



# A variational framework for the inverse Henderson problem of statistical mechanics

Fabio Frommer<sup>1</sup> · Martin Hanke<sup>1</sup>

Received: 20 April 2022 / Revised: 8 June 2022 / Accepted: 27 June 2022 /

Published online: 17 July 2022

© The Author(s) 2022, corrected publication 2022

## Abstract

The inverse Henderson problem refers to the determination of the pair potential which specifies the interactions in an ensemble of classical particles in continuous space, given the density and the equilibrium pair correlation function of these particles as data. For a canonical ensemble in a bounded domain, it has been observed that this pair potential minimizes a corresponding convex relative entropy functional, and that the Newton iteration for minimizing this functional coincides with the so-called inverse Monte Carlo (IMC) iterative scheme. In this paper, we show that in the thermodynamic limit analogous connections exist between the specific relative entropy introduced by Georgii and Zessin and a proper formulation of the IMC iteration in the full space. This provides a rigorous variational framework for the inverse Henderson problem, valid within a large class of pair potentials, including, for example, Lennard-Jones-type potentials. It is further shown that the pressure is strictly convex as a function of the pair potential and the chemical potential, and that the specific relative entropy at fixed density is a strictly convex function of the pair potential. At a given reference potential and a corresponding density in the gas phase, we determine the gradient and the Hessian of the specific relative entropy, and we prove that the Hessian extends to a symmetric positive semidefinite quadratic functional in the space of square integrable perturbations of this potential.

---

This article belongs to the themed collection: Mathematical Physics and Numerical Simulation of Many-Particle Systems; V. Bach and L. Delle Site (eds.). The research leading to this work has been done within the Collaborative Research Center TRR 146; corresponding funding by the DFG is gratefully acknowledged.

---

✉ Martin Hanke  
hanke@math.uni-mainz.de

Fabio Frommer  
fabiofrommer@uni-mainz.de

<sup>1</sup> Institut für Mathematik, Johannes Gutenberg-Universität Mainz, 55099 Mainz, Germany

**Keywords** Coarse-graining · Relative entropy · Pair correlation function · Cluster function · Pressure · Fréchet derivative · Gibbs point processes

**Mathematics Subject Classification** 82B21 · 82B80 · 60G55

## 1 Introduction

Numerical simulation has established itself as an independent and indispensable branch of research in the natural sciences, on equal footing with theory and experiment. To be truly useful, numerical approaches have to face and master the multiscale nature which is ubiquitous in almost all relevant applications. This is particularly true for soft matter, where spatial scales may bridge from the electron scale up to the millimeter scale of biomaterials or polymers. Concerning examples we refer to the excellent survey by Noid [31], the collected volume edited by Monticelli and Salonen [28], or the recent special issue [40] of *Journal of Physics: Condensed Matter*.

An important technique to advance numerical algorithms to the particular needs of multiscale applications consists in *coarse-graining*, cf., e.g., Noid [32], or Peter and Kremer [33]: small-scale features are discarded on the coarser scale by replacing detailed descriptions of molecules or matter by artificial *beads* of a certain shape. The simulation then focusses only on these beads and their interaction with the other constituents of the system. When and where necessary, fine details can be reinserted back into the simulation for better accuracy as, e.g., in the AdResS scheme ([6, 34]).

Of course, to evaluate the equations of motion for the coarse-grained model, it is necessary to derive the prevailing effective forces on these beads. Concerning the transition from an atomic microscale to a molecular macroscale in thermodynamic equilibrium, for example, there are essentially two ways to settle this problem. On the one hand, one can employ an *ab initio* bottom-up approach and evaluate and assemble the resulting forces from the eliminated details of the fine-grained description, cf. Ercolessi and Adams [5], or run a fine-grained simulation, compute the corresponding forces, and somehow approximate them on the coarse level as suggested, e.g., by Izvekov and Voth [20], and Wang et al [46]. The other alternative is to follow a top-down strategy and use the given structural information about the location of the beads on the coarse scale to formulate an *inverse problem*: Which are the appropriate forces or interactions on the coarse level that define ensembles with the same structural properties?

