Elements on the "dimer in a WCA solvent" code

Gabriel Stoltz (stoltz@cermics.enpc.fr)

February 7, 2010

1 Overall structure

The source files are in the directoty `src/`, the input files are in the directory `input/`, the output files are in the directory `output/`, and the present documentation is in `doc/`. The codes should be compiled and run within the `src/` directory as they stand. The input file read by the executable should be self-explanatory, otherwise the user should have a look at the corresponding file `input_METHOD.cpp` which has been compiled depending on the method (see the `makefile` command file for a list of files required for the compilation).

Besides header files (extension `.hpp`), the source files include `gaussian.cpp` which is used to produce gaussian random numbers; `vector.cpp` and `matrix.cpp` which are simple vector and matrix classes defined in order to require the least amount of `C++` libraries; `main.cpp` which is the (concise) main files, in which a call to the function `load` is performed. This is the master function in the codes, which is defined in a file whose name depends on the method at hand; finally, forces and energy computations are performed with `hamiltonian.cpp`.

2 Model and input files

2.1 Microscopic interactions and thermodynamic conditions

The parameters describing the microscopic interactions and the thermodynamic state of the system are the same for all input files:

- `Friction coefficient` γ used in the Langevin dynamics;
- `Inverse temperature` β;
- `Number of particles` n, the total number being then $N = n^2$, and the number of solvent particles is $N - 2$;
- `Elementary cell size` a, so that the density of solvent is $\rho = (1 - 2/N)a^{-2}$;
- `WCA equilibrium distance` σ;
- `WCA energy epsilon` ε;
- `Dimer barrier height` h;
- `Dimer bond length` w.

Besides, time step Δt (see the field `Time step`) is needed in all input files.

2.2 Canonical sampling

A simple canonical sampling can be obtained by `make sampling`, which produces `sampling++`. The output files are: `energy` gives the energy as a function of time; `coord` is the instantaneous distribution of the values of the reaction coordinate; `xmakemol.xyz` is a movie of the system readable with the software `XMakeMol` (see `www.nongnu.org/xmakemol/`).

The additional parameters are

- `Dynamics (0=Ovrdmpd, 1=Lngvn)` depending on whether Langevin dynamics or overdamped dynamics should be used;
- `Number of iterations`;
- `Xmakemol output frequency` is the number of steps before a new configurations is written down;
- `Other outputs frequency` is the frequency at which the energy and the value of the reaction coordinate are saved.

2.3 Extended bridge sampling

The code is compiled with the instruction `make mbar`, which produces `mbar++`. The additional parameters for the production part are

- `Number config. / restraint` is the required number of configurations per restraining center;
- `Preliminary thermalization` is the number of steps of the dynamics before configurations are acquired;
- `Number of subsampling steps` is the number of steps before a new configuration is saved;
- `Restraining potential` K is spring constant for the restraining potential;
- `Lower restraint center` $z_{\min}$ is the opsition of the lowest center;
- `Upper restraint center` $z_{\max}$ is the position of the largest center;
- `Number of restraint cent.` gives the number N_z of centers considered;

while the parameters for the post-processing part are

- `Lower value of RC` $\xi_{\min}$ and
- `Upper value of RC` $\xi_{\max}$ give the range of values of the reaction coordinate for which the free energy is required;
- `Number of PMF RC values` N_ξ gives the number of bins in this interval;
- `Tolerance for SC convergence` is the stopping criterion for the simple fixed-point iteration used to solve the nonlinear equation on the ratios of normalizing constants.

There are several output files: `energy` and `coord` list, in each line, the energies and the values of the reaction coordinate for the configurations sampled at a given value of the restraining center; `ratio` gives the final ratios of partition functions (N_z values), while `PMF` is the final potential of mean force profile in the interval $[x_{\min}, \xi_{\max}]$, obtained by computing the average of indicator functions of bins of size $\Delta z = (\xi_{\max} - \xi_{\min})/N_\xi$.

2.4 Thermodynamic integration

The code is compiled with the instruction `make ti`, which produces `ti++`. The parameters of the dynamics are

- `Dynamics (0=Ovrdmpd, 1=Lngvn)` depending on whether projected Langevin dynamics or projected overdamped dynamics should be used;
- `Number of iterations`;
- `Number of thermalization steps`: number of steps performed each time the value of the reaction coordinate changes before the computation of the approximation of the mean force by ergodic averages starts;
- `Output frequency` is the number of time steps before the outputs are written out (see below for a list);
- `Lower value of RC` and `Upper value of RC` define the biasing region where the mean force is estimated (the bias being 0 outside this zone);
- `Number of PMF RC values` gives the number of bins in which the reaction coordinate space is separated;

The output files are the following:

- `coord` lists the values of the reaction coordinate;
- `energy` gives the total energy and the potential energy;
- `current_mean_force` is the current ergodic approximation of the mean force. In the overdamped case, the first column is obtained by averaging the local mean force (known analytically here), the second by averaging the plain Lagrange multipliers, the third by averaging the Lagrange multipliers with elimination of the martingale part, the fourth with the variance reduction method using time reversed increments. In the Langevin case, the first column is obtained by averaging the local mean force, the second one by averaging the Lagrange multipliers used in the Rattle part, the third by computing the momentum average local mean force, the fourth and fifth ones being the averages of the Lagrange multipliers of the position and momentum constraints, respectively.
- `Lagrange` gives the instantaneous values of the mean force and Lagrange multipliers which are averaged in the previous file (same ordering).
- `mean_force` and `PMF` give the final mean force and free energy profiles respectively.

2.5 Adaptive dynamics

The code is compiled with the instruction `make abf`, which produces `abf++`. There are several output files: `energy` gives the energy as a function of time; `coord` is the instantaneous distribution of the values of the reaction coordinate, while `coord_first` is the same information for the first system only; `mean_force` is the current mean force (the values of the reaction coordinate bins are in the first line, and each subsequent line is a profile some time later than the previous one), while `PMF` is the current associated free energy profile obtained by numerical integration; finally, `histo` is the cumulated histogram of visits in a bin.

The parameters of the dynamics are

- `Dynamics (0=Ovrdmpd, 1=Lngvn)` depending on whether Langevin dynamics or overdamped dynamics should be used;
- `Number of iterations`;
- `Number of replicas` gives the number of systems simulated in parallel but contributing to the same free energy profile;
- `Mean force/bias output freq.` gives the frequency at which mean force, free energy profiles, distribution of reaction coordinates and cumulated histogram are written out (counted in number of iteration steps)
- `Other outputs frequency` is the number of time steps before the current energy and reaction coordinate value of the first system are written out;
- `Lower value of RC` and `Upper value of RC` define the biasing region where the mean force is estimated (the bias being 0 outside this zone);
- `Number of PMF RC values` gives the number of bins in which the reaction coordinate space is separated;
- `Selection` is set to 1 when selection is required;
- `Section intensity` is the prefactor before the selection term (provided selection is required).

2.6 Nonequilibrium dynamics

The code is compiled with the instruction `make jarz`, which produces `jarz++`.

- `Number of thermalization steps` is the subsampling time used to sample the initial conditions from a projected dynamics;
- `Work distribution output freq.` is the number of steps after which the whole work distribution is written out;
- `Other output frequency` is the number of steps after which the energies and the value of the reaction coordinate are saved;
- `Lower value of RC` is the initial value of the reaction coordinate at the beginning of the switching;

- `Upper value of RC` is the final value of the reaction coordinate;
- `Switching time` gives the time to perform the switching;
- `Number of replicas`

The output files are the following:

- `coord` gives the value of the reaction coordinate of the first system as a function of time;
- `energy` gives the total energy and the potential energies as a function of time;
- `PMF` gives the free energy profile as a function of time, with the estimate using the local mean force on the second column, and the estimates obtained with the Lagrange multipliers on the third one;
- `Work` is the work distribution at the required times, for the works computed with the local mean force;
- `Work_lagrange` is the work distributions obtained for the works computed from the Lagrange multipliers;
- `current_work` gives the current work for the first system, computed using the local mean force on the first column, and the Lagrange multipliers on the second.

3 Useful analytical expressions

Reaction coordinate and derivatives. The chosen reaction coordinate is

$$\xi(q) = \frac{|q_1 - q_2| - r_0}{2w},$$

where $q_1, q_2 \in \mathrm{R}^d$, d denoting the dimension of the ambient physical space ($d = 2$ here). Its gradient is

$$\nabla\xi(q) = \frac{1}{2w} \begin{pmatrix} \dfrac{q_1 - q_2}{|q_1 - q_2|} \\ -\dfrac{q_1 - q_2}{|q_1 - q_2|} \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

and

$$|\nabla\xi(q)|^2 = \frac{1}{2w^2}.$$

The local mean force is

$$\begin{aligned} f(q) &= \frac{\nabla\xi(q)\cdot\nabla V(q)}{|\nabla\xi(q)|^2} - \frac{1}{\beta}\,\mathrm{div}\left(\frac{\nabla\xi(q)}{|\nabla\xi(q)|^2}\right) \\ &= \frac{w}{|q_1 - q_2|}\left[(q_1 - q_2)\cdot\Big(\partial_{q_1}V(q) - \partial_{q_2}V(q)\Big) - \frac{2(d-1)}{\beta}\right]. \end{aligned}$$

Finally, notice that there is no Fixman term since $|\nabla\xi|$ is constant.

