RESEARCH ARTICLE | MAY 29 2026

# Direct Boltzmann inversion method from particle configurations at arbitrary state points

Olivier Coquand ; Davide Paolino ; Ludovic Berthier



*J. Chem. Phys.* 164, 204112 (2026)
https://doi.org/10.1063/5.0336283

The Journal of Chemical Physics ARTICLE pubs.aip.org/aip/jcp

# Direct Boltzmann inversion method from particle configurations at arbitrary state points

Cite as: J. Chem. Phys. 164, 204112 (2026); doi: 10.1063/5.0336283
Submitted: 30 March 2026 • Accepted: 29 April 2026 •
Published Online: 29 May 2026



Olivier Coquand,[1] Davide Paolino,[2] and Ludovic Berthier[2,a)]

## AFFILIATIONS

[1] Laboratoire de Modélisation Pluridisciplinaire et Simulations, Université de Perpignan Via Domitia, 52 avenue Paul Alduy, F-66860 Perpignan, France
[2] Gulliver, CNRS UMR 7083, ESPCI Paris, PSL Research University, 75005 Paris, France

[a)] **Author to whom correspondence should be addressed:** ludovic.berthier@espci.fr

## ABSTRACT

We introduce a direct Boltzmann inversion method to infer the interaction potential in particle systems using as input particle configurations generated at an arbitrary state point of the system. Unlike iterative Boltzmann inversion, the proposed method does not require performing a new Monte Carlo simulation at each step of the iteration process. It relies instead on enforcing consistency between two independent estimates of the pair correlation function, respectively obtained from interparticle distances and from pairwise forces. As a result, the approach is computationally inexpensive and straightforward to implement. Because it relies on the sole expression of interparticle forces, our method naturally applies to any state point, including when the density is large and alternative methods may fail. Here, we present the basic principles of the method and benchmark its performance on a diverse set of test potentials studied using computer simulations. Practical aspects and detailed implementation of the method are also discussed. Owing to its simplicity and generality, the method should be broadly applicable, from the construction of coarse-grained interaction potentials to the inference of effective interactions in non-equilibrium systems.





## I. INTRODUCTION

In liquid-state theory,[1] two cornerstone quantities are the interaction potential, $u(r)$, and the radial distribution function, $g(r)$. The first dictates the microscopic dynamics of each particle, while the second describes the structure. The latter can easily be obtained numerically or experimentally using, for example, scattering or microscopy techniques. In a classic paper, Henderson proved that for equilibrium systems governed by pairwise interactions, there exists a one-to-one correspondence between the pair potential and the radial distribution function.[2] A fundamental consequence is that knowledge of one function uniquely determines the other, at least in principle. As is well-known, there is no exact analytic formula to predict directly $g(r)$ from $u(r)$, and the derivation of approximate closure relations was a basic endeavor of liquid-state theory.[1] The parallel development of computer simulations of particle systems[3,4] has allowed for an essentially exact determination of $g(r)$ given any interaction potential $u(r)$. From a numerical viewpoint, the forward problem of determining $g(r)$ for a given $u(r)$ can therefore be considered as being essentially solved.

By contrast, the inverse problem[5] of computing the interaction potential from a known $g(r)$ has attracted somewhat less attention, and there exists no fully satisfying solution either analytically[1,6] or numerically. This so-called Boltzmann inversion problem is, nevertheless, of interest, as it can be used for several purposes. An obvious application is the determination of an unknown interaction potential from an experimental dataset.[7] For pairwise interactions, this can be performed by studying only two particles, but there is no such simple method available when an arbitrary bulk state point is accessible or in the presence of many-body interactions. Another application is the determination of an effective coarse-grained interaction potential representing a more complex system with a larger number of degrees of freedom. This is a useful task, as the determined coarse-grained model can then be studied numerically over much larger length scales and time scales.[8–10] A related application is the deduction of optimal effective pair potentials for systems interacting with non-pairwise additive interactions. A final application could be the determination of effective interactions in systems evolving far from equilibrium[11,12] in order to provide an equilibrium interpretation of physics phenomena observed far from equilibrium.

Over the years, a number of numerical methods have been developed to solve the inverse Boltzmann problem. Existing methods can be broadly divided into model-based and model-free approaches. Model-based methods essentially fit the measured radial distribution function by optimizing a parametrized family of potentials. Although widely used,[11,13,14] these approaches rely heavily on guessing an accurate parametrization and become less tractable as the number of parameters increases.[15] Model-free techniques, by contrast, do not impose a functional form for $u(r)$. These algorithms typically refine an initial potential guess iteratively until the simulated structural data match the reference one.[16] The main differences between different algorithms lie in the iteration rules,[17] whose complexity typically trades off with the number of iterations needed to converge. For example, inverse Monte Carlo[18] converges in relatively few steps, but this requires long simulations at each step for accurate statistics as well as a good initial guess.[19] In contrast, iterative Boltzmann inversion[10,20,21] relies on a very simple update rule, making it one of the most robust and widely used algorithms, applied in fields from liquid interfaces[22] to polymers[8] and biomolecules.[9] Its convergence is typically much slower than inverse Monte Carlo, often requiring an order of magnitude more iteration steps.[23] Recent proposals using machine learning to reconstruct interactions are promising[24] but still lack the accuracy of iterative algorithms[25] or must rely on intense training.[26]

A common feature of all iterative methods mentioned earlier is the need to perform a computer simulation at each iteration step using the updated potential in order to compare the generated data to the target $g(r)$. This represents the main computational bottleneck when performing the inversion, potentially limiting scalability to larger systems or higher accuracy.

Recently, an elegant strategy was proposed to bypass this bottleneck.[12,27–29] The method uses two expressions for the radial distribution function, the usual one using pairwise distances and a second one known as the test-particle insertion formula[1,30] that requires knowledge of the interaction potential. The quality of the updated potential is now assessed by comparing the radial distribution function generated by the test-particle insertion to the reference one coming from the initial dataset. An important shortcoming of this approach is that the test-particle insertion method is computationally much more expensive than conventional $g(r)$ calculations. An even more severe limitation is that particle insertions become virtually impossible in dense fluids, and the proposed approach is thus by construction limited to small enough densities to allow particle insertions.[28]

Here, we propose an alternative approach to directly obtain the interaction potential without performing a novel computer simulation at each step of the iteration, and this approach can be applied to any state point, including dense fluids. Our central idea is to replace the particle insertion expression of the radial distribution with an alternative expression based on interparticle forces.[31,32] This force-formula for $g(r)$ involves both interparticle distances and forces. Computationally, it is just as cheap as the conventional histogram method, and it is applicable and accurate at arbitrary state points, with no limit on the studied density. Our study, therefore, considerably simplifies and extends the applicability of the recent efforts performed in Refs. 27 and 29 to directly solve the Boltzmann inversion problem numerically.

Here, we explain how to implement this direct inversion method numerically to obtain the interaction potential from a given dataset that consists of a series of particle configurations. The method converges rapidly, accurately, and is computationally inexpensive. We implement and demonstrate its efficiency in a variety of situations for which the interaction potential is known, and we show that our inversion method accurately recovers the correct potential in all cases in just a few minutes, thus efficiently solving the Boltzmann inversion problem. We leave applications to experimental data, coarse-graining, and non-equilibrium situations for a future study.

The paper is organized as follows. The central idea behind the direct inversion method is presented in Sec. II. The practical implementation of the algorithm is discussed in Sec. III. We implement, test, and validate the algorithm on a number of known potentials at various state points in Sec. IV. We discuss the results and present some outlook for future study in Sec. V.

## II. DIRECT BOLTZMANN INVERSION ALGORITHM

### A. Radial distribution function (RDF)

A many-body system of $N$ particles can be completely characterized by the knowledge of all the $n$-particle distribution functions $g_N^{(n)}(\boldsymbol{r}_1,\ldots,\boldsymbol{r}_n)$.[1] The computation of all these quantities for $N>2$ is often difficult, but fortunately, the knowledge of low-order particle distribution functions $n\leq 2$ is often sufficient to evaluate the equation of state and other thermodynamic properties. In particular, if the system is both isotropic and homogeneous, the pair distribution function $g_N^{(2)}(\boldsymbol{r}_1,\boldsymbol{r}_2)$ only depends on the separation $r_{12}=|\boldsymbol{r}_2-\boldsymbol{r}_1|$. In this case, it is usually called the radial distribution function (RDF) and referred to as $g(r)$, taking the form

$$g(r)=\left\langle \frac{1}{\rho N}\sum_{i=1}^{N}\sum_{j\neq i}\delta(r-r_{ij})\right\rangle, \tag{1}$$

where $r_{ij}=|\boldsymbol{r}_i-\boldsymbol{r}_j|$, $\delta(x)$ is the Dirac delta function, $\rho=N/V$ is the number density, and $\langle\cdots\rangle$ represents an ensemble average.