In this work, we treat one of the simplest incarnations of the latter approach. Let us presume that the probabilities of the snapshots of the coarse-grained bead ensembles are in good agreement with a model which only uses additive pairwise translation invariant interactions of the beads. Such interactions can be formulated in terms of a scalar pair potential which only depends on the relative position of the respective pair of beads. The pair correlation function, which measures the empirical likelihood to observe two beads at a given relative position, appears to be an adequate piece of data to be used for finding the corresponding potential, because both the data and the unknown consist of a scalar function of the space variable in that case. In fact, in a celebrated paper, Henderson [18] argued that the pair potential is uniquely specified

this way, i.e., under given conditions of temperature and density, no two different pair potentials can give rise to the same piece of empirical data. Finding a pair potential that reproduces the given pair correlation function is therefore sometimes called the *inverse Henderson problem*.

The numerical solution of the inverse Henderson problem is demanding because of the lack of a mathematical formula for computing the pair correlation function for a given pair potential, or vice versa. In the old days people have developed approximate identities like the Percus–Yevick or hypernetted chain integral equations for this purpose, cf. Hansen and McDonald [17], or have used parametric ansatz functions like, e.g., Lennard–Jones potentials, and optimized the corresponding parameters numerically. Today, the state of the art is to use nonparametric potentials and employ iterative schemes which, in each iteration, simulate the equilibrium ensemble for the current guess of the pair potential, and use the associated data fit to somehow generate a new guess. Well-known examples are the iterative Boltzmann inversion (IBI), cf. Soper [42] and Reith, Pütz, and Müller-Plathe [35], and the inverse Monte Carlo method (IMC) by Lyubartsev and Laaksonen [26]. We recommend the valuable reviews by Toth [44] and Rühle et al [37] for a comparison of these and further methods.

We emphasize that the setting of the inverse Henderson problem is generally accepted to be far too simplistic to capture all the relevant features of a real system, cf., e.g., [25, 32, 45], mostly because many-body interactions are neglected. In particular, thermodynamic properties of the coarse-grained model may differ from the real system, especially at other temperatures or densities.

But its simplicity offers a great opportunity for a mathematical analysis, which in turn may lead to a better understanding of other, more flexible coarse-graining techniques that are routinely being employed in practice. Still, only few rigorous mathematical results have yet been obtained, e.g., in [2, 7, 16, 23, 24, 30], the reason being, again, the lack of explicit formulae to attack the problem.

The aim of this paper is to point out and advocate an alternative access point for theoretical investigations, which goes back to a nice observation by Shell [41] from within the chemical physics community: He argues that the Henderson potential minimizes the (information theoretic) relative entropy

$$S_{\text{rel}} = \sum_{\gamma} \mathcal{P}_*(\gamma) \log \frac{\mathcal{P}_*(\gamma)}{\mathcal{P}(\gamma)}, \quad (1.1)$$

where—in his words—the summation is over all possible (coarse grained) ensemble configurations  $\gamma$ ,  $\mathcal{P}_*(\gamma)$  denotes the target (or observed) probability of  $\gamma$ , and  $\mathcal{P}(\gamma)$  is the corresponding probability of a model with a given pair potential. (Compare, for example, Georgii [12] for background on the concept of relative entropy in stochastics.) Subsequently, Murtola, Karttunen, and Vattulainen [29] pointed out that the functional (1.1) is convex, and that the Newton iteration for minimizing the relative entropy coincides with the aforementioned IMC iteration (see also Rosenberger et al [36]).

Like Henderson’s paper [18], the results in [29, 36, 41] lack mathematical rigor, because their arguments are restricted to bounded domains, whereas an unambigu-

ous definition of a “translation invariant ensemble” is only possible in the full space. Concerning the Henderson theorem, this shortcoming has recently been fixed in [7], building on fundamental work by Ruelle and by Georgii. Here, we focus on the proposal by Shell and his colleagues, provide a rigorous justification of their results, and elaborate further on them.

To be specific, we outline in Sect. 2 that the (appropriately formulated) relative entropy (1.1) divided by the volume of the bounded domain converges in the thermodynamic limit to the specific relative entropy, first introduced for continuum systems by Gallavotti and Miracle-Sole [8] in the case of hard-core interactions, and further investigated by Georgii and Zessin in [9, 10, 13] for general additive pair interactions. Under mild assumptions on the model and target ensembles to be utilized, we prove that this specific relative entropy is a strictly convex functional on the corresponding set of pair potentials (which include Lennard-Jones type and hard-core potentials), and that its (unique) minimizer is the particular potential which solves the inverse Henderson problem (see Sects. 3 and 4).

