Computation of pair distribution functions and three-dimensional densities with a reduced variance principle

Daniel Borgis, Roland Assaraf, Benjamin Rotenberg, Rodolphe Vuilleumier

Computation of pair distribution functions and three-dimensional densities with a reduced variance principle

Daniel Borgis[1], Roland Assaraf[2], Benjamin Rotenberg[3],
and Rodolphe Vuilleumier[1]

[1]*Pôle de Physico-Chimie Théorique, École Normale Supérieure, UMR 8640*
CNRS-ENS-UPMC, 24 rue Lhomond, 75005 Paris, France

[2]*CNRS, UPMC Univ. Paris 06, ESPCI, UMR 7616 LCT, 75005 Paris, France*

[3]*CNRS, UPMC Univ. Paris 06, ESPCI, UMR 7195 PECSA, 75005 Paris, France*

Abstract

No fancy statistical objects here, we go back to the computation of one of the most basic and fundamental quantities in the statistical mechanics of fluids, namely the pair distribution functions. Those functions are usually computed in molecular simulations by using histogram techniques. We show here that they can be estimated using a global information on the instantaneous forces acting on the particles and that this leads to a reduced variance compared to the standard histogram estimators. The technique is extended successfully to the computation of three-dimensional solvent densities around tagged molecular solutes, quantities that are noisy and very long to converge using histograms.

1   Introduction

What we all learned from our mentor is that, for a given statistical mechanics problem, it is important (and intellectually rewarding) to first derive the proper, irrefutable, mathematical formulation, to understand as much as possible the formulas in mechanical terms, and, at the end, to be able to express the relevant formulas in computational forms that can be tackled with efficient and accurate simulation algorithms. Such goal

1

can be hardly reached without the help of a chalk and a blackboard, and without questioning every single line.[1]

We hope that the present contribution is as much as possible in this vein, although it stays away from any complex problems in computational statistical mechanics that have attracted the community recently [1,2], in particular quantum-classical statistical mechanics [3,4], statistical mechanics with constraints [5–7] or phase-space exploration and characterization for complex systems using accelerated sampling methods [8,9]. We go back here to the computation of the most elementary objects in liquid-state theories, namely the two-particle pair-distribution functions in bulk liquids, or the three-dimensional one-particle liquid density around a tagged molecular object. [10,11] The evaluation of such quantities, as prescribed in simulation textbooks [12, 13], is usually done through histograms, or possibly improved histograms, that rely in any case on the representation of the Dirac δ-function on a discrete grid. With such method, the variance of the results diverges when the grid spacing goes to zero. Elaborating on the work of Assaraf and collaborators [14, 15] for the evaluation quantum electronic densities with zero-variance/zero-bias principles, we propose a different approach to the computation of pair distribution functions and one-particle densities that leads to a much reduced variance of the results with respect to straight histogram counting and furthermore achieves a finite variance when the grid spacing tends to zero. As will be seen, smooth curves can be obtained even with very limited statistics.

Their will be a fair amount of explicit rigorous mathematical manipulations in this contribution, as it suits to a rigorous scientist who denies any physical (not to speak of chemical) intuition and thinks that Nature can only speak through well defined mathematical objects and their relations. We must confess, however, that some points of the derivation will not have the degree of rigor that it deserves; we will elude in particular some mathematical proofs of convergence and variance properties –in part because our degree of understanding and current formalization is not yet complete. Nevertheless, all the basic principles will be set.

The paper is organized as follows. Section 2 is devoted to the computation of

---

[1]This contribution is dedicated to Prof. Giovanni Ciccotti, in honour of his $70th$ birthday

pair distribution functions whereas section 4 focuses on three-dimensional densities. Section 4 concludes.