### B. Conventional RDF estimate from pairwise distances

Given the configurations of a homogeneous system, the radial distribution function $g(r)$ can be computed using a distance histogram (DH) procedure that amounts to a simple discretization of Eq. (1). In practice, one counts the number of particles $N(r)$ within a spherical shell of radius $r$ and thickness $\Delta r$, normalized by the shell volume $\Omega_d r^{d-1}\Delta r$ and the particle density $\rho$, where $\Omega_d$ is the solid angle in $d$ dimensions ($\Omega_2=2\pi$ and $\Omega_3=4\pi$). Averaging over an ensemble of configurations $\langle\cdots\rangle$ yields

$$g_{DH}(r)=\frac{\langle N(r)\rangle}{\rho\Omega_d r^{d-1}\Delta r}. \tag{2}$$

Note that Eq. (2) depends explicitly on the choice of the bin size $\Delta r$, and the variance of the RDF obtained via this method diverges as $1/\Delta r^2$ for a given amount of data. The number of operations needed to measure $g_{DH}(r)$ scales quadratically with the number of particles,

$\mathcal{O}(N^2)$ (this is a sum over particle pairs), and it assumes no prior knowledge of the interparticle potential. Nevertheless, Henderson's theorem implies that the RDF $g(r)$ evaluated via interparticle distances is uniquely related to the potential $u(r)$ used to generate the configurations for pairwise interactions.[2]

### C. Computing the RDF from interparticle forces

Inspired by a previous study on quantum electronic densities, Borgis and co-workers proposed a new formula to estimate the pair correlation function of a homogeneous and isotropic system when the forces acting on each particle are known,[31]

$$g_{\text{force}}(r) = 1 - \frac{1}{\rho N}\left\langle \sum_{i<j} \beta(\mathbf{f}_i - \mathbf{f}_j)\cdot \frac{\boldsymbol{r}_{ij}}{\Omega_d r_{ij}^d}\Theta(r_{ij}-r)\right\rangle, \tag{3}$$

where $\Theta(x)$ is the Heaviside function and $\beta = 1/k_BT$, with $k_B$ the Boltzmann constant. In this expression, $\mathbf{f}_i$ represents the total force acting on particle $i$. This formula can be qualitatively understood as resulting from an integration by parts of the original expression in Eq. (1), transforming the delta function into a Heaviside one, while the derivative on the Boltzmann distribution produces the force $\mathbf{f}_i$. While this formula was originally obtained using precisely this integration by parts,[31] we provide an alternative derivation in Appendix A that first derives (producing the forces) and then re-integrates (producing the Heaviside functions) the function $g(r)$ with respect to the variable $r$. This derivation is close to the one given in Ref. 4. All derivations are, of course, mathematically equivalent.

In particular, the derivations make it clear that no assumption is made on the nature of the potential $u(r)$ (for instance, it is not necessarily pairwise additive). The formula in Eq. (3) differs from the original study[31] by a factor of 2, as already corrected in Ref. 33. An important assumption is that the system is at thermal equilibrium so that the ensemble average is represented by an integration of the configuration space, weighted by the Boltzmann distribution. This formula can, therefore, not be applied to a non-equilibrium driven steady state, for instance.

An important difference with respect to the histogram method is that a given pair of particles, $(i,j)$, contributes at all distances $r \le r_{ij}$ instead of contributing to a single bin at distance $r = r_{ij}$. This helps reduce the variance of the $g(r)$ estimate,[34] which no longer depends on the spatial discretization $\Delta r$ chosen to represent the RDF. As a result, the radial distribution function can be evaluated with an arbitrary spatial resolution without increasing the statistical noise.

It is important for the inversion method discussed in this study to realize that the evaluation of $g_{\text{force}}(r)$ at $r$ depends on all distances and is therefore no more local in position space. This unfortunately implies that a small error on the reconstructed potential $u(r)$ at some given $r$ may affect the function $g_{\text{force}}(r)$ at other $r$ values.

Another notable difference is that the function $g_{\text{force}}(r)$ obtained in Eq. (3) properly converges, by construction, to 1 in the limit $r \to \infty$ but displays a spurious non-zero value in the $r \to 0$ limit, which only vanishes in the limit of an infinite dataset.[34] These two differences will play an important role in the practical implementation discussed below. However, the key idea is that both $g_{\text{force}}(r)$ and $g_{\text{DH}}(r)$ provide unbiased estimates of the same analytical quantity $g(r)$, and both estimates must lead to the same result $g(r)$ in the limit of infinite statistics. The coincidence of both estimates, however, only happens when the correct expression of the forces is employed in Eq. (3). Enforcing this condition will allow us to directly determine forces from an ensemble of particle configurations.

### D. The direct inversion algorithm

We are now in a position to explain our direct inversion algorithm, which we have sketched in Fig. 1.

It is useful to recall first the traditional iterative Boltzmann inversion method. This starts by the determination of a reference RDF, $g_{\text{ref}}(r)$, obtained from the dataset using the distance histogram method. The iteration loop can then start. An initial guess of the interaction potential, $u_0(r)$, is used in a Monte Carlo simulation to determine the RDF $g_0(r)$ using the histogram method. The difference between $g_0$ and $g_{\text{ref}}(r)$ is used to make an improved guess for the potential, $u_1(r)$, leading to $g_1(r)$. The process is then repeated until the guess $u_t(r)$ made at step $t$ leads to a measured $g_t(r)$ that is equal to $g_{\text{ref}}$. At this point, the algorithm has reached a fixed point, and the potential $u_t(r)$ is no longer updated.

Our approach follows the same philosophy and starts with the construction of the reference RDF, $g_{\text{ref}}(r)$, and a guess, $u_0(r)$, is

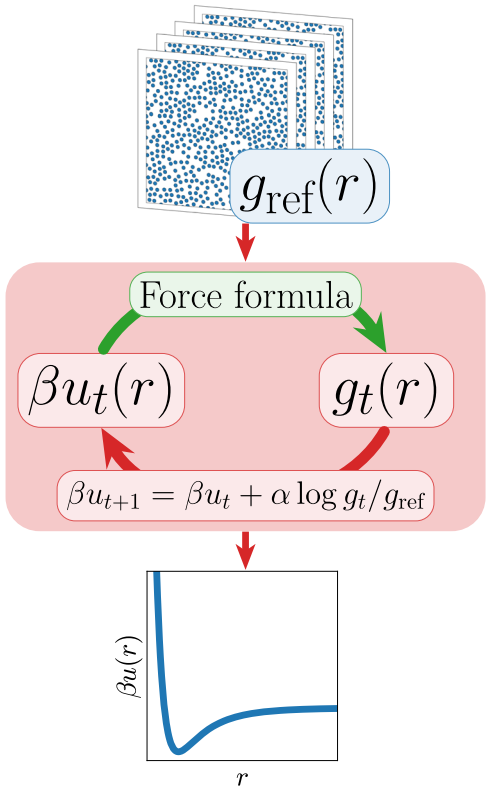


**FIG. 1.** Sketch of the direct Boltzmann inversion method. The available dataset (series of images) produces a reference RDF $g_{\text{ref}}(r)$. A guess potential $u_t(r)$ is used to generate the RDF $g_t(r)$ using the force formula, and the difference between $g_{\text{ref}}(r)$ and $g_t(r)$ is used to make an improved guess for the potential $u_{t+1}(r)$. The fixed point provides the correct inverted potential $u(r)$.

again made for the potential. Crucially, we now obtain the RDF $g_0(r)$ from the guess $u_0(r)$ using the force formula applied on the original dataset. (A very similar idea was proposed before by Stones and co-workers,[27] but they employed the test-particle insertion estimate of the RDF instead of the force formula.) The difference between $g_0$ and $g_{\text{ref}}(r)$ is used to make an improved guess for the potential, $u_1(r)$. The process is then repeated until convergence. In contrast to iterative Boltzmann inversion, the only dataset used in the iteration is the original one, as the process does not require the generation of new data via Monte Carlo simulations.

