Chemical Space Explorer
ChemSPX

Ignas Pakamorė

School of Chemistry
College of Science and Engineering
University of Glasgow

September 2022

1 Input Parameters for ChemSPX

Input/Output

- `calculation_name` [string]
  Optional. Not in use.

- `in_file` [string]
  Must be specified.
  Except if `init_data_sampling` set to LHSEQ.
  Reference data file

- `verbose` [int]
  Optional. If set to 0 console printing is reduced to a minimum.
  Default: 1.

- `out_dir` [string]
  Must be specified. Calculation output directory.
  Default: '/OUT'.

- `write_initial` [True *or* False]
  Optional. Writes initial data files: initial_points.csv and initial_fx.csv. The first one contains coordinates of generated initial data points; the later $f(x)$ of the coordinates.
  Default: False.

- `print_every` [int]
  Optional. Controls printed updates in the main loop.
  Default: 1.

- `write_f_every` [int]
  Optional. Every n^{th} iteration file with optimised data point coordinates is written.
  Default: 1.

- `PCA` [True *or* False]
  Optional. Every n^{th} iteration principle component of optimised output data are calculated.
  Default: False.

- `pca_n_components` [int]
  Optional. The number of components for PCA.
  Default: 2.

(INITIAL) DATA SAMPLING AND RESTART

- `init_data_sampling` [string]
  Must be specified. Initial data sampling methods. If a restart is used calculations are started from the positions defined in the restart file. This can be used to start calculations from customary positions.
  Options: LHS, LHSEQ, void, restart
  Default: LHS

- `sample_number` [int]
  Must be specified. Number of initial sample points

- `restart_file_name` [string]
  Optional. .csv file containing data point positions for the calculations.

- `method` [string]
  Must be specified. Method for space sampling. If subspace is selected: change 0 in the reference data file to 1. As well, `split_value` must be defined (*Default* 0.1).
  Options: full_space, sub_space_c and sub_space.

- `Apply_BD` [True *or* False]
  Must be specified. Apply boundary conditions. If True, UBL and LBL must be specified.

- `UBL` [list]
  Optional. Upper boundary limit. specify a list of numbers (*int* or *float*) separated by a comma. Must match the number of ref. data axes.
  e. g 1, 2, 3

- `LBL` [list]
  Optional. Lower boundary limit. specify a list of numbers (*int* or *float*) separated by a comma. Must match the number of ref. data axes.
  e. g -1, -2, -3

OPTIMISATION

- `OPT_method` [string]
  Must be specified.
  Options: GA, BO, GRID

- `optimisation_cycles` [int]
  Must be specified. The number of optimisation cycles done by the optimiser. Not applicable for the GRID method.

- `iteration_num` [int]
  Must be specified. The number of total main loop iteration cycles done by the program.

CONVERGENCE

- `check_conv_every` [int]
  Optional. Check the convergence every n^{th} iteration cycle.
  Default: 10.

- `conv_fx` [float *or* int]
  Must be specified. Convergence value for average function $\langle f(x) \rangle$.

- `conv_del_fx` [float *or* int]
  Must be specified. Convergence value for derivative of the average function: $\Delta\langle f(x) \rangle$.

- `conv_vec` [float *or* int]
  Must be specified. Convergence value for average vector change $\langle |\Delta\vec{v}| \rangle$.

FUNCTION

- `f(x)` [string]
  Must be specified. The optimisation function.
  Options: BallTree_Force, Force, BallTree_COS, external.

- `power` [float *or* int]
  Optional. The exponent used in the pseudo-force equation.
  Default: 1.

- `k` [int string]
  Optional. k number of nearest neighbours used in BallTree_Force function.
  Default: 'all'.

FUNCTION MAP

- `map_function` [True *or* False]
  Optional. Returns density of underlying distribution of function.

- `map_grid_size` [int]
  Optional. LHS grid sample size. Only applicable for map_type force.

- `map_type` [string]
  Optional. Type of map.
  Options: density, force.

DEFAULT PARAMETERS: GA

- `mut_prob` [float]
  Mutation probability.
  Default: 0.1.

- `cross_prob` [float]
  Crossover probability.
  Default: 0.5.

- `parent_po` [float]
  Parent population.
  Default: 0.3.

- `elit_ratio` [float]
  Elite ratio.
  Default: 0.01.

- `max_iteration_without_improv` [int]
  The maximum number of iterations without improvement.
  Default: 50.

- `crossover_type` [string]
  Cross over type.
  Default: Uniform.

- `n_processes` [int]
  The number of processes. Only recommended for heavy objective functions.
  This parameter is GLOBAL *i. e.* might start in parallel with other processes.
  Default: 1.

DEFAULT PARAMETERS: BO

These parameters are not controlled via input file *or* dictionary. To change them see functions.py module, CSPX_BO class, and run_bayassian function.

- `acq_func`
  Acquisition function.
  Default: gp_hedge.

- `initial_point_generator`
  Initial point generator.
  Default: lhs.

- `random_state`
  Random state.
  Default: None.

- `verbose`
  Verbose.
  Default: False.

DEFAULT PARAMETERS: Ball-Tree

- `metric` [string]
  Metric.
  Default: euclidean.

- `leaf_size` [int]
  Leaf size.
  Default: 20.