Constrained overdamped Langevin processes. The constrained dynamics

$$\left\{\begin{array}{rcl} dq_t &=& -\nabla V(q_t)\,dt + \sqrt{\dfrac{2}{\beta}}\,dW_t + \nabla\xi(q_t)\,d\Lambda_t, \\ \xi(q_t) &=& z, \end{array}\right.$$

is integrated numerically with the scheme

$$\left\{\begin{array}{rcl} q^{n+1} &=& q^n - \nabla V(q^n)\,\Delta t + \sqrt{\dfrac{2\Delta t}{\beta}}\,G^n + \nabla\xi(q^{n+1})\,\Delta\Lambda^{n+1}, \\ \xi(q^{n+1}) &=& z. \end{array}\right.$$

Therefore,

$$\left(1 - \frac{\Delta\Lambda^{n+1}}{w|q_2^{n+1} - q_1^{n+1}|}\right)(q_2^{n+1} - q_1^{n+1}) = \widetilde{q}_2^n - \widetilde{q}_1^n,$$

with

$$\widetilde{q}^n = q^n - \nabla V(q^n)\,\Delta t + \sqrt{\frac{2\Delta t}{\beta}}\,G^n.$$

and $G^n = (G_1^n, G_2^n, \dots)$ with $G_i^n \in \mathbb{R}^d$. Since $|q_2^{n+1} - q_1^{n+1}| = 2wz + r_0$ is fixed, the lagrange multiplier reads

$$\Delta\Lambda^{n+1} = w\Big(2wz + r_0 - |\widetilde{q}_2^n - \widetilde{q}_1^n|\,\Big) = 2w^2\Big(z - \xi(\widetilde{q})\Big).$$

Besides, $q_1^{n+1} + q_2^{n+1} = \widetilde{q}_2^n + \widetilde{q}_1^n$, so that

$$q_1^{n+1} = \frac{1}{2}\left(1 + \frac{1}{1 - \Delta\Lambda^{n+1}/[w(2wz + r_0)]}\right)\widetilde{q}_1^n + \frac{1}{2}\left(1 - \frac{1}{1 - \Delta\Lambda^{n+1}/[w(2wz + r_0)]}\right)\widetilde{q}_2^n,$$

and a similar expression for q_2^{n+1}. Some care is however required to treat correctly the periodic boundary conditions. With the variance reduction procedure,

$$\Delta\widetilde{\Lambda}^{n+1} = \Delta\Lambda^{n+1} + \sqrt{\frac{2\Delta t}{\beta}}\,\frac{\nabla\xi(q^n)\cdot G^n}{|\nabla\xi(q^n)|^2} = \Delta\Lambda^{n+1} + w\frac{q_1^n - q_2^n}{|q_1^n - q_2^n|}\cdot(G_1^n - G_2^n).$$

The expressions in the case of nonequilibrium switching processes are similar. It suffices to consider an additional term related to the variation of the constraint:

$$\left\{\begin{array}{rcl} q^{n+1} &=& q^n - \nabla V(q^n)\,\Delta t + \sqrt{\dfrac{2\Delta t}{\beta}}\,G^n + \nabla\xi(q^{n+1})\,\Delta\Lambda^{n+1}, \\ \xi(q^{n+1}) &=& z^{n+1}, \end{array}\right.$$

and

$$\Delta\Lambda_{\rm noneq}^{n+1} = \Delta\Lambda^{n+1} + \sqrt{\frac{2\Delta t}{\beta}}\,\frac{\nabla\xi(q^n)\cdot G^n}{|\nabla\xi(q^n)|^2} = \Delta\Lambda^{n+1} + w\frac{q_1^n - q_2^n}{|q_1^n - q_2^n|}\cdot(G_1^n - G_2^n) - \frac{z^{n+1} - z^n}{|\nabla\xi(q^n)|^2}.$$

Constrained Langevin processes. We detail how the Rattle step

$$\left\{\begin{array}{rcll} p^{n+1/2} &=& p^n - \dfrac{\Delta t}{2}\nabla V(q^n) + \nabla\xi(q^n)\,\lambda^{n+1/2}, & \\ \xi(q^{n+1}) &=& z, & (C_q) \\ q^{n+1} &=& q^n + \Delta t\, M^{-1}\, p^{n+1/2}, & \\ p^{n+1} &=& p^{n+1/2} - \dfrac{\Delta t}{2}\nabla V(q^{n+1}) + \nabla\xi(q^{n+1})\,\lambda^{n+1}, & \\ \nabla\xi(q^{n+1})^T M^{-1} p^{n+1} &=& 0, & (C_p) \end{array}\right.$$

is implemented. In the code, $M = \mathrm{Id}$. Denoting

$$\widetilde{q}^n = q^n + \Delta t\, p^n - \frac{\Delta t^2}{2}\,\nabla V(q^n),$$