In all inversion schemes, an improved guess of the potential from the measured difference between two RDF estimates needs to be made. Here, we follow the update rule proposed by Schommers,[20]

$$\beta u_{t+1}(r) = \beta u_t(r) + \alpha \log\left(\frac{g_t(r)}{g_{\text{ref}}(r)}\right), \tag{4}$$

where $\alpha \in [0,1]$ is an empirical regularization factor used to control the stability of the iteration procedure. By construction, the potential $u_t(r)$ is not updated when the RDF at step $t$ agrees with the target reference.

This update formula remains, however, largely empirical. For consistency, we summarize Schommers' justification of Eq. (4) in Appendix B. It has the desired property that the fixed point $(g^*(r), u^*(r))$ necessarily corresponds to $u^*(r) = u(r)$ and $g^*(r) = g_{\text{ref}}(r)$. We discuss the stability of this fixed point in Appendix C.

## III. PRACTICAL IMPLEMENTATION

While the workflow presented in Fig. 1 faithfully describes all steps of the direct Boltzmann inversion algorithm, there are several practical issues that arise during its implementation, which we now discuss.

### A. Generation of reference RDF from data

The reference radial distribution function $g_{\text{ref}}(r)$ is constructed from the microscopic configurations available in the original dataset. At this point, the distance-histogram method in Eq. (2) needs to be used, since it does not require knowledge of the pair potential. As discussed in Sec. II B, the formula contains an adjustable parameter $\Delta r$ used to bin the pair distribution that requires a careful discussion.

Ideally, one would like $\Delta r$ to be as small as possible in order to reconstruct an interaction potential $u(r)$ on a sufficiently fine grid to allow the numerical determination of the derivative $u'(r)$ involved in force estimates. However, the distance-histogram RDF becomes noisy when $\Delta r$ is chosen too small for the amount of available data.

To mitigate these effects, we construct $g_{\text{ref}}(r)$ in two steps. We first use the conventional method that decreases $\Delta r$ until the RDF becomes noisy. We then smooth and interpolate the measured function to finally obtain a smooth, finely discretized reference RDF $g_{\text{ref}}(r)$. The smoothing and interpolation procedures are illustrated in Fig. 2.

In practice, we apply spline interpolation to the measured RDF. This reduces noise by fitting the histogram data with a continuous function that can then be evaluated at any desired resolution. Given a dataset $\mathcal{D} = \{(r_i, g_i)\}_{i=1}^{M}$, the spline is obtained by minimizing the following loss functional:

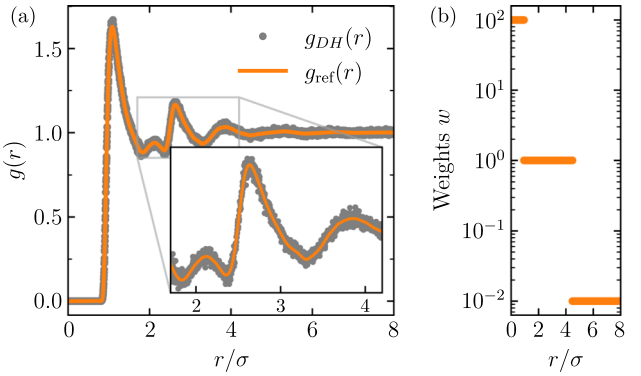


**FIG. 2.** (a) The RDF measured by the distance-histogram $g_{DH}(r)$ is smoothed and finely discretized to produce the reference $g_{\text{ref}}(r)$. (b) Weights used for the spline interpolation. The data was taken for the shoulder potential at $\rho\sigma^2 = 0.56$ and $\beta\epsilon = 0.5$.

$$\mathcal{L}[f] = \sum_{i=1}^{M} w_i |g_i - f(r_i)|^2 + \lambda \int \left|\frac{\partial^2 f}{\partial r^2}\right|^2 \mathrm{d}r, \tag{5}$$

using piecewise cubic polynomials for $f(r)$. The first term measures deviation between the fit function and the data points, while the second penalizes curvature, enforcing smoothness of the interpolated function $f(r)$. Here, $\lambda$ tunes the trade-off between data fidelity (first term) and smoothness (second term).

While conceptually easy, applying a single smoothing function across the entire radial distribution function is challenging because the RDF behaves very differently across different scales. At short distances, the hard-core region is undersampled and exhibits a sharp rise from zero that splines may fail to capture. At a larger range, the RDF peaks may suffer from statistical noise in the distance-histogram method, and this may obscure fine structural details. Finally, at long distances where $g(r) \approx 1$, noise dominates, making the interpolation potentially unreliable due to overfitting.

To address these issues, we implement a region-specific weighting scheme $w_i$ with three distinct zones: the hard-core region (from $r = 0$ to the midpoint to the first peak) receives a weight $10^2$ times stronger than the intermediate region with unit weight. Instead, the noisy long-range region is deliberately suppressed with a weight reduced to $10^{-2}$. Finally, the smoothing parameter $\lambda$ is determined automatically by the `scipi.signal` library[35] through generalized cross-validation. This approach maintains simplicity and transferability across diverse RDF shapes while delivering robust performances for the inversion procedure. As shown in Fig. 2, the obtained reference $g_{\text{ref}}(r)$ is simultaneously faithful to the measured data, smooth, and can be discretized over an arbitrary fine spatial grid.

### B. Choice of cutoff distances

Our goal is to reconstruct a function $u(r)$ finely discretized over a range of distances. While in principle we would like that range to be between $r = 0$ and $r = \infty$, in practice one needs to restrict this idealized range to a finite one and introduce two cutoff values to reconstruct $u(r)$ over a finite range $r \in [r_{\text{low}}, r_{\text{cut}}]$.

An upper cutoff $r_{cut}$ is needed when forces and potentials become very small at large distances. In that case, there will not be enough signal in the measured dataset to reconstruct very small numbers. In a typical application where that cutoff is not known, the chosen value could be refined iteratively by starting from a high value and progressively lowering $r_{cut}$ while monitoring stability and convergence of the inversion algorithm.

For strongly repulsive potentials, a lower bound $r_{low}$ is also needed. Since the hard core part of the potential is not explored by the particles, it is obvious that the dataset contains no information about the functional form of $u(r)$ at very short distances. On the other hand, we also find that taking $r_{low}$ too large can lead to an incorrect determination of the potential $u(r)$. The reason is related to the discussion in Sec. II C. When $r_{low}$ is too large, we miss pairs of particles in the dataset for which $r_{ij} < r_{low}$, and this may affect the entire function $g_{force}(r)$. In turn, an incorrect estimate of $g_{force}(r)$ will lead to an incorrect fixed point of the algorithm.

In practice, we go over the dataset and measure, over independent configurations, the minimum distance found in each configuration. We then set $r_{low}$ as the average of this quantity. This ensures that very few particle pairs are missed and that we do not attempt to reconstruct the potential in a region where statistics is poor.

Particle pairs with separations $r_{ij} < r_{low}$ are excluded by the Heaviside function in the calculation of $g_{force}(r)$ in Eq. (3). However, they still contribute significantly to the total forces $\mathbf{f}_i$ and $\mathbf{f}_j$. To account for these contributions, the derivative of the potential (that is needed for the force calculations) in the region $r < r_{low}$ is extrapolated as

$$u'(r) = u'(r_{low})\left(\frac{r_{low}}{r}\right)^2. \tag{6}$$

This expression ensures continuity of the force at $r = r_{low}$ and effectively introduces a soft-core repulsion that prevents nonphysical artifacts during the inversion process. We checked that the specific choice made in Eq. (6) is arbitrary and largely irrelevant, as several other choices can also be employed.

### C. Initialization of the potential

To run the iterative inversion scheme, one should finally choose an initial potential guess, $\beta u_0(r)$. While the final result should be independent of this initial choice, it is obvious that starting from a totally incorrect guess may lead to a much longer iterative process or even introduce numerical instabilities. Moreover, the Schommers scheme in Eq. (4) assumes that the initial potential guess is a sufficiently good approximation of the unknown pair potential.[20] For these reasons, typically, the starting potential is chosen as

$$\beta u_0(r) = -\ln g_{ref}(r), \tag{7}$$

which is also called the potential of mean-force. This guess coincides with the true potential in the case of weak interactions. This choice provides a model-agnostic initial guess, which, moreover, reproduces the most salient features of the true potential, thus being close enough in the sense of Schommers (see Appendix B). We adopt this choice in the following.

### D. Improved iteration formula to preserve physical constraints

Since the force-based estimator in Eq. (3) can yield spurious finite values as $r \to 0$,[34] the estimate $g_t(r)$ may, in particular, become negative during the inversion process, especially when the initial potential guess is poor. To prevent logarithmic singularities in the Schommers update rule (4), in that case, we apply a vertical shift to $g_t(r)$ to ensure the argument of the logarithm always remains positive.