2 Pair distribution functions

Let us consider a fluid mixture for which microscopic configurations are generated by computer simulations in the canonical ensemble. The pair distribution function (or radial distribution function, RDF) between molecules of type a and b is defined as

$$g_{ab}(r) = \frac{\epsilon_{ab}}{4\pi r^2} \left\langle \sum_{i=1}^{N_a} \sum_{j=1}^{N_b'} \delta(r - r_{ij}) \right\rangle, \tag{1}$$

where $\epsilon_{ab} = (1 - \frac{1}{2}\delta_{ab})\frac{V}{N_a N_b}$, $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$, and $r_{ij} = |\mathbf{r}_{ij}|$. The prime in the second sum indicates that $i = j$ should be excluded in the case $a = b$. This function is generally computed through histograms with bins of of finite width, Δr, replacing the δ-function by $\frac{1}{\Delta r} h_{\Delta r}(r)$, where $h_{\Delta r}(r)$ is the characteristic function equal to 1 between r and $r + \Delta r$ and 0 otherwise. This statistical estimation is of infinite variance as $\Delta r \to 0$ since the instantaneous density in each bead oscillates between 0 and $\mathcal{O}(1/4\pi r^2 \Delta r)$.

Accounting for the rotational invariance in the bulk fluid, the pair distribution can be also expressed as

$$g_{ab}(r) = \frac{\epsilon_{ab}}{4\pi} \int d\Omega \left\langle \sum_{i=1}^{N_a} \sum_{j=1}^{N_b'} \delta(\mathbf{r} - \mathbf{r}_{ij}) \right\rangle, \tag{2}$$

with $\mathbf{\Omega} = \mathbf{r}/r$. Following the ideas of Assaraf et al for electron densities [14, 15], use can be made of the Poisson equality to replace the three-dimensional delta-function

$$\delta(\mathbf{r} - \mathbf{r}_{ij}) = -\frac{1}{4\pi} \Delta_{\mathbf{r}_i} \frac{1}{|\mathbf{r} - \mathbf{r}_{ij}|}. \tag{3}$$

Insertion of the Laplacian with respect to either $\mathbf{r}_i$ or $\mathbf{r}_j$ in the canonical average and integration by part yields (after symmetrization)

$$h_{ab}(r) = -\frac{\beta \epsilon_{ab}}{4\pi} \int d\Omega \left\langle \sum_{i=1}^{N_a} \sum_{j=1}^{N_b'} \frac{\mathbf{r}_{ij} - \mathbf{r}}{|\mathbf{r}_{ij} - \mathbf{r}|^3} \cdot \frac{1}{2}(\mathbf{F}_j - \mathbf{F}_i) \right\rangle \tag{4}$$

where $h_{ab} = g_{ab} - 1$.

Using the Gauss theorem for the electric field created by a uniformly charged sphere of radius r at the location r_{ij}

$$\int d\mathbf{\Omega} \frac{\mathbf{r}_{ij} - r\mathbf{\Omega}}{|\mathbf{r}_{ij} - r\mathbf{\Omega}|^3} = \frac{\mathbf{r}_{ij}}{r_{ij}^3} H(r_{ij} - r),\tag{5}$$

with H the Heaviside function, we get from eqn 4

$$h_{ab}(r) = -\frac{\beta \epsilon_{ab}}{4\pi} \left\langle \sum_{i=1}^{N_a} \sum_{j=1}^{N_b'} \frac{1}{2}(\mathbf{F}_j - \mathbf{F}_i) \cdot \frac{\mathbf{r}_{ij}}{r_{ij}^3} H(r_{ij} - r) \right\rangle \tag{6}$$

This is a key formula of this work. Compared to the standard histogram procedure, it now involves the force acting on the particles in addition to their positions. It also implies a quite different numerical procedure: Here, for each configuration, every particles pair contributes to all distances $r < r_{ij}$ instead of just to $r = r_{ij}$. Furthermore application of the formula requires a predefined grid but does not necessarily imply the limit of infinitely small grid separation, $\Delta r \to 0$. The only requirement is that the chosen grid, not necessarily regular, is adapted to cope with the variations of the pair distribution at all distances.

The new procedure is illustrated in Figure 1 for the radial distribution function, $g(r)$, of a pure Lennard-Jones fluid composed of 864 particles at a reduced density $\rho^* = 0.8$ and reduced temperature $T^* = 1.35$, computed by molecular dynamics simulation. We have displayed the RDF computed from ONE SINGLE equilibrated configuration using either histograms or eqn 6 and we compare those "instantaneous" curves to the converged result after 10000 time steps. In both approaches we used the same regular grid with $\Delta r = 0.005\sigma$. It can be seen that the curve obtained by eqn 6 is already very smooth and quite close to the final converged result. The histogram curve does contain the converged one within its fluctuations but appears very noisy. This is further illustrated in Fig. 2 where we plot the variances