From Sect. 5 onward, we restrict ourselves to low densities (the “gas phase”), where the specific relative entropy is a differentiable function of the pair potential. We calculate its Hessian, and in Sect. 6, we verify that the Newton iteration for minimizing the specific relative entropy does indeed coincide with the IMC iteration formulated in the thermodynamic limit. In Sect. 7, we investigate the Hessian in more detail and show that it can be represented by a self-adjoint positive semidefinite operator in  $L^2$ . The mapping properties of this operator can be analyzed somewhat further for the particular class of Lennard-Jones-type pair potentials; we conclude with a corresponding result in Sect. 8.

We hope that this variational framework for the inverse Henderson problem opens a possibility to discuss the convergence of the IMC iteration, or to come up with measures to stabilize or regularize this popular iterative scheme.

## 2 The relative entropy in the thermodynamic limit

To reformulate Shell’s approach within a rigorous mathematical framework we start with the assumption that the target ensemble is given by a translation invariant probability measure  $\mathbb{P}_*$  on the configuration space

$$\Gamma = \{ \gamma \subset \mathbb{R}^d : \Delta \subset \mathbb{R}^d \text{ bounded} \Rightarrow \#(\gamma \cap \Delta) < \infty \}$$

with density  $\rho_*$  and finite locally second moments, compare Georgii [9]. For the model ensemble, we restrict ourselves to ensembles of classical particles with additive pairwise interactions defined by a measurable even pair potential  $u : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ . Concerning the latter, we assume that there exists  $r_0 > 0$  and decreasing positive functions  $\varphi : (0, r_0) \rightarrow \mathbb{R}_0^+$  and  $\psi : [0, \infty) \rightarrow \mathbb{R}^+$  with

$$\int_0^{r_0} r^{d-1} \varphi(r) dr = +\infty \quad \text{and} \quad \int_0^\infty r^{d-1} \psi(r) dr < \infty, \quad (2.1)$$

such that

$$\begin{aligned} u(x) &\geq \varphi(|x|) && \text{for } 0 < |x| < r_0, \\ |u(x)| &\leq \psi(|x|) && \text{for } |x| \geq r_0, \end{aligned} \tag{2.2}$$

holds true almost everywhere; for our convenience, we take  $\psi$  to be a bounded and decreasing function defined for all  $r \geq 0$ . We denote by  $\mathcal{U}_0$  the subset of the above pair potentials, which also belong to  $L^\infty_{\text{loc}}(\mathbb{R}^3 \setminus \{0\})$ , and define  $\mathcal{U}$  to be the union of  $\mathcal{U}_0$  with the hard-core potentials, which satisfy (2.2) with  $\psi$  as above and  $\varphi$  replaced by  $+\infty$ ; accordingly,  $r_0$  is taken to be the hard-core radius in this case. Technically, we do not distinguish between potentials (and functions in general) which differ on sets of Lebesgue measure zero.

**Remark 2.1** The set  $\mathcal{U}$  is convex. To see this let  $u_i \in \mathcal{U}, i = 1, 2$ , be such that (2.1) and (2.2) hold for  $r_{0,i} > 0$  and decreasing functions  $\varphi_i$  and  $\psi_i$  satisfying (2.1), respectively. Without loss of generality, we may assume that  $r_{0,1} \leq r_{0,2}$ . Let  $u = tu_1 + (1 - t)u_2$  for some fixed  $t \in (0, 1)$ . We distinguish two cases. If  $u_2$  is a hard-core potential, then  $u$  is also a hard-core potential with hard-core radius  $r_{0,2}$ . In this case, we can choose  $r_0 = r_{0,2}$ ,

$$\begin{aligned} \varphi(r) &= +\infty && \text{for } 0 < r < r_{0,2}, \\ \psi(r) &= t\psi_1(r) + (1 - t)\psi_2(r) && \text{for } r \geq 0, \end{aligned}$$

to achieve the assumption (2.2) for  $u$ . On the other hand, if  $u_2$  is no hard-core potential, then  $u_2$  is bounded on  $[r_{0,1}, r_{0,2})$ , and there exists  $c \geq 1$ , such that

$$|u_2(x)| \leq c \psi_2(|x|) \quad \text{for } r_{0,1} \leq |x| < r_{0,2}.$$

It follows that  $u$  satisfies the inequalities (2.2) with  $r_0 = r_{0,1}$  and

$$\begin{aligned} \varphi(r) &= t\varphi_1(r) + (1 - t)\varphi_2(r) && \text{for } 0 < r < r_{0,1}, \\ \psi(r) &= t\psi_1(r) + (1 - t)c \psi_2(r) && \text{for } r \geq 0. \end{aligned}$$