Specifically, we define $r_{min}$ as the radial distance where the current estimate $g_t(r)$ reaches its minimum over the inversion window $[r_{low}, r_{cut}]$. We then implement the modified update rule

$$\beta u_{t+1} = \beta u_t + \alpha \log\left(\frac{g_t(r) - \delta g}{g_{ref}(r)}\right), \tag{8}$$

where the offset $\delta g \equiv g_t(r_{min}) - g_{ref}(r_{min})$ ensures that the numerator is well-behaved and matches the positive reference $g_{ref}(r)$ at its minimum value. In practice, $r_{min}$ typically coincides with $r_{low}$, the lower bound of the inversion window, as expected. Instances where $g_t(r)$ becomes negative at larger radial distances are rare and are generally restricted to the earliest iteration steps. As the potential converges toward a fixed point, we checked that $\delta g$ also becomes very small, and Eq. (8) becomes, in that case, mathematically equivalent to the conventional Schommers expression, in particular near the fixed point.

### E. Convergence metrics

Finally, an iterative algorithm requires a criterion to determine when convergence has been reached. Given some distance metric $D(f, g)$ between two functions $f(x)$ and $g(x)$ defined on a discretized series of $x$ values, one could theoretically stop at step $T$ when the current estimate $g_T(r)$ is closer to the target than a desired precision $\nu$, namely,

$$D(g_T, g_{ref}) \le \nu. \tag{9}$$

However, we specifically adopt a convergence criterion based on the distance between successive iterations,

$$D(g_T, g_{T-1}) \le \nu. \tag{10}$$

This second definition has several advantages: it is more easily transferable across different pair distribution shapes and does not assume that the distance to the reference decreases monotonically.

Because our method is computationally inexpensive, we have set a stringent precision requirement of $\nu = 10^{-10}$ to stop the iteration loop. This ensures the solution has reached the fixed point. This very low value eliminates any residual error that might exist when more costly methods are used and the iteration is stopped early.

In practice, we verified that both criteria yield nearly identical results, confirming the robustness of the fixed-point approach. The distance $D(f, g)$ is chosen as the mean-squared error

$$D(f,g) = \frac{1}{n}\sum_{i=1}^{n} |f(r_i) - g(r_i)|^2, \tag{11}$$

computed over the window $[r_{low}, r_{cut}]$ discretized in $n$ bins.

## IV. VALIDATION USING KNOWN POTENTIALS

The ultimate goal of our inversion algorithm is to reconstruct the pair potential $u(r)$ starting from equilibrium configurations of the system. As a first test, we use configurations obtained through numerical simulations of known potentials in order to assess the quality of the inversion method and detect possible pitfalls.

### A. Four simulated potentials

We analyze different two-dimensional systems of monodisperse particles of mass $m$, interacting with various widely used potentials, each one posing different challenges for their inversion. The first one is the conventional Lennard-Jones (LJ) potential,[36] truncated and shifted at $r_c^{\mathrm{LJ}} = 2.5\sigma$,

$$u_{\mathrm{LJ}}(r) = \begin{cases} 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} + C\right], & r \le r_c^{\mathrm{LJ}}, \\ 0 & \text{otherwise}, \end{cases} \tag{12}$$

where $C$ is an additive constant so that $\beta u_{\mathrm{LJ}}(r_c^{\mathrm{LJ}}) = 0$. The Lennard-Jones potential presents both a repulsive soft core and an attractive long-range tail.

The second tested potential is the Weeks–Chandler–Anderson (WCA) potential,[37] which is nothing but the LJ potential truncated at its minimum $r_c^{\mathrm{WCA}} = 2^{1/6}\sigma$,

$$u_{\mathrm{WCA}}(r) = \begin{cases} 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] + C', & r < r_c^{\mathrm{WCA}}, \\ 0 & \text{otherwise}, \end{cases} \tag{13}$$

where $C' = \epsilon$ is such that $u_{\mathrm{WCA}}(r_c) = 0$, thus creating a purely repulsive potential that is often used in high-density regimes where the role of attractive forces diminishes.

Third, we introduce a long-range power-law potential that scales as the inverse of the cubic distance,

$$u_{R3}(r) = \begin{cases} \epsilon\left(\frac{\sigma}{r}\right)^{3} + C'' & r \le r_c^{R3} = 5\sigma, \\ 0 & \text{otherwise}, \end{cases} \tag{14}$$

where again $C''$ is an additive constant to ensure $u_{R3}(r_c^{R3}) = 0$. Such an $r^{-3}$ potential describes softer repulsive interactions compared to LJ and WCA, with slower decay relevant for dipole-like or screened electrostatic effects, common in soft colloidal systems.[38]

Finally, we introduce a shoulder potential[39]

$$u_{\mathrm{SH}}(r) = \begin{cases} \epsilon\left(\frac{\sigma}{r}\right)^{n} - \frac{\epsilon}{2}\tanh\frac{k_0}{\sigma}(r - r_0) + C''', & r < r_c^{\mathrm{SH}}, \\ 0 & \text{otherwise}, \end{cases} \tag{15}$$

that features two characteristic length scales. It has a hard core $r^{-n}$ and an outer softer shell represented by the hyperbolic tangent. The parameters are fixed as in Ref. 29, namely, $n = 14$, $k_0 = 10$, and $r_0 = 2.5\sigma$ with a cutoff at $r_c^{\mathrm{SH}} = 2.8\sigma$. Once more, the potential is shifted to vanish at its cutoff, fixing the value of the constant $C'''$.

We consider square systems in two dimensions with periodic boundary conditions and box size $L = 60\sigma$, while the number of particles $N$ and the temperature $T$ vary between different simulations. The values for $\rho = N/L^2$, the temperature $T$, and other simulation parameters will be reported in the figure captions accompanying the inversion results.

In practice, we first equilibrate and then sample equilibrium configurations in the canonical ensemble via molecular dynamics simulations using the open-source software LAMMPS.[40] Equations of motion are integrated using the Verlet algorithm, and the temperature is fixed via the Nosé–Hoover thermostat.[3] For each potential, the target radial distribution function $g_{\mathrm{ref}}(r)$ is computed from 500 independent configurations using the distance-histogram method.



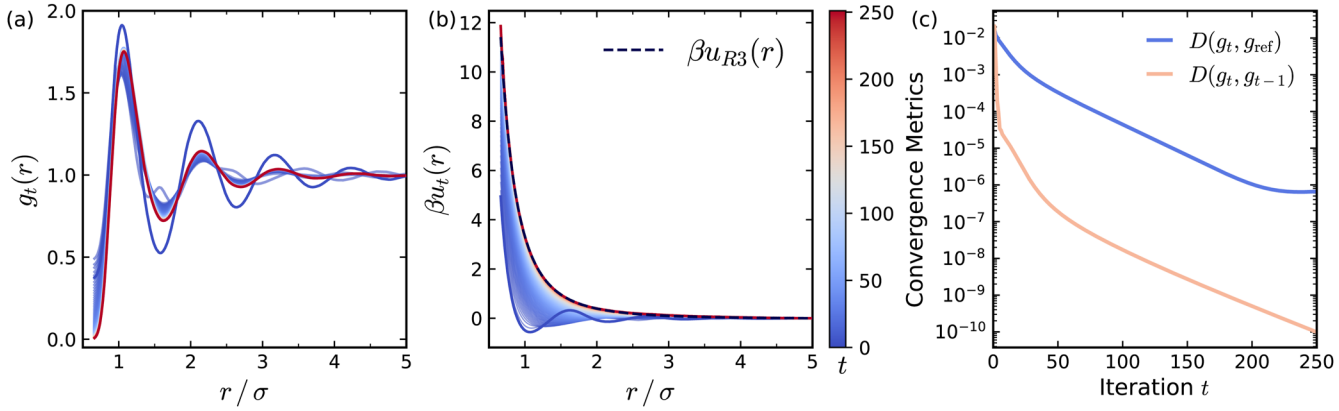


**FIG. 3.** Convergence of the iterative procedure for an inverse cubic potential at $\rho\sigma^2 = 0.80$ and $\beta\epsilon = 10/3$, using $\alpha = 0.5$. (a) Evolution of $g_t(r)$ toward the reference (red line). (b) Effective potentials $\beta u_t(r)$ reconstructed on the window $[0.663\sigma, 5\sigma]$; the dashed line shows the analytical potential and the thick blue line indicates the initial guess $\beta u_0 = -\log g_{\mathrm{ref}}(r)$. (c) Convergence metrics as a function of the iteration number $t$.