$$v(r) = \frac{1}{T} \sum_t g_t(r)^2 - \left(\frac{1}{T} \sum_t g_t(r) \right)^2 \tag{7}$$

obtained after $T = 1000$ simulation steps; $g_t(r)$ is the "instantaneous" pair distribution function measured at step number t. It can be verified that the variance measured with the "force" approach is indeed much reduced with respect to the histogram approach
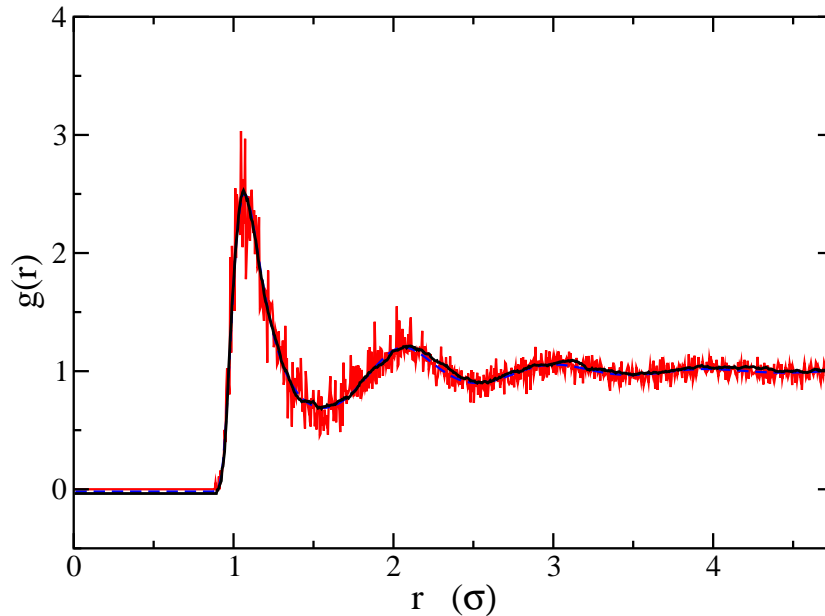
Figure 1: Radial distribution function obtained for a SINGLE equilibrated configuration of a Lennard-Jones liquid composed of 864 particles using either the force approach, eqn 6, or the standard histogram technique, with a grid spacing $\Delta r = 0.005\sigma$ The dashed blue line indicates the converged result after 10000 simulation steps.

and appears independent of the chosen grid size; the histogram method leads to a variance that is inversely proportional to Δr.

In Fig. 3 is displayed the oxygen-oxygen radial distribution obtained from 100 configurations of a DFT-MD simulation of 128 water molecules at ambient thermodynamic conditions with BLYP functional, after preliminary equilibration. It can be seen that, even for the relatively fine grid chosen and with a very limited number of steps, the "force RDF" is very smooth. It should also be noted that, even if the agreement with the converged RDF is already and overall very good after such a short trajectory, one observes slight discrepancies, in particular in the height of the first peak. At this stage, the force method is able to improve the variance of the RDF, but does not correct for the bias in the statistics. Looking for an improved method with minimum variance AND minimum bias is still another matter. Nonetheless, Fig. 3 is meant to show that
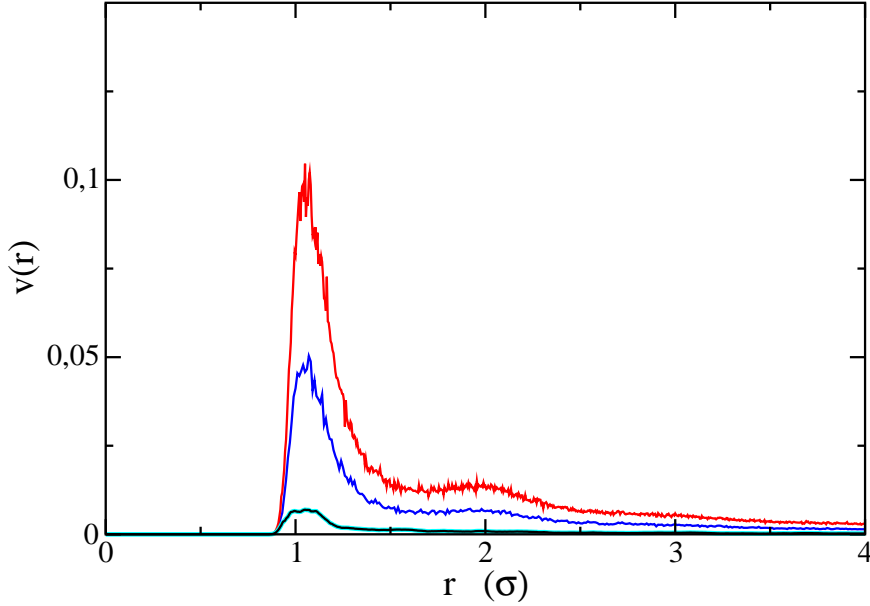
Figure 2: Variance on the value of the radial distribution function depicted in figure 1 after 1000 simulation steps using either the force approach, eqn 6, or the standard histogram technique, with a grid spacing $\Delta r = 0.01\sigma$ (black and blue lines, respectively) or $\Delta r = 0.005\sigma$ (cyan and red lines, respectively).