In either case, we have verified that  $u \in \mathcal{U}$ , hence  $\mathcal{U}$  is convex. ◊

Let  $\Lambda = [-\ell, \ell]^d$  be a bounded box in  $\mathbb{R}^d$ . In analogy to Shell, who considered the relative entropy framework for canonical ensembles in  $\Lambda$ , we take the restriction  $\mathbb{P}_{*,\Lambda}$  of  $\mathbb{P}_*$  as target, and the grand canonical ensemble in  $\Lambda$  with chemical potential  $\mu \in \mathbb{R}$  and pair potential  $u \in \mathcal{U}$  as model. For a configuration  $\gamma \in \Gamma$  of  $N = N(\gamma)$  particles located at  $\{x_1, \dots, x_N\} \subset \Lambda$ , the probability density of the model is thus given by

$$\mathcal{P}(\gamma) = \frac{1}{\Xi_\Lambda} e^{\beta\mu N(\gamma)} e^{-\beta U(\gamma)},$$

where  $\beta > 0$  is the inverse temperature,

$$U(\gamma) = U(x_1, \dots, x_N) = \sum_{1 \leq i < j \leq N} u(x_i - x_j) \tag{2.3}$$

is the interaction energy, and

$$\Xi_\Lambda = \sum_{N=0}^\infty \frac{e^{\beta\mu N}}{N!} \int_{\Lambda^N} e^{-\beta U(x_1, \dots, x_N)} dx_1 \cdots dx_N$$

is the corresponding grand canonical partition function.

Note that most of the quantities we deal with depend on  $\beta$ ; however, since we will keep the temperature fixed throughout this paper, we refrain from making this dependency explicit in our notation.

Let  $S(\mathbb{P}_{*,\Lambda})$  denote the entropy associated with  $\mathbb{P}_{*,\Lambda}$ . In the spirit of (1.1), we then specify the relative entropy

$$\begin{aligned} S_{\text{rel},\Lambda} &= S(\mathbb{P}_{*,\Lambda}) + \int_{\Gamma_\Lambda} (\log \Xi_\Lambda - \beta\mu N(\gamma) + \beta U(\gamma)) d\mathbb{P}_{*,\Lambda}(\gamma) \\ &= S(\mathbb{P}_{*,\Lambda}) + \log \Xi_\Lambda - \beta\mu \mathbb{E}_{*,\Lambda}[N] + \beta \mathbb{E}_{*,\Lambda}[U], \end{aligned}$$

where  $\mathbb{E}_{*,\Lambda}[\cdot]$  denotes expectation with respect to  $\mathbb{P}_{*,\Lambda}$ . Take note that the entropy and the expected interaction energy may be  $+\infty$ . Dividing by  $\beta|\Lambda|$ , we arrive at

$$\frac{1}{\beta|\Lambda|} S_{\text{rel},\Lambda} = \frac{1}{\beta|\Lambda|} S(\mathbb{P}_{*,\Lambda}) + \frac{1}{\beta|\Lambda|} \log \Xi_\Lambda - \mu\rho_* + \frac{1}{|\Lambda|} \mathbb{E}_{*,\Lambda}[U]. \tag{2.4}$$

Now, we want to derive the analog of this identity for the thermodynamic limit  $\ell \rightarrow \infty$ . Concerning this limit it is known that there is a sequence  $(\ell_k)_k$  with  $\ell_k \rightarrow \infty$ , such that the probability densities associated with the corresponding grand canonical ensembles converge in a local topology to a translation invariant tempered Gibbs measure on  $\Gamma$ , cf. Ruelle [39], for brevity called  $(\mu, u)$ -Gibbs measure in the sequel. In general, different sequences may lead to different  $(\mu, u)$ -Gibbs measures, e.g., when different phases coexist. For any such sequence, however, the limit

$$p(\mu, u) = \lim_{\ell \rightarrow \infty} \frac{1}{\beta|\Lambda|} \log \Xi_\Lambda \tag{2.5}$$