### B. Inversion algorithm at work

Having obtained the reference RDF $g_{\text{ref}}(r)$ as explained in Sec. III A, we now enter the iteration loop. In this part, we use 125 independent configurations produced by the molecular dynamics simulations.

We illustrate the results in Fig. 3 for the inverse cubic potential $u_{R3}(r)$. We use $\alpha = 0.5$ and reconstruct the potential over the range $[0.663\sigma, 5\sigma]$, using the potential of mean force as the initial guess, $\beta u_0(r) = -\log g_{\text{ref}}(r)$. We show the evolution of both the radial distribution function $g_t(r)$ [Fig. 3(a)] and the reconstructed potential $\beta u_t(r)$ [Fig. 3(b)] as the iteration progresses. Convergence is monitored using both $D(g_t, g_{\text{ref}})$ and $D(g_t, g_{t+1})$ [see Fig. 3(c)].

From the evolution of the potential, we observe that the nonphysical double-well present in the initial guess is rapidly lost after a few iteration steps. Nonetheless, matching the precise values and functional form of the reference potential requires many additional iterations.

Convergence metrics show that both the difference between successive iterations and the difference relative to the target drop below ten parts per million after roughly 160 iterations. Notably, the former drops by a further two orders of magnitude in twenty steps, signaling the arrest of the iteration procedure and the approach to the fixed point.

In terms of computational performance, the entire procedure for this system requires about ten minutes on a standard laptop. Each step of the iteration simply requires the calculation of the radial distribution function over 125 configurations composed of $N = 2916$ particles each. This could even be shortened if a less precise inverted potential were needed, but this method therefore represents a computationally very cheap inversion method. To achieve a similar performance using iterative Boltzmann inversion would have required performing 200 long and independent Monte Carlo simulations at each iteration step.

### C. Results for other potentials

We now collect our results for the different interaction potentials introduced in Sec. IV A. In order to ease the discussion, we set the inversion cutoff $r_{\text{cut}}$ to match the simulation cutoff $r_c$ for each potential. We checked that increasing $r_{\text{cut}}$ beyond $r_c$ does not affect the quality of the inversion, with converged potential curves that coincide well with the ones presented here.

The long-range $r^{-3}$ potential, often problematic for inverse methods due to its slow decay, is also reconstructed, as already shown in Fig. 3.

All the other results are shown in Fig. 4. The overall agreement between the reconstructed and target potentials is excellent, even for particularly challenging systems.

Both the Lennard-Jones potential [Fig. 4(a)], characterized by steep short-range repulsion and a shallow attractive tail, and the WCA potential [Fig. 4(b)], which is non-zero over a very narrow region, are accurately resolved, highlighting the robustness

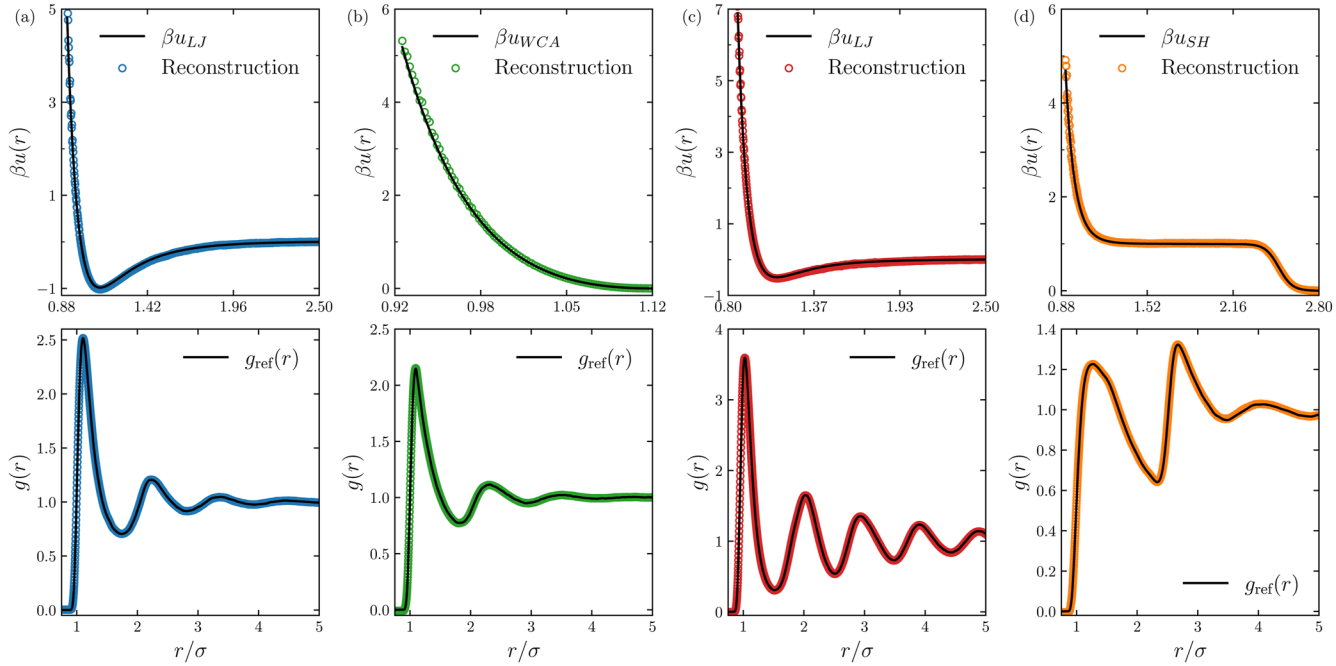


**FIG. 4.** Potential reconstruction results for various interaction types. The top row displays the reconstructed pair potentials $\beta u(r)$, while the bottom row shows the corresponding radial distribution functions $g(r)$. In all panels, the solid black lines represent the reference, and the colored scatter points denote the final reconstruction results: (a) Lennard-Jones (LJ) at $\rho\sigma^2 = 0.56, \beta\epsilon = 1.0, \alpha = 0.4$, reconstructed over the window $[0.915\sigma, 2.5\sigma]$; (b) Weeks–Chandler–Andersen (WCA) at $\rho\sigma^2 = 0.56, \beta\epsilon = 1.0, \alpha = 0.2$, reconstruction window $[0.920\sigma, 2^{1/6}\sigma]$; (c) LJ at $\rho\sigma^2 = 0.92, \beta\epsilon = 1/2, \alpha = 0.5$, reconstruction window $[0.863\sigma, 2.5\sigma]$; (d) Shoulder potential at high density $\rho\sigma^2 = 0.28, \beta\epsilon = 1.0, \alpha = 0.2$, reconstruction window $[0.911\sigma, 2.8\sigma]$.

of the method in handling both attractive regimes and short-range interactions.

Remarkably, the method remains robust even at high densities [ $\rho\sigma^2 = 0.92$, see Fig. 4(c)], accurately recovering the potential in a state point where the test-particle insertion method would typically encounter severe convergence issues.[28]

Finally, the shoulder potential [Fig. 4(d)], featuring both a steep repulsive core and an intermediate plateau, is recovered, demonstrating the ability of the method to capture potentials with multiple intrinsic length scales.

Across all test cases, potential reconstructions exceed expectations, with discrepancies appearing primarily in the derived force profiles. However, this limitation comes from the numerical differentiation used to compute forces. In applications where the goal is force reconstruction, additional post-processing, such as smoothing or filtering of the underlying potential, could significantly improve the quality of the extracted force field, if needed.

## V. SUMMARY AND OUTLOOK

In conclusion, we have introduced a simple, robust, computationally cheap, and flexible Boltzmann inversion algorithm capable of reconstructing pair potentials directly from an ensemble of configurations obtained at arbitrary state points.

The method exploits the robustness of iterative Boltzmann inversion, but a substantial reduction in computational cost is achieved by replacing expensive Monte Carlo simulations at each iteration step with a simple evaluation of the RDF using the force formula over the initial dataset. This modification enables hundreds of iterations within seconds or minutes, which provides superior accuracy compared to traditional inversion schemes. The benchmark tests contained in our paper demonstrate that the method successfully recovers interaction potentials across a wide variety of physical situations, including steeply repulsive cores, short-ranged interactions, and potentials with multiple or very large characteristic length scales.