it is certainly in the field of ab-initio studies, where the generation of the trajectories themselves is computationally very expensive, that the force method described here to compute the RDF's reveals its full potential, given that the forces on the nuclei at each time step are readily available from the simulations.

In order to gain some physical insight into the above formula, one can write

$$H(r_{ij} - r) = \int_r^\infty dr' \delta(r' - r_{ij}) \tag{8}$$

Replacing into eq. 6 and inverting integral and canonical average, we get

$$\rho_b h_{ab}(r) = -\beta \int_r^\infty dr' F(r') \tag{9}$$

with the mean force density defined by

$$F(r) = \frac{\rho_b \epsilon_{ab}}{4\pi r^2} \left\langle \sum_{i=1}^{N_a} \sum_{j=1}^{N_b'} \frac{1}{2} (\mathbf{F}_j - \mathbf{F}_i) \cdot \frac{\mathbf{r}_{ij}}{r_{ij}} \delta(r - r_{ij}) \right\rangle. \tag{10}$$
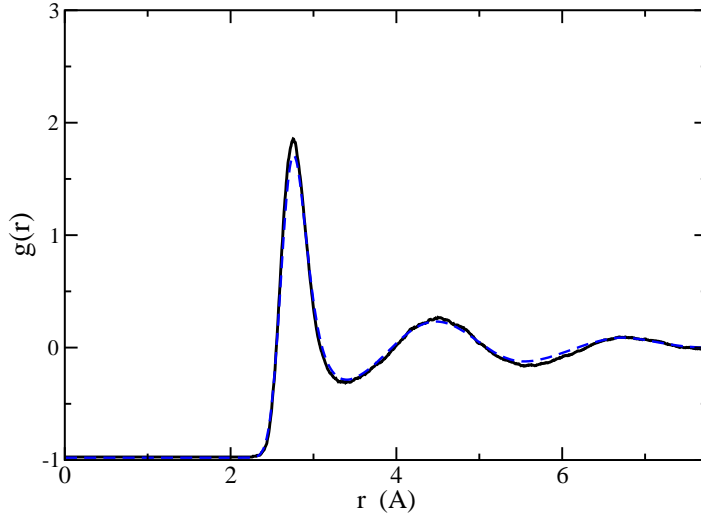
6

Figure 3: Oxygen-oxygen radial distribution function averaged over 100 configurations extracted from a DFT-MD trajectory with 128 water molecules at ambient liquid conditions. The dashed blue line indicates the converged result obtained by averaging over 38600 configurations.

We are back to a histogram procedure, but for $F(r)$ instead of $g_{ab}(r)$ directly.

Denoting by $\bar{F}(r)$ the *constrained* (or conditional) mean force at a given distance, $\bar{F}(r) = F(r)/\rho_b g_{ab}(r)$, we get by differentiation of eqn 9, and division by $g_{ab}(r)$:

$$\frac{1}{g_{ab}(r)}\frac{dg_{ab}}{dr} = \bar{F}(r) = -\beta\frac{dw_{ab}(r)}{dr}, \tag{11}$$

which is the definition of the potential of mean force (PMF), $g_{ab}(r) = \exp(-\beta w_{ab}(r))$. A fundamental difference of the present approach with respect to standard PMF calculations is the use of the force density $F(r)$ instead of the mean force $\bar{F}(r)$. Eqn 9 is thus reminescent of the usual PMF formula but not equivalent to it in practice: Here, the integration of the force density $F(r)$ yields $g_{ab}(r)$ directly rather than its logarithm and this quantity is computed in its integrality during the simulation, rather than step by step using constraints or restraints on the $a - b$ distance as in usual PMF calculations. Note also that eqs 6 and 9-10 are rigorously equivalent only in the limit of infinitely small grid size.