is always the same well-defined and nonnegative finite number, namely the pressure of the ensemble in the thermodynamic limit, cf. Ruelle [38]. Therefore, passing in (2.4) to the thermodynamic limit  $\ell \rightarrow \infty$ , we can (uniquely) define the relative entropy  $S_{\text{rel}}(\mu, u)$  of all these  $(\mu, u)$ -Gibbs measures with respect to the target model  $\mathbb{P}_*$  as

$$\frac{1}{\beta} S_{\text{rel}}(\mu, u) = \frac{1}{\beta} S_* + p(\mu, u) - \mu\rho_* + E(u, \mathbb{P}_*), \tag{2.6}$$

where the individual terms on the right-hand side of (2.6) are the corresponding limits of the respective terms in (2.4):  $S_* = S(\mathbb{P}_*)$  is the specific entropy and  $E(u, \mathbb{P}_*)$  is the specific interaction energy (with respect to the potential  $u$ ) of the target ensemble, both of which have been shown to be well-defined in  $\mathbb{R} \cup \{+\infty\}$ , cf. [11, 13]. If  $\mathbb{P}_*$  satisfies a Ruelle condition (compare (A.1) in the appendix), then the specific entropy

is finite; this is the case, e.g., when  $\mathbb{P}_*$  is a  $(\mu_*, u_*)$ -Gibbs measure for some  $u_* \in \mathcal{U}$  and  $\mu_* \in \mathbb{R}$ , compare [39, Corollary 5.3]. For this latter particular case, it has further been shown in [7] that

$$E(u, \mathbb{P}_*) = \frac{1}{2} \int_{\mathbb{R}^d} u(x) \rho_*^{(2)}(x) \, dx, \tag{2.7}$$

where  $\rho_*^{(2)}(x)$  is the pair correlation function associated with  $\mathbb{P}_*$ . Here, and throughout this paper, we deliberately use the short-hand notation  $\rho^{(2)}(x)$  instead of  $\rho^{(2)}(x, 0)$  for the pair correlation function of a translation invariant point process.

In [10], Georgii studied the relative entropy  $S_{\text{rel}}(\mu, u)$  in detail and established the following *Gibbs variational principle*:

**Theorem A** (Gibbs variational principle) *Let  $\mathbb{P}_*$  be a translation invariant probability measure on  $\Gamma$  with finite locally second moments. Then, the relative entropy (2.6) is nonnegative real or  $+\infty$  for every  $u \in \mathcal{U}$  and  $\mu \in \mathbb{R}$ . There holds  $S_{\text{rel}}(\mu, u) = 0$ , if and only if  $\mathbb{P}_*$  is a  $(\mu, u)$ -Gibbs measure.*

We have utilized this result in [7] to prove a rigorous version of the Henderson theorem:

**Theorem B** (Henderson theorem) *Let  $u_1, u_2 \in \mathcal{U}$  and  $\mu_1, \mu_2 \in \mathbb{R}$ . If  $\mathbb{P}_1$  and  $\mathbb{P}_2$  are  $(\mu_1, u_1)$ - and  $(\mu_2, u_2)$ -Gibbs measures, respectively, which share the same density and the same pair correlation function, then  $u_1 = u_2$  and  $\mu_1 = \mu_2$ .*

Combining Georgii’s Gibbs variational principle and the Henderson theorem, we can formulate an alternative version of the Gibbs variational principle. This version is the one that we will mostly use below.

**Theorem 2.2** (Gibbs variational principle, alternative form) *If the target  $\mathbb{P}_*$  is a  $(\mu_*, u_*)$ -Gibbs measure for some  $u_* \in \mathcal{U}$  and  $\mu_* \in \mathbb{R}$ , then the relative entropy  $S_{\text{rel}}(\mu, u)$  becomes minimal, if and only if  $u = u_*$  and  $\mu = \mu_*$ .*

**Proof** According to Theorem A, it remains to investigate the case when  $S_{\text{rel}}(\mu, u) = 0$  for some  $u \in \mathcal{U}$  and  $\mu \in \mathbb{R}$ . The Gibbs variational principle states that  $\mathbb{P}_*$  is then a  $(\mu_*, u_*)$ - and  $(\mu, u)$ -Gibbs measure at the same time. In particular this means that these two Gibbs measures share the same density and pair correlation function. The assertion thus follows from the Henderson theorem. □

Georgii investigated the relative entropy for fixed interaction and chemical potentials  $u$  and  $\mu$ , and varied the target model  $\mathbb{P}_*$ . In connection with the inverse Henderson problem, our interest is sort of dual to this: we assume that  $\mathbb{P}_*$  is fixed, and consider the relative entropy as a function of  $\mu$  and  $u$ .