Beyond its technical success, the approach holds significant promise for practical applications. Its efficiency makes it well-suited for analyzing experimental data, where structural measurements are often available but the underlying interactions may remain elusive. Moreover, it provides a viable tool for probing non-equilibrium systems through effective equilibrium mappings and for developing coarse-grained models in soft matter and biomolecular contexts. In the latter case, an explicit derivation starting from the full many-body Hamiltonian would be requested. For these systems, our method would be particularly valuable, since the coarse-graining could easily be performed at any state point, and the evolution of effective interactions with thermodynamic parameters could thus be followed explicitly.[41] All these directions will be the focus of future study, where we plan to apply the algorithm to experimental systems, non-equilibrium active matter, and coarse-graining problems.

## ACKNOWLEDGMENTS

We acknowledge R. Jack for useful discussions, D. Aarts and C. Valeriani for exchanges over their related study, F. van Wijland for discussions related to the stability of the algorithm, and D. Frenkel and B. Guiselin for their interest in the force formula. We acknowledge Yuan Liu, Qiuju Chen, Xurui Li, and Jianxiang Tian for the study performed in the early stages of this project.[42] We acknowledge the financial support of the French Agence Nationale de la Recherche (ANR), under Grant Nos. ANR-20-CE30-0031 (project THEMA) and ANR-24-CE30-0442 (project GLASSGO).

## AUTHOR DECLARATIONS

### Conflict of Interest

The authors have no conflicts to disclose.

### Author Contributions

**Olivier Coquand**: Investigation (equal); Writing – original draft (equal). **Davide Paolino**: Investigation (equal); Writing – original draft (equal). **Ludovic Berthier**: Investigation (equal); Writing – original draft (equal).

## DATA AVAILABILITY

The data that support the findings of this study are openly available on Zenodo.[43]

## APPENDIX A: ANOTHER DERIVATION OF THE FORCE FORMULA

The radial distribution function (RDF) provides a statistical description of fluid structure. Its definition is

$$g(\boldsymbol{r}) = \frac{1}{\rho N}\left\langle \sum_{i\neq j} \delta(\boldsymbol{r} - \boldsymbol{r}_{ij}) \right\rangle, \qquad \boldsymbol{r}_{ij} = \boldsymbol{r}_i - \boldsymbol{r}_j, \tag{A1}$$

where $\rho$ is the number density and $N$ is the number of particles. The configurational average is taken over the Boltzmann distribution for a system with potential energy $U = U(\{\boldsymbol{r}_k\}, k = 1 \dots N)$,

$$\langle \cdots \rangle = \frac{1}{Z}\int \mathrm{d}^N\boldsymbol{r}\ (\cdots)\, e^{-\beta U}, \tag{A2}$$

$$Z = \int \mathrm{d}^N\boldsymbol{r}\ e^{-\beta U}, \tag{A3}$$

where $\beta = 1/k_B T$.

We compute the radial derivative $\partial_r g(\boldsymbol{r})$ by differentiating inside the expectation value,

$$\partial_r g(\boldsymbol{r}) = \frac{1}{\rho N}\sum_{i\neq j} \langle \partial_r \boldsymbol{r} \cdot \nabla_{\boldsymbol{r}} \delta(\boldsymbol{r} - \boldsymbol{r}_{ij}) \rangle. \tag{A4}$$

Using $\partial_r \boldsymbol{r} = \boldsymbol{r}/r$ and exploiting symmetry between indices $i$ and $j$, we express this as

$$\partial_r g(\boldsymbol{r}) = \frac{1}{\rho N}\sum_{i\neq j} \left\langle \frac{\boldsymbol{r}}{r} \cdot \frac{(\nabla_{\boldsymbol{r}_j} - \nabla_{\boldsymbol{r}_i})}{2} \delta(\boldsymbol{r} - \boldsymbol{r}_{ij}) \right\rangle. \tag{A5}$$

Splitting the expectation value into two contributions and focusing only on one of them, we integrate by parts,

$$\left\langle \frac{\boldsymbol{r}}{r} \cdot \frac{1}{2}\nabla_{\boldsymbol{r}_j} \delta(\boldsymbol{r} - \boldsymbol{r}_{ij}) \right\rangle \tag{A6}$$

$$= \int \frac{\mathrm{d}^N\boldsymbol{r}}{Z} e^{-\beta U} \frac{\boldsymbol{r}}{r} \cdot \frac{1}{2}\nabla_{\boldsymbol{r}_j} \delta(\boldsymbol{r} - \boldsymbol{r}_{ij}) \tag{A7}$$

$$= -\int \frac{\mathrm{d}^N \boldsymbol{r}}{Z} \frac{1}{2} \nabla_{\boldsymbol{r}_j} e^{-\beta U} \cdot \frac{\boldsymbol{r}}{r} \delta(\boldsymbol{r} - \boldsymbol{r}_{ij}) \tag{A8}$$

$$= -\int \frac{\mathrm{d}^N \boldsymbol{r}}{Z} \frac{\beta}{2} \mathbf{f}_j e^{-\beta U} \cdot \frac{\boldsymbol{r}}{r} \delta(\boldsymbol{r} - \boldsymbol{r}_{ij}), \tag{A9}$$

where $\mathbf{f}_j = -\nabla_{\boldsymbol{r}_j} U$ is the total force acting on particle $j$. The boundary term is zero assuming that the Boltzmann weight (and, therefore, the pair potential) decays sufficiently fast. A similar expression holds for the second term with $\nabla_{\boldsymbol{r}_i}$. Combining both contributions, we find

$$\partial_r g(\boldsymbol{r}) = \frac{1}{\rho N} \sum_{i \neq j} \left\langle \beta \frac{(\mathbf{f}_i - \mathbf{f}_j)}{2} \cdot \frac{\boldsymbol{r}}{r} \delta(\boldsymbol{r} - \boldsymbol{r}_{ij}) \right\rangle. \tag{A10}$$

For homogeneous, isotropic systems, the RDF depends only on the radial distance $r = |\boldsymbol{r}|$. We extract this radial dependence by averaging over solid angles,

$$\int_{\mathbb{S}_{d-1}} \mathrm{d}\boldsymbol{\Omega}_{\hat{\boldsymbol{r}}} \partial_r g(\boldsymbol{r}) = \Omega_d \partial_r g(r) = \Omega_d g'(r), \tag{A11}$$

where $\Omega_d = 2\pi^{d/2}/\Gamma(d/2)$ is the surface of the $d$-dimensional unit sphere. At this point, the only term that depends on the angular components is the Dirac distribution, which satisfies the identity,

$$\int_{\mathbb{S}_{d-1}} \mathrm{d}\boldsymbol{\Omega}_{\hat{\boldsymbol{r}}}\ \delta(\boldsymbol{r} - \boldsymbol{r}_{ij}) = \frac{1}{r^{d-1}} \delta(r - r_{ij}). \tag{A12}$$

Altogether, using the symmetry between $i$ and $j$ to restrict the sum, we obtain

$$g'(r) = \frac{1}{\rho N} \left\langle \sum_{i<j} \beta(\mathbf{f}_i - \mathbf{f}_j) \cdot \frac{\boldsymbol{r}_{ij}}{\Omega_d r_{ij}^d} \delta(r - r_{ij}) \right\rangle. \tag{A13}$$

We can then integrate from 0 to $r$, and noting that $g(0) = 0$ for quasi-hard core potentials, we obtain

$$g(r) = \frac{1}{\rho N} \left\langle \sum_{i<j} \beta(\mathbf{f}_i - \mathbf{f}_j) \cdot \frac{\boldsymbol{r}_{ij}}{\Omega_d r_{ij}^d} \Theta(r - r_{ij}) \right\rangle. \tag{A14}$$

Alternatively, one can integrate $g'(r)$ from $r$ to $\infty$ and impose the condition $g(r \to \infty) = 1$, which holds for systems lacking long-range order,

$$g(r) = 1 - \frac{1}{\rho N} \left\langle \sum_{i<j} \beta(\mathbf{f}_i - \mathbf{f}_j) \cdot \frac{\boldsymbol{r}_{ij}}{\Omega_d r_{ij}^d} \Theta(r_{ij} - r) \right\rangle. \tag{A15}$$

These expressions are the two equivalent forms of the Borgis formula, expressing the RDF directly in terms of interparticle forces. These expressions differ from the original study[31] by a factor 2 that was omitted, as already pointed out in Ref. 33.

These expressions do not rely on a specific form of the potential energy $U(\{\boldsymbol{r}_k\})$, other than its isotropy and the presence of a (quasi-)hard core, only on the existence of well-defined forces $\mathbf{f}_i$ acting on each particle. Therefore, these relations hold for any system where forces can be defined, including those with many-body interactions or non-pairwise potentials.