We note in passing that after eqn 4, a second integration by part can be performed, yielding

$$\rho_b h_{ab}(r) = \frac{\beta \epsilon_{ab}}{4\pi} \int d\Omega \left\langle \sum_{i=1}^{N_a} \sum_{j=1}^{N'_b} \frac{1}{2} (\Phi_i + \Phi_j) \frac{1}{|\mathbf{r} - \mathbf{r}_{ij}|} \right\rangle \tag{12}$$

with $\Phi_i = \beta \mathbf{F}_i^2 - \Delta_{\mathbf{r}_i} U$.

Using now the Gauss theorem for the electrostatic potential created by a uniformly charged sphere of radius r at a location $\mathbf{r}_{ij}$

$$\int d\Omega \frac{1}{|\mathbf{r}_{ij} - r\Omega|} = F(r, r_{ij}) = \min(\frac{1}{r}, \frac{1}{r_{ij}}), \tag{13}$$

eqn 12 can also be re-written as

$$\rho_b h_{ab}(r) = \frac{\beta \epsilon_{ab}}{4\pi} \left\langle \sum_{i=1}^{N_a} \sum_{j=1}^{N'_b} \frac{1}{2} (\Phi_i + \Phi_j) F(r, r_{ij}) \right\rangle . \tag{14}$$

As for eqn 6, this formula has to be evaluated at all points r of a pre-defined grid by summation over all particle pairs for each microscopic configuration. The key quantities to be evaluated for each particle are now the Φ_i's, that are scalar quantities involving the second derivatives of the potential felt by each particle (minus the mean square force times β).

In analogy to eqs 9-10, one can define also the continuous function $\Phi(r)$,

$$\Phi(r) = \frac{\rho_b \epsilon_{ab}}{4\pi r^2} \left\langle \sum_{i=1}^{N_a} \sum_{j=1}^{N'_b} \frac{1}{2} (\Phi_j + \Phi_i) \delta(r - r_{ij}) \right\rangle , \tag{15}$$

that can be evaluated through standard histograms for a given simulation. This function being computed, it is easy to show that the radial distribution can be extracted at the end through the one-dimensional integral relation

$$\rho_b h_{ab}(r) = \beta \int_r^\infty dr' r' (1 - r'/r) \Phi(r') \tag{16}$$

This second formula appears much less familiar than the mean force formula of eqn 10. We found its application, as well as that of the direct formula 14, to be very unstable and thus quite disappointing. We attribute this to the large fluctuation of the Φ_i's that appear as the difference of two large numbers. We do not have a clear and final understanding of this fact yet. For the time being, we limit ourselves to the force formulation, eqs 6 and 9, for which the reduced variance properties was shown to work quite fine.

3 Three-dimensional densities

The related problem that we address now is the computation of the three-dimensional density of a molecular solvent around a fixed molecule of arbitrary shape located at the origin using molecular simulations such as Monte-Carlo or Molecular Dynamics. This density is defined by

$$\rho(\mathbf{r}) = \left\langle \sum_{i=1}^{N} \delta(\mathbf{r} - \mathbf{r}_i) \right\rangle. \tag{17}$$

This quantity is usually computed by discretizing space into a 3D-grid and monitoring the filling of 3D-histograms of volume ΔV in the course of the simulation. This statistical process is indeed of infinite variance as $\Delta V \to 0$ since the instantaneous density in each bead oscillates betweeen 0 and $\mathcal{O}(1/\Delta V)$. The exploration of the three-dimensional volume requires much more statistics than that required for radial functions and it is indeed known that three-dimensional densities are hard to get accurately that way. To bypass that problem, the Dirac function in eqn 17 may be again transformed according to the Poisson's identity of eqn 3. Substitution in the average above and integration by part with respect to the Boltzmann weight gives

$$\Delta\rho(\mathbf{r}) = \rho(\mathbf{r}) - \rho_0 = -\frac{\beta}{4\pi} \left\langle \sum_{i=1}^{N} \frac{\mathbf{r} - \mathbf{r}_i}{|\mathbf{r} - \mathbf{r}_i|^3} \cdot \mathbf{F}_i \right\rangle \tag{18}$$