### 3 Strict convexity of the pressure and the relative entropy

It is well-known that the pressure is a convex function of the chemical potential, cf. [38, Theorem 3.4.6]. This convexity is strict, whenever a Gibbs variational principle is valid,

cf., e.g., Hughes [19, Sect. 4.3]. In the sequel, we show that under our assumptions the pressure is also strictly convex in  $\mu$  and  $u$ .

**Theorem 3.1** *The pressure  $p = p(\mu, u)$  of (2.5) is a strictly convex function of  $\mu \in \mathbb{R}$  and  $u \in \mathcal{U}$ . Moreover, for  $u \in \mathcal{U}_0$ , there holds*

$$\frac{p(\mu, u)}{\mu} \rightarrow +\infty \quad \text{as } \mu \rightarrow +\infty. \quad (3.1)$$

**Proof** To begin with, we first observe that  $\mathbb{R} \times \mathcal{U}$  is convex by virtue of Remark 2.1.

Now let  $\mu_1, \mu_2 \in \mathbb{R}$  and  $u_1, u_2 \in \mathcal{U}$  be arbitrarily chosen, not both being equal at the same time, and define

$$\mu = t\mu_1 + (1-t)\mu_2 \quad \text{and} \quad u = tu_1 + (1-t)u_2$$

for some fixed  $t \in (0, 1)$ . When choosing for  $\mathbb{P}_*$  a corresponding  $(\mu, u)$ -Gibbs measure, then it follows from (2.6) and the Gibbs variational principle of Theorem 2.2 that

$$p(\mu, u) = -\frac{1}{\beta} S_* + \mu\rho_* - E(u, \mathbb{P}_*). \quad (3.2)$$

In particular, the specific entropy  $S_*$  is finite because  $\mathbb{P}_*$  is a Gibbs measure, and hence, so is the specific interaction energy. Likewise, we obtain

$$p(\mu_1, u_1) > -\frac{1}{\beta} S_* + \mu_1\rho_* - E(u_1, \mathbb{P}_*)$$

and

$$p(\mu_2, u_2) > -\frac{1}{\beta} S_* + \mu_2\rho_* - E(u_2, \mathbb{P}_*),$$

which gives

$$t p(\mu_1, u_1) + (1-t) p(\mu_2, u_2) > -\frac{1}{\beta} S_* + \mu\rho_* - E(u, \mathbb{P}_*)$$

because of the linearity of the specific interaction energy. A comparison with (3.2) thus shows that the pressure is strictly convex.

Consider now a fixed pair potential  $u \in \mathcal{U}_0$ . It has been shown in the proof of [9, Lemma 7.1] that for any  $\rho > 0$  there exists a translation invariant probability measure  $\mathbb{P}_\rho$  on  $\Gamma$  with density  $\rho$ , such that  $S(\mathbb{P}_\rho) < \infty$ , and  $E(u, \mathbb{P}_\rho) < \infty$ . Choosing  $\mathbb{P}_* = \mathbb{P}_\rho$  in (2.6), Georgii's Gibbs variational principle (Theorem A) yields the inequality

$$p(\mu, u) - \mu\rho \geq -\frac{1}{\beta} S(\mathbb{P}_\rho) - E(u, \mathbb{P}_\rho) =: c_\rho > -\infty \quad (3.3)$$



for every  $\mu \in \mathbb{R}$ . In other words,

$$\frac{p(\mu, u)}{\mu} \geq \rho + \frac{c_\rho}{\mu},$$

and hence,

$$\liminf_{\mu \rightarrow +\infty} \frac{p(\mu, u)}{\mu} \geq \rho.$$

Since  $\rho > 0$  has been arbitrary, this implies (3.1). □

We thus have shown that the relative entropy (2.6) is the sum of a strictly convex functional and an affine function of  $(\mu, u)$ . Accordingly, the relative entropy is also a strictly convex functional of  $\mu$  and  $u$ , as long as it is finite.

Another immediate consequence of Theorem 2.2 is the following inequality for the pressure.