## APPENDIX B: JUSTIFICATION OF ITERATION FORMULA

To ensure the self-consistency of the paper, we summarize the key arguments from Schommers[20] that underpin the iterative Boltzmann inversion approach and discuss their implications for the reliability of the algorithm. While the mathematical foundations of the inverse problem have been studied extensively,[16] a general and rigorous convergence proof for the IBI algorithm itself is lacking.

For simplicity, we study in dimensionless units, absorbing $\beta$ into the potential $U(r)$. Consider a system with a true pair interaction potential $U^{(0)}(r)$ and its corresponding pair correlation function $g^{(0)}(r)$. These are related via the cavity function $\gamma^{(0)}(r)$, defined as

$$g^{(0)}(r) = \gamma^{(0)}(r) e^{-U^{(0)}(r)}, \tag{B1}$$

where $\gamma(r)$ captures all correlations beyond the potential of mean force. Now, suppose we have guessed a potential $U^{(1)}(r) \neq U^{(0)}(r)$, which yields a pair correlation function $g^{(1)}(r)$ through a corresponding $\gamma^{(1)}(r)$. By Henderson's uniqueness theorem, the resulting $g^{(1)} \neq g^{(0)}$. However, if the $\gamma$ functions are close, i.e., $\gamma^{(1)}(r) = \gamma^{(0)}(r) + \Delta\gamma(r)$ with $|\Delta\gamma(r)| \ll |\gamma^{(0)}(r)|$, then we can write

$$\gamma^{(1)}(r) = g^{(1)}(r) e^{U^{(1)}(r)} = g^{(0)}(r) e^{U^{(0)}(r)} + \Delta\gamma(r). \tag{B2}$$

Rearranging and taking logarithms gives

$$U^{(1)}(r) - U^{(0)}(r) = \log\left(\frac{g^{(0)}(r)}{g^{(1)}(r)}\right) + \log\left(1 + \frac{\Delta\gamma(r)}{\gamma^{(0)}(r)}\right). \tag{B3}$$

Expanding to first order leads to

$$U^{(1)}(r) - U^{(0)}(r) \approx \log\left(\frac{g^{(0)}(r)}{g^{(1)}(r)}\right) + \mathcal{O}\left(\frac{\Delta\gamma}{\gamma^{(0)}}\right). \tag{B4}$$

This relation provides two useful insights. First, it connects the difference between the true and guessed potentials to the ratio of their pair correlation functions, motivating the IBI update rule Eq. (4). Second, the error in this estimate is controlled by the relative deviation $\Delta\gamma/\gamma^{(0)}$, rather than by the difference between the potentials themselves. This means that convergence hinges on the similarity of the many-body environments between iterations, not on an initially accurate guess for $U(r)$.

The condition of closeness in $\gamma$, rather than in $U$, is significantly weaker and more practical. For example, in the low-density limit, the virial expansion gives $\gamma(r) = 1 + \mathcal{O}(\phi)$, where $\phi$ is the packing fraction. Therefore, for dilute systems, $\gamma(r)$ is close to unity regardless of the details of the potential, making the iterative correction reliable. This explains why initializing IBI with $U(r) = 0$ often works for dilute systems.[27]

For denser systems, one may worry that strong many-body correlations could cause $\gamma(r)$ to differ significantly between iterations, potentially hindering or destabilizing convergence. To address this concern, we recall the arguments by Soper,[44] who addressed the question of convergence at arbitrary density. First of all, using

the notation $\Delta U = U^{(1)} - U^{(0)}$ and $\Delta g = g^{(1)} - g^{(0)}$, Henderson's theorem provides a global constraint,

$$\int \Delta U(r)\Delta g(r)\, d^d r < 0, \tag{B5}$$

where the integral is over all space. This inequality indicates that, on average, it is possible to predict the sign of the variation in the pair correlation function given a change in the potential, and vice versa. The relative variation of $g$ and $U$ is at the core of the iteration rule in IBI.

To understand it better, one can decompose $g(r)$ as $g(r) = g^{(p)}(r) + g^{(m)}(r)$, where $g^{(p)}(r) = e^{-U(r)}$ is the dilute contribution and $g^{(m)}(r)$ captures many-body effects via the potential of mean force $W(r)$: $g^{(m)}(r) = e^{-W(r)} - e^{-U(r)}$. It is easy to show that, for all $r$, $\Delta U(r)\Delta g^{(p)}(r) < 0$ so that

$$\int \Delta U(r)\Delta g^{(p)}(r)\, d^d r < 0. \tag{B6}$$

Combining this with the previous equation yields a bound on the many-body contribution,

$$\int \Delta U(r)\Delta g^{(m)}(r)\, d^d r < -\int \Delta U(r)\Delta g^{(p)}(r)\, d^3 r. \tag{B7}$$

This shows that the relative variations of the many-body part $\Delta g^{(m)}$ and that of $\Delta U$ at a given $r$ are not directly related. Yet, the variation of the many-body contribution $g^{(m)}$ with $\Delta U$ is bounded by a positive number controlled by the dilute limit. For this reason, the use of the potential of the mean force in Schommers' update rule in Eq. (4) is justifiable even at finite density, although a rigorous convergence proof remains elusive.

## APPENDIX C: STABILITY ANALYSIS

Even in the absence of a convergence proof, one can show that the iteration is stable. Namely, one can show that a small perturbation of the potential near the correct fixed point disappears by iterating Eq. (4), rather than exploding. To establish stability, we adopt a continuous-time view of the iteration. The iterative update can then be rewritten as

$$\frac{\mathrm{d}U_t(r)}{\mathrm{d}t} = \alpha \log\left(\frac{g_t(r)}{g_{\mathrm{ref}}(r)}\right). \tag{C1}$$

Here, $t$ is now a real variable, and the positive constant $\alpha$ (often introduced for numerical convergence) does not affect the stability analysis. The radial distribution function $g_t(r)$ is a functional of the current potential $U_t$,

$$g[U](r) = e^{-W[U](r)} = e^{-U(r)}\gamma[U](r), \tag{C2}$$

where $W[U](r)$ is the potential of mean force and $\gamma[U](r)$ the cavity distribution function[1] associated with $g[U](r)$.

Now consider a potential close to the target $U_{\mathrm{ref}}(r)$,

$$U(r) = U_{\mathrm{ref}}(r) + \delta U(r), \qquad \frac{|\delta U|}{|U_{\mathrm{ref}}|} \ll 1. \tag{C3}$$

Since $g[U_{\mathrm{ref}}](r) = g_{\mathrm{ref}}(r)$, the evolution of the perturbation is

$$\begin{aligned}\frac{\mathrm{d}\delta U(r)}{\mathrm{d}t} &= \alpha \log\left(\frac{g[U_{\mathrm{ref}} + \delta U](r)}{g_{\mathrm{ref}}(r)}\right) \\ &= \alpha \log\left(\frac{e^{-U_{\mathrm{ref}}-\delta U}\,\gamma[U_{\mathrm{ref}} + \delta U](r)}{e^{-U_{\mathrm{ref}}}\,\gamma[U_{\mathrm{ref}}](r)}\right) \\ &= -\alpha\,\delta U(r) + \alpha \log\left(\frac{\gamma[U_{\mathrm{ref}} + \delta U](r)}{\gamma[U_{\mathrm{ref}}](r)}\right) + O(\delta U^2).\end{aligned} \tag{C4}$$

Expanding $\gamma$ to first order in $\delta U$, we obtain

$$\frac{\mathrm{d}\delta U(r)}{\mathrm{d}t} = -\alpha\,\delta U + \alpha\,\frac{1}{\gamma[U_{\mathrm{ref}}]}\,\frac{\delta\gamma}{\delta U}\bigg|_{U_{\mathrm{ref}}}\delta U + O(\delta U^2). \tag{C5}$$

In the low-density limit, $\gamma[U](r) \approx 1$ for all potentials; hence, $\delta\gamma/\delta U \approx 0$, and we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\delta U \simeq -\alpha\,\delta U, \tag{C6}$$

giving

$$\delta U(r) = \delta U_0(r)\, e^{-\alpha t}. \tag{C7}$$

Therefore, at low density, any sufficiently small initial perturbation decays exponentially, confirming the stability of the iteration near the fixed point. As for the convergence formula mentioned earlier, it is difficult to extend this argument beyond the dilute limit, which thus simply serves as a guide for denser systems.

