liblammps.so
```

5) It can be convenient to put the commands from steps 3 and 4 into a shell script, instead of typing them all the time.

II. Running The Code

1) Make sure you have the following required files in the directory:
   a) INPUT – Used to change settings of the algorithm
   b) EXTRAINDS – Optional solutions to start fitting from (can be empty)
   c) TEMPPOT – Potential file used for testing individuals
   d) REFPOT – Potential file used as a reference
   e) CONFIGS – Contains atomic positions, forces, and cohesive energies
2) Edit INPUT according to your needs.
3) Execute with:

```
./ga
```

III. Input Files

1) INPUT

This file contains 4 main sections as follows:

a) Genetic algorithm settings

b) LAMMPS commands to prepare the potential evaluation. This includes setting up the box, declaring atomic masses, creating atoms, atom styles, units, etc..

c) LAMMPS commands for non-tabulated parameters.

d) LAMMPS commands for tabulated parameters.

Here is an example of the file for fitting 2 atom types:

```
THIS LINE IS IGNORED
  generations: 2
  extrainds: 0
  population_size: 100
  mutation_rate: 0.006
  elite_percentage: 0.5
  tabulated_parameters: 10
  nontabulated_parameters: 2
  M: 25
  N: 64
  force_weight: 1
  energy_weight: 1
  stress_weight: 1
  intervals: 0.001 10 0.001 100 0.001 100 -1 1 0.001 100 0.001 100 0.001 10 0.001 5000 0.0001 10 0.0001 5000 0 1 -5 5
  tabulated_parameter_lines: 17 17 17 17 17 17 17 17 17 17
  tabulated_parameter_columns: 3 4 5 6 7 8 9 10 13 14
  decimal_place_accuracies: 4 4 4 4 4 4 4 4 4 4
  symmetric_lines: 21 25
  symmetric_coeffs: As Ga Ga As Ga As
----------------------------------------LAMMPS SETUP COMMANDS
  log none
  neighbor 1.0 multi
  boundary p p p
  units metal
  atom_style hybrid atomic charge
  atom_modify map array
  ### ATOM/LATTICE/BOX CREATION MUST BE DONE IN THE FOLLOWING ORDER!
  ### Define the first basis
  lattice custom 5.469 basis 0 0 0 basis 0 0.5 0.5 basis 0.5 0 0.5 basis 0.5 0.5 0
  ### Declare the box
  region box block 0 2 0 2 0 2 units lattice
  create_box 2 box
  ### Create atom type 1 in the box
  create_atoms 1 box
  ### Define the second basis
  lattice custom 5.469 basis 0.25 0.25 0.25 basis 0.25 0.75 0.75 basis 0.75 0.25 0.75 basis 0.75 0.75 0.25
  ### Create atom type 2 in the basis
  create_atoms 2 box
  mass 1 69.723
  mass 2 74.9216
  compute P all pe # THIS IS REQUIRED TO FIT ENERGIES
  compute S all pressure thermo_temp # THIS IS REQUIRED TO FIT STRESSES
----------------------------------------LAMMPS NON-TABULATED PARAMETER COMMANDS
  pair_style hybrid/overlay tersoff lj/cut/coul/dsf 0.25 12.00000 12.00000
  set type 1 charge PARAM # Put "PARAM" for parameters that will be changed
  pair_coeff * * lj/cut/coul/dsf PARAM 1.000 # Absence of "PARAM" means no change
----------------------------------------LAMMPS TABULATED PARAMETER COMMANDS
  pair_coeff * * tersoff TEMPPOT Ga As # Put "TEMPPOT" for the potential
  #pair_style tersoff # Can use the "#" symbol to comment out commands in LAMMPS
  #pair_coeff * * TEMPPOT Si
----------------------------------------END OF FILE
```

III. Input Files (continued)

1) INPUT (continued)

It is crucial that the the lines beginning in "−−−−−" are NOT changed, as the code recognizes these exact lines to separate different types of input.

It is crucial that no extra lines are made before the LAMMPS setup commands.

GA COMMANDS

generations: Number of generations to run.
extrainds: Extra individuals to add to population from EXTRAINDS.
population_size: Size of population.
mutation_rate: Rate of mutation.
elite_percentage: Percent of individuals to take as elite every generations.
intervals: Intervals for all parameters (***must fit this all on one line***).
symmetric_lines: Location of lines to apply symmetric mixing rules.

- ***The lines specified here correspond to lines in REFPOT that will be copied to the next line.***
- ***If no symmetric mixing rules will be used, leave a high number (higher than number of lines in potential file***).