**Corollary 3.2** *For  $\mu_1, \mu_2 \in \mathbb{R}$  and  $u_1, u_2 \in \mathcal{U}$ , there holds*

$$\begin{aligned} (\mu_2 - \mu_1)\rho_1 - \frac{1}{2} \int_{\mathbb{R}^d} (u_2 - u_1)(x)\rho_1^{(2)}(x) \, dx \\ \leq p(\mu_2, u_2) - p(\mu_1, u_1) \leq (\mu_2 - \mu_1)\rho_2 - \frac{1}{2} \int_{\mathbb{R}^d} (u_2 - u_1)(x)\rho_2^{(2)}(x) \, dx. \end{aligned} \tag{3.4}$$

whenever  $\rho_i$  and  $\rho_i^{(2)}$  are the density and pair correlation function of a  $(\mu_i, u_i)$ -Gibbs measure, respectively. Both inequalities are strict, unless  $\mu_1 = \mu_2$  and  $u_1 = u_2$ . In particular, if  $u_1 = u_2$  and  $\mu_1 < \mu_2$ , then  $\rho_1 < \rho_2$ .

**Proof** The two inequalities (3.4) follow readily from Theorem 2.2 by choosing for  $\mathbb{P}_*$  the corresponding  $(\mu_i, u_i)$ -Gibbs measures, respectively. They are strict, unless  $\mu_1 = \mu_2$  and  $u_1 = u_2$ . □

### 4 The relative entropy functional for fixed density

Returning to the inverse Henderson problem formulated in the introduction, we now constrain our model ensembles to have the same density  $\rho_*$  as the target ensemble. We do so because numerical simulations are often done in the canonical ensemble, where the density is the crucial quantity, and because this is also the setting used by Henderson and by Shell, as well as in the definition of the IMC scheme mentioned before. Since the attainable densities for hard-core potentials are bounded, we need to distinguish the case whether  $u$  is a hard-core potential or not. We focus our analysis on the latter case and mention the necessary modifications for hard-core potentials in Remark 4.4 later in this section.

The first fundamental problem to settle concerns the question whether and how the prescribed density  $\rho_*$  can be attained by some  $(\mu, u)$ -Gibbs measure for a given

$u \in \mathcal{U}_0$ . When the chemical potential is sufficiently small, i.e., when the system is in the *gas phase* (see Sect. 5 for a specification of this term), then it is known that there is a one-to-one relation between the corresponding chemical potentials and the associated densities, and in this case, the inverse map  $\rho \mapsto \mu$  can even be computed by means of cluster expansions, cf., e.g., Jansen, Kuna, and Tsagkarogiannis [22]. Outside the gas phase, where phase transitions may occur, the problem becomes more difficult. Adopting a method from Chayes and Chayes [2], we can establish the following result, which holds for the full range of possible densities.

**Theorem 4.1** *Let  $u \in \mathcal{U}_0$  and  $\rho_* > 0$  be fixed. Then, there is a unique chemical potential  $\mu_* = \mu_*(u) \in \mathbb{R}$ , for which there exists a  $(\mu_*, u)$ -Gibbs measure with density  $\rho_*$ .*

**Proof** For every  $\mu \in \mathbb{R}$  and the given  $u \in \mathcal{U}_0$ , let  $\mathbb{P}_{\mu, u}$  be a  $(\mu, u)$ -Gibbs measure, and denote by  $p(\mu)$  and  $\rho(\mu)$  the associated pressure and density, respectively.

It is well-known, cf. [38, Theorem 4.3.1], that for small chemical potentials, the pressure is a differentiable function with

$$p'(\mu) = \rho(\mu). \quad (4.1)$$

In other words, if  $\rho_*$  is sufficiently small, then the corresponding chemical potential  $\mu_*$  from the formulation of the theorem is given as the unique minimum of the function

$$\Theta(\mu) = p(\mu) - \mu\rho_*, \quad (4.2)$$

the latter being strictly convex by virtue of Theorem 3.1. We will proceed by showing that the minimizer of  $\Theta$  is also the appropriate chemical potential to choose for larger values of  $\rho_*$ .