$\mathbf{F}_i = -\nabla_{\mathbf{r}_i} U$ representing the force on molecule i, U is the total interaction potential. A second integration by part yields

$$\Delta\rho(\mathbf{r}) = -\frac{\beta}{4\pi} \left\langle \sum_{i=1}^{N} \frac{\Phi_i}{|\mathbf{r} - \mathbf{r}_i|} \right\rangle \tag{19}$$

with $\Phi_i = \beta \mathbf{F}_i^2 - \Delta_{\mathbf{r}_i} U$. Those two formulas have a different content that eqn 17 since every particle now contributes instantaneously to every grid point in space, even those at long distances, and not only to the closest ones.

The straight application of those formulas is time consuming however, with a cost of order $N \times N_g$, where N_g is the number of grid points, for each solvent configuration. This might not be a limitation in ab-initio molecular dynamics where the cost of a time step is already very high. In molecular mechanics however, the cost should be compared to N^2 or even $N \ln(N)$ for advanced algorithms. A compromise has therefore to be found.

Inspired by the developments of the previous section, we can introduce a delta-function in $\mathbf{r}'$ into eqn (18)

$$\Delta\rho(\mathbf{r}) = -\frac{\beta}{4\pi}\left\langle\sum_{i=1}^{N}\int d\mathbf{r}'\delta(\mathbf{r}'-\mathbf{r}_i)\frac{\mathbf{r}-\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|^3}\cdot\mathbf{F}_i\right\rangle \tag{20}$$

and with the definition of the *force density* at point $\mathbf{r}'$

$$\mathbf{F}(\mathbf{r}') = \left\langle\sum_{i=1}^{N}\delta(\mathbf{r}'-\mathbf{r}_i)\mathbf{F}_i\right\rangle \tag{21}$$

one gets

$$\Delta\rho(\mathbf{r}) = -\frac{\beta}{4\pi}\int d\mathbf{r}'\frac{\mathbf{r}-\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|^3}\cdot\mathbf{F}(\mathbf{r}'). \tag{22}$$

In practice, $\mathbf{F}(\mathbf{r}')$ can be computed by histograms during a simulation and at the end Fourier-transformed to $\mathbf{F}(\mathbf{k})$ using a discrete three-dimensional fast Fourier transform (FFT). $\rho(\mathbf{r})$ is then obtained by inverse FFT of

$$\Delta\rho(\mathbf{k}) = \frac{i\beta}{k^2}\mathbf{k}\cdot\mathbf{F}(\mathbf{k}). \tag{23}$$

Defining in the same way

$$\Phi(\mathbf{r}') = \left\langle\sum_{i=1}^{N}\delta(\mathbf{r}'-\mathbf{r}_i)(\beta\mathbf{F}_i^2-\Delta_{\mathbf{r}_i}U)\right\rangle \tag{24}$$

one has equivalently

$$\Delta\rho(\mathbf{r}) = -\frac{\beta}{4\pi}\int d\mathbf{r}'\frac{\Phi(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \tag{25}$$

or

$$\Delta\rho(\mathbf{k}) = -\beta\frac{\Phi(\mathbf{k})}{k^2}. \tag{26}$$

Noting that in fact $\Phi(\mathbf{r}) = \nabla_{\mathbf{r}}\cdot\mathbf{F}(\mathbf{r})$, or in k-space, $\Phi(\mathbf{k}) = -i\mathbf{k}\cdot\mathbf{F}(\mathbf{k})$, the above equations involving $\mathbf{F}$ or Φ are indeed equivalent. Eqn (26) involves the direct analytical calculation of the divergence of the mean force whereas eqn (23) involves its numerical estimation by differentiation in k-space.