the value of $\lambda^{n+1/2}$ is determined by the condition

$$\xi\left(\widetilde{q}^n + \Delta t\,\lambda^{n+1/2}\nabla\xi(q^n)\right) = z,$$

which can be rewritten as

$$\left\|\widetilde{q}_1^n - \widetilde{q}_2^n + \frac{\Delta t\,\lambda^n}{w}\,e_{12}^n\right\|^2 = z^2,$$

with

$$e_{12}^n \equiv e_{12}(q^n) = \frac{q_1^n - q_2^n}{|q_1^n - q_2^n|}.$$

This equation is a second-order equation in the $\Delta t\,\lambda$ variable:

$$(\Delta t\,\lambda)^2 + 2we_{12}^n\cdot(\widetilde{q}_1^n - \widetilde{q}_2^n)\,\Delta t\,\lambda + w^2\left(\|\widetilde{q}_1^n - \widetilde{q}_2^n\|^2 - z^2\right) = 0,$$

whose solutions are

$$\lambda_\pm = \frac{w}{\Delta t}\Big(-e_{12}^n\cdot(\widetilde{q}_1^n - \widetilde{q}_2^n) \pm \sqrt{\Delta}\Big),$$

with

$$\Delta = \Big[e_{12}^n\cdot(\widetilde{q}_1^n - \widetilde{q}_2^n)\Big]^2 - 4\left(\|\widetilde{q}_1^n - \widetilde{q}_2^n\|^2 - z^2\right).$$

The smallest λ is chosen, so that $\lambda^n = \lambda_-$ when $e_{12}^n\cdot(\widetilde{q}_1^n - \widetilde{q}_2^n) \geq 0$ and $\lambda^n = \lambda_+$ otherwise.

The Lagrange multiplier $\lambda^{n+1/2}$ is determined by

$$\nabla\xi(q^{n+1})^T M^{-1} p^{n+1} = 0,$$

which can be rewritten as

$$e_{12}^{n+1}\cdot\left(p_1^{n+1} - p_2^{n+1}\right) = e_{12}^{n+1}\cdot\left(\widetilde{p}_1^{n+1} - \widetilde{p}_2^{n+1} + \frac{\lambda^{n+1}}{w}e_{12}^{n+1}\right) = 0,$$

with

$$\widetilde{p}^{n+1} = p^{n+1/2} - \frac{\Delta t}{2}\nabla V(q^{n+1}).$$

Then,

$$\lambda^{n+1} = -w\,e_{12}^{n+1}\cdot\left(\widetilde{p}_1^{n+1} - \widetilde{p}_2^{n+1}\right).$$

The comparison with the local rigid mean force (equal to the local mean force) requires the computation of

$$f_{\mathrm{rgd}}(q,p) = G_M^{-1}\nabla\xi^T M^{-1}\nabla V(q) - G_M^{-1}(q)\mathrm{Hess}_q(\xi)(M^{-1}p, M^{-1}p).$$

The first term is the same as for the local mean force. For the second term, some straightforward computations give

$$\begin{aligned} &G_M^{-1}(q)\mathrm{Hess}_q(\xi)(M^{-1}p, M^{-1}p) = w\nabla^2\xi(q) : p\otimes p \\ &= \frac{w}{|q_1-q_2|}\left[(p_{1,x}-p_{2,x})^2 + (p_{1,y}-p_{2,y})^2 - \Big(\frac{x_1-x_2}{|q_1-q_2|}(p_{1,x}-p_{2,x}) + \frac{y_1-y_2}{|q_1-q_2|}(p_{1,y}-p_{2,y})\Big)^2\right]. \end{aligned}$$

Notice that

$$\frac{x_1-x_2}{|q_1-q_2|}(p_{1,x}-p_{2,x}) + \frac{y_1-y_2}{|q_1-q_2|}(p_{1,y}-p_{2,y}) = 2w\, e_{12}\cdot(p_1-p_2).$$

This quantity is therefore equal to 0 for constrained processes (and in general proportional to the momentum or velocity constraint).

Nonequilibrium Langevin processes. We describe how the momentum constrained is handled. It reads, in the nonequilibrium case,

$$\nabla\xi(q^{n+1})^T M^{-1} p^{n+1} = \frac{z(t_{n+1})-z(t_n)}{\Delta t}.$$

with

$$p^{n+1} = p^{n+1/2} - \frac{\Delta t}{2}\nabla V(q^{n+1}) + \nabla\xi(q^{n+1})\,\lambda^{n+1}.$$

Then,

$$\lambda^{n+1} = w\left[-e_{12}^{n+1}\cdot(\widetilde{p}_1^{n+1} - \widetilde{p}_2^{n+1}) + \frac{z(t_{n+1})-z(t_n)}{\Delta t}\right]$$

with

$$\widetilde{p}^{n+1} = p^{n+1/2} - \frac{\Delta t}{2}\nabla V(q^{n+1}).$$