## REFERENCES


[1] J.-P. Hansen and I. R. McDonald, *Theory of Simple Liquids* (Academic Press, Oxford, 2013).
[2] R. L. Henderson, "A uniqueness theorem for fluid pair correlation functions," Phys. Lett. A **49**, 197 (1974).
[3] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, 2nd ed. (Oxford University Press, Oxford, 2017).
[4] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications* (Elsevier, 2023).
[5] J. T. Chayes, L. Chayes, and E. H. Lieb, "The inverse problem in classical statistical mechanics," Commun. Math. Phys. **93**, 57 (1984).
[6] S. Chakrabarty and Z. Nussinov, "High-temperature correlation functions: Universality, extraction of exchange interactions, divergent correlation lengths, and generalized Debye length scales," Phys. Rev. B **84**, 064124 (2011).
[7] C. P. Royall, A. A. Louis, and H. Tanaka, "Measuring colloidal interactions with confocal microscopy," J. Chem. Phys. **127**, 044507 (2007).
[8] G. Maurel, F. Goujon, B. Schnell, and P. Malfreyt, "Multiscale modeling of the polymer–silica surface interaction: From atomistic to mesoscopic simulations," J. Phys. Chem. C **119**, 4817 (2015).
[9] H. I. Ingólfsson, C. A. Lopez, J. J. Uusitalo, D. H. de Jong, S. M. Gopal, X. Periole, and S. J. Marrink, "The power of coarse graining in biomolecular simulations," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **4**, 225 (2014).
[10] T. C. Moore, C. R. Iacovella, and C. McCabe, "Derivation of coarse-grained potentials via multistate iterative Boltzmann inversion," J. Chem. Phys. **140**, 224104 (2014).
[11] S. Torquato and H. Wang, "Precise determination of pair interactions from pair statistics of many-body systems in and out of equilibrium," Phys. Rev. E **106**, 044122 (2022).

[12]C. R. Rees-Zimmerman, C. M. Barriuso Gutierrez, C. Valeriani, and D. G. A. L. Aarts, "Effective interactions in active Brownian particles," Soft Matter **22**, 803 (2026).
[13]H. Wang and S. Torquato, "Equilibrium states corresponding to targeted hyperuniform nonequilibrium pair statistics," Soft Matter **19**, 550 (2023).
[14]J. Tian and L. Berthier, "Determination of pairwise interactions via the radial distribution function in equilibrium systems interacting with the Mie potential," Results Phys. **52**, 106782 (2023).
[15]S. Izvekov, M. Parrinello, C. J. Burnham, and G. A. Voth, "Effective force fields for condensed phase systems from ab initio molecular dynamics simulation: A new method for force-matching," J. Chem. Phys. **120**, 10896 (2004).
[16]F. Delbary, M. Hanke, and D. Ivanizki, "A generalized Newton iteration for computing the solution of the inverse Henderson problem," Inverse Probl. Sci. Eng. **28**, 1166 (2020).
[17]M. P. Bernhardt, M. Hanke, and N. F. A. van der Vegt, "Iterative integral equation methods for structural coarse-graining," J. Chem. Phys. **154**, 084118 (2021).
[18]A. P. Lyubartsev and A. Laaksonen, "Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach," Phys. Rev. E **52**, 3730 (1995).
[19]T. Murtola, E. Falck, M. Karttunen, and I. Vattulainen, "Coarse-grained model for phospholipid/cholesterol bilayer employing inverse Monte Carlo with thermodynamic constraints," J. Chem. Phys. **126**, 075101 (2007).
[20]W. Schommers, "A pair potential for liquid rubidium from the pair correlation function," Phys. Lett. A **43**, 157 (1973).
[21]D. Reith, M. Pütz, and F. Müller-Plathe, "Deriving effective mesoscale potentials from atomistic simulations," J. Comput. Chem. **24**, 1624 (2003).
[22]M. Jochum, D. Andrienko, K. Kremer, and C. Peter, "Structure-based coarse-graining in liquid slabs," J. Chem. Phys. **137**, 064102 (2012).
[23]S. Jain, S. Garde, and S. K. Kumar, "Do inverse Monte Carlo algorithms yield thermodynamically consistent interaction potentials?," Ind. Eng. Chem. Res. **45**, 5614 (2006).
[24]F. Berressem and A. Nikoubashman, "Boltzmann: Predicting effective pair potentials and equations of state using neural networks," J. Chem. Phys. **154**, 124123 (2021).
[25]M. Ruiz-Garcia, C. M. Barriuso G., L. C. Alexander, D. G. A. L. Aarts, L. M. Ghiringhelli, and C. Valeriani, "Discovering dynamic laws from observations: The case of self-propelled, interacting colloids," Phys. Rev. E **109**, 064611 (2024).
[26]S. Bag and R. Mandal, "Interaction from structure using machine learning: In and out of equilibrium," Soft Matter **17**, 8322 (2021).
[27]A. E. Stones, R. P. A. Dullens, and D. G. A. L. Aarts, "Model-free measurement of the pair potential in colloidal fluids using optical microscopy," Phys. Rev. Lett. **123**, 098002 (2019).
[28]C. R. Rees-Zimmerman, A. Heafield, D. Ellerbeck, A. E. Stones, R. P. A. Dullens, and D. G. A. L. Aarts, "Inverting $g(r)$ to $u(r)$: The test-particle insertion method," JCIS Open **20**, 100156 (2025).
[29]C. R. Rees-Zimmerman, J. Martín-Roca, D. Evans, M. A. Miller, D. G. A. L. Aarts, and C. Valeriani, "Numerical methods for unraveling inter-particle potentials in colloidal suspensions: A comparative study for two-dimensional suspensions," J. Chem. Phys. **162**, 074103 (2025).
[30]A. E. Stones and D. G. A. L. Aarts, "Measuring many-body distribution functions in fluids using test-particle insertion," J. Chem. Phys. **159**, 194502 (2023).
[31]D. Borgis, R. Assaraf, B. Rotenberg, and R. Vuilleumier, "Computation of pair distribution functions and three-dimensional densities with a reduced variance principle," Mol. Phys. **111**, 3486 (2013).
[32]B. Rotenberg, "Use the force! Reduced variance estimators for densities, radial distribution functions, and local mobilities in molecular simulations," J. Chem. Phys. **153**, 150902 (2020).
[33]A. Purohit, A. J. Schultz, and D. A. Kofke, "Force-sampling methods for density distributions as instances of mapped averaging," Mol. Phys. **117**, 2822 (2019).
[34]S. W. Coles, E. Mangaud, D. Frenkel, and B. Rotenberg, "Reduced variance analysis of molecular dynamics simulations by linear combination of estimators," J. Chem. Phys. **154**, 191101 (2021).
[35]P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt *et al.*, "SciPy 1.0: Fundamental algorithms for scientific computing in Python," Nat. Methods **17**, 261 (2020).
[36]J. E. Lennard-Jones, "Cohesion," Proc. Phys. Soc. **43**, 461 (1931).
[37]J. D. Weeks, D. Chandler, and H. C. Andersen, "Role of repulsive forces in determining the equilibrium structure of simple liquids," J. Chem. Phys. **54**, 5237 (1971).
[38]K. Zahn, R. Lenke, and G. Maret, "Two-stage melting of paramagnetic colloidal crystals in two dimensions," Phys. Rev. Lett. **82**, 2721 (1999).
[39]N. V. Gribova, Yu. D. Fomin, D. Frenkel, and V. N. Ryzhov, "Waterlike thermodynamic anomalies in a repulsive-shoulder potential system," Phys. Rev. E **79**, 051202 (2009).
[40]A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. in 't Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen, R. Shan, M. J. Stevens, J. Tranchida, C. Trott, and S. J. Plimpton, "LAMMPS—A flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales," Comput. Phys. Commun. **271**, 108171 (2022).
[41]A. A. Louis, "Beware of density dependent pair potentials," J. Phys.: Condens. Matter **14**, 9187 (2002).
[42]Y. Liu, Q. Chen, X. Li, and J. T. Tian, "Fixed-configuration inversion of interaction potentials from equilibrium configurations at fixed state points" (unpublished) (2026).
[43]P. Davide, O. Coquand, and L. Berthier (2026). "paolinodavide/forceIBI: Code for 'irect Boltzmann inversion method from particle configurations at arbitrary state points,'" Zenodo (v0.1.1). https://doi.org/10.5281/zenodo.20056599
[44]A. K. Soper, "Empirical potential Monte Carlo simulation of fluid structure," Chem. Phys. **202**, 295 (1996).