As an illustration of this formalism, we display in Fig. 4 the three-dimensional water density around a given water molecule computed from the DFT-MD simulation with 128 water molecules already described in Fig. 3. Here the whole trajectory was exploited, *i.e.* 38600 stored configurations separated by $5\,fs$. The simulation was
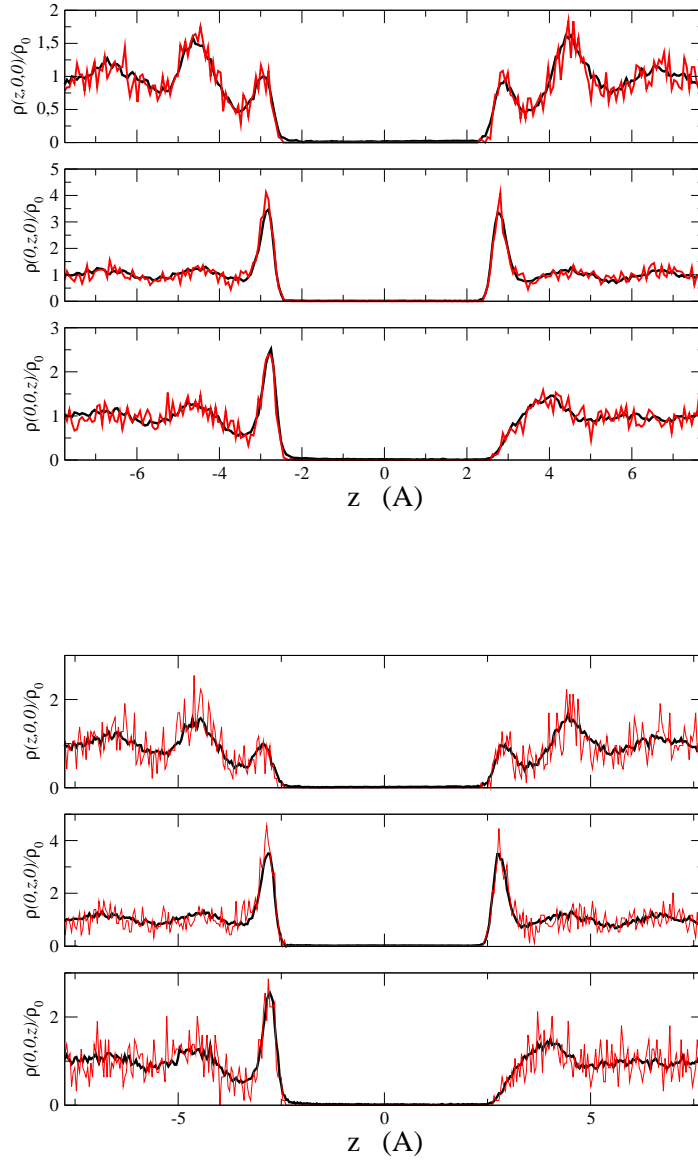
Figure 4: Three-dimensional water density around a fixed water molecule from 36800 DFT-MD configurations of 128 molecules, represented in three orthogonal directions: $\vec{z}$ along the H_1OH_2 bissector, $\vec{y}$ perpendicular to the molecular plane, $\vec{x} = \vec{z} \times \vec{y}$. The red and black lines are the results of the histogram and force methods, respectively. The grid spacing is $\Delta r = 0.06$ Å for the top figure and $\Delta r = 0.04$ Å for the bottom one.

generated with CP2K code using the conventional BLYP functional. It is seen that the integrated force method described by eqs. 22-23 yields smooth densities that are basically independent of the chosen histogram bin size, whereas the straight histogram method yields noisy densities with fluctuations increasing with the fineness of the grid. Restricting the sampling, which is rather high here, to a smaller number of configurations yields analogous results with coarser grids.

We note also that, as already found for the RDF's, the use of the second derivative approach using the function $\phi(\mathbf{r})$ in eqs. 24-26 instead of the above first-derivative approach with the forces yields numerical instabilities in the computation that we attribute to the very large statistical fluctuations of the molecular quantities Φ_i's; we still do not fully understand this behaviour.

4 Conclusions

We have derived in this paper two rather unusual formulas to compute the pair distribution functions in a bulk liquid mixture, namely eqs. 6 and 14. They require at each time step and for each pair distance realization to accumulate statistics over a whole pre-defined grid rather than just updating one histogram bin corresponding to that distance, as it is usually done. The first formula was shown to be related to the computation of the potential of mean force between two particles. It leads in practice to a computation on-the-fly that is quite different from the constraints/restraints method that is usually prescribed. The variance of the results is reduced by a large factor with respect to histogram techniques and was shown to be insensitive to the grid size. The second formula, that was associated to a zero-variance principle in the electronic density case [14,15], seems impaired by the large fluctuations of the second-order quantities Φ_i and did not prove useful at this stage. We think that the force approach is already very valuable in the context of ab-initio molecular dynamics simulations to improve the usually poor statistics due to short runs and small number of particles. It benefits moreover from the fact that in such simulations the forces on individual nuclei are already available.