symmetric_coeffs: This specifies the atom types to be used for the copied line. In the example above, the pair coefficients from line 21 will be copied to line 22 and the atom types "As Ga Ga" will be used from the atom types for line 22 in REFPOT.

LAMMPS SETUP COMMANDS

Normal LAMMPS commands can be used in this section. The purpose of this section is to set up the simulation box, create atoms, declare atom types, etc.. It is crucial that atom type 1 is created first, atom type 2 is created second, and so forth. This is because the atom data read from CONFIGS corresponds to VASP format, which has the atom types ordered in this way. *It is best to try creating the structure in LAMMPS first and viewing the list of atom IDs and types before you start fitting.* For only one atom type, this is much simpler because you can just create all the atoms once in a given basis, without worrying how they are ordered.

LAMMPS NON-TABULATED PARAMETER COMMANDS

This section also consists of normal LAMMPS commands. The purpose of this section is to declare the parameters that are *not* used in a tabulated fashion. Insert "PARAM" for any parameter that you wish to fit, and leave a numerical value for any parameter that you wish to keep constant. Make sure that the number of non-tabulated parameters in the first section corresponds to the number of times "PARAM" is used here.

LAMMPS TABULATED PARAMETER COMMANDS

This section also consists of normal LAMMPS commands. The purpose of this section is to declare the parameters are used in a tabulated fashion. Simply enter "TEMPPOT" as the parameter table that you wish to include in fitting, and make sure TEMPPOT is in the directory.

2) EXTRAINDS

This file contains individuals in the form of binary strings that can be used in the initial population of another GA run. The number of individuals added to the run is specified by the " `extrainds`" flag in INPUT. This flag can be zero and the EXTRAINDS file can be empty, but it must be present in the directory.

3) TEMPPOT

This is a LAMMPS tabulated potential file. It does not matter what the values of the parameters are for starting out, since this file reads from REFPOT to test individual fitnesses.

4) REFPOT

This is also a LAMMPS tabulated potential file, but this is used as a reference to change TEMPPOT when testing individuals. ***The parameters that are to be unchanged must be given the desired value in this file. There must be exactly 5 spaces in between parameters and each parameter MUST have a decimal point. The code works by finding decimal points in the lines, so that every value with a decimal point denotes a column.***

5) CONFIGS

Contains information on atomic positions, forces and cohesive energies obtained from DFT. The file is formatted as follows (for 8 atoms):

```
# Empty space
--------------- CONFIGURATION 1 -----------------
x       y       z      fx      fy      fz
x       y       z      fx      fy      fz
x       y       z      fx      fy      fz
x       y       z      fx      fy      fz
x       y       z      fx      fy      fz
x       y       z      fx      fy      fz
x       y       z      fx      fy      fz
x       y       z      fx      fy      fz
Ecoh
Sxx Syy Szz Sxy Sxz Syz
```

The first two lines are ignored. The next N lines are the atomic positions and forces. The last line is the cohesive energy of the configuration. ***To fit multiple configurations, simply concatenate the above in the same file.***

IV. Output FILES

1) LASTGEN

Contains parameter values and objective function values for all the individuals in the last generation.

2) HISTORY

Contains the lowest objective function value for each generation.

V. Submitting Multiple Runs on The Cluster

1) Generate trial files
2) Use “runtrials.pbs” to loop through all the files and submit a GA run for each
3) Use “gather.pbs” to gather all the best individuals in all of the LASTGEN files into the “lastgen_house” directory.
4) Use “genextrainds.m” to generate an EXTRAINDS file to be used for further fitting.