The formalism has been extended to the computation of three-dimensional liquid densities around given molecular objects, quantities that are submitted to large statistical fluctuations when tackled by histogram techniques. Mere application of the Poisson identity for the three-dimensional Dirac function and successive integrations by parts, as proposed by Assaraf *et al* for electronic densities [14, 15], yield two formal expressions of the density, eqs. 18 and 19, which, to our knowledge, are unexploited in a statistical mechanics context. Straight application of the formulas requires the definition of a pre-determined 3D-grid and, for all grid points at each time step, an appropriate summation over all particles. The associated computation cost is largely affordable in an AIMD context. For classical simulations, in order not to impair the N^2 cost of the simulations, we have proposed an alternative, approximate scheme in which the mean forces (or mean second-order derivatives) are accumulated locally by histograms, and a final convolution is performed at the end with the help of 3D Fourier transforms. The first order scheme, using the forces, was shown to produce smooth densities with a much reduced variance compared to histograms. We hope to be able to clarify why the second-order scheme, aimed at giving a zero variance, turns out not to function so well in practice.

In any case, we think that the first order formalism described above will already prove very useful for *e.g.* the characterization of water densities around complex solutes such as protein, in order to estimate solvation free-energies, and to map local properties such as local entropies and enthalpies [16]. The introduction of the angular variables in the formalism in order to describe the molecules orientations in addition to their positions will be a necessary generalization in that context. Last but not least, extension to densities along general collective variables seems possible by using an auxiliary vector field as introduced in Ref. [7].

References

[1] M. Ferrario, G. Ciccotti, and K. Binder. *Computer Simulations in Condensed Matter: From Materials to Chemical Biology. Volume 1.* Lect. Notes Phys. 703, Springer, Berlin Heidelberg, 2006.

[2] M. Ferrario, G. Ciccotti, and K. Binder. *Computer Simulations in Condensed Matter: From Materials to Chemical Biology. Volume 2.* Lect. Notes Phys. 703, Springer, Berlin Heidelberg, 2006.

[3] S. Bonella, M. Monteferrante, C. Pierleoni, and G. Ciccotti. *J. Chem. Phys.*, 133:164104 and 164105, 2010.

[4] M. Monteferrante, S. Bonella, and G. Ciccotti. *J. Chem. Phys.*, 138:054118, 2013.

[5] G. Ciccotti and P. Echenique. *Eur. Phys. J.-Special Topics*, 200:1, 2011.

[6] C. Hartmann, C. Schuette, and G. Ciccotti. *J. Chem. Phys.*, 132:111103, 2010.

[7] G. Ciccotti, R. Kapral, and E. Vanden-Eijnden. *ChemPhysChem*, 6:1809, 2005.

[8] G. Ciccotti and S. Meloni. *Phys. Chem. Chem. Phys.*, 13:5952, 2011.

[9] S. Bonella, S. Meloni, and G. Ciccotti. *Eur. Phys. J. B*, 85:97, 2012.

[10] D. A. McQuarrie. *Statistical Mechanics.* University Science Books, Mill Valley, California, 2000.

[11] J. P. Hansen and I. R. McDonald. *Theory of Simple Liquids.* Academic Press, Elsevier, Oxford, 2006.

[12] M. P. Allen and D. J. Tildesley. *Computer simulation of liquids.* Oxford Science Publications, 1989.

[13] D. Frenkel and B. Smit. *Understanding Molecular Simulations: From Algorithms to Applications.* Academic Press, 2001.

[14] R. Assaraf, M. Caffarel, and A. Scemama. *Phys. Rev. E*, 75:035701, 2007.

[15] J. Toulouse, R. Assaraf, and C. J. Umrigar. *J. Chem. Phys.*, 126:24412, 2007.

[16] C. N. Nguyen, T. K. Young, and M. K. Gilson. *J. Chem. Phys.*, 137:044101, 2012.