To see this, we first observe that

$$\Theta(\mu) \geq -\mu\rho_*,$$

because the pressure is nonnegative, whereas

$$\Theta(\mu) \geq c_{(1+\varepsilon)\rho_*} + \varepsilon\mu\rho_* \quad (4.3)$$

for any suitable  $\varepsilon > 0$  and corresponding constant  $c_{(1+\varepsilon)\rho_*}$  by virtue of (3.3) (with  $\rho = (1 + \varepsilon)\rho_*$ ). This shows that  $\Theta$  is bounded from below and that

$$\Theta(\mu) \rightarrow +\infty, \quad \text{whenever } |\mu| \rightarrow \infty.$$

Therefore,  $\Theta$  attains its minimum for a uniquely defined value  $\mu = \mu_*$ .

For any  $\mu \in \mathbb{R}$ ,  $\mu \neq \mu_*$ , we now conclude from (3.4) and (4.2) that

$$(\mu_* - \mu)\rho(\mu) < p(\mu_*) - p(\mu) = \Theta(\mu_*) + \mu_*\rho_* - \Theta(\mu) - \mu\rho_*,$$

i.e.,

$$(\mu_* - \mu)(\rho(\mu) - \rho_*) < \Theta(\mu_*) - \Theta(\mu) < 0.$$

This means that

$$\begin{aligned} \rho(\mu) < \rho_* & \text{ for } \mu < \mu_* , \\ \rho(\mu) > \rho_* & \text{ for } \mu > \mu_* . \end{aligned} \tag{4.4}$$

Now, let  $(\mu_k^+)_k$  be a strictly decreasing sequence and  $(\mu_k^-)_k$  a strictly increasing sequence of chemical potentials, both of which converge to  $\mu_*$ . By virtue of Lemma A.1 from the appendix, there exist  $(\mu_*, u)$ -Gibbs measures  $\mathbb{P}^-$  and  $\mathbb{P}^+$  with densities

$$\rho(\mathbb{P}^-) = \lim_{k \rightarrow \infty} \rho(\mu_k^-) \quad \text{and} \quad \rho(\mathbb{P}^+) = \lim_{k \rightarrow \infty} \rho(\mu_k^+).$$

The inequalities (4.4) imply that

$$\rho(\mathbb{P}^-) \leq \rho_* \leq \rho(\mathbb{P}^+),$$

and hence, there is some  $t \in [0, 1]$ , for which  $\mathbb{P} = t\mathbb{P}^- + (1 - t)\mathbb{P}^+$  has density

$$\rho(\mathbb{P}) = t\rho(\mathbb{P}^-) + (1 - t)\rho(\mathbb{P}^+) = \rho_* .$$

Since the set of  $(\mu_*, u)$ -Gibbs measures is convex,  $\mathbb{P}$  has all the desired properties from the statement of this theorem, and the proof is done. □

In the light of the above theorem, we can now restrict our attention to  $(\mu, u)$ -Gibbs measures with density  $\rho_*$ , i.e., with chemical potential  $\mu = \mu_*(u)$ , when looking at the relative entropy functional. Further, we drop the constant offset  $S_*$  in (2.6), as it is independent of  $u$ . This leads to the functional

$$\Phi(u) = p_*(u) - \mu_*(u)\rho_* + E(u, \mathbb{P}_*) \tag{4.5}$$

for  $u \in \mathcal{U}_0$ , where we have set

$$p_*(u) = p(\mu_*(u), u) \tag{4.6}$$

for brevity. By a slight abuse of wording we will keep calling  $\Phi$  the relative entropy functional. As we will see next, although  $p_*$  may fail to be convex, in general, the functional  $\Phi$  is strictly convex, again.

**Theorem 4.2** *Let  $\mathbb{P}_*$  be as in Theorem A. Then, the functional  $\Phi : \mathcal{U}_0 \rightarrow \mathbb{R}$  of (4.5) is strictly convex as long as the specific interaction energy  $E(u, \mathbb{P}_*)$  is finite.*

**Proof** Let  $u_1$  and  $u_2$  be two different pair potentials from  $\mathcal{U}_0$ . For any fixed  $0 < t < 1$ , define

$$u = tu_1 + (1 - t)u_2 ,$$

and, as in Theorem 4.1, denote by  $\mu = \mu_*(u)$  the chemical potential associated with  $u$  and density  $\rho_*$ . Finally, let  $\rho^{(2)}$  be the pair correlation function of the associated  $(\mu, u)$ -Gibbs measure constructed in Theorem 4.1.

Then, we obtain from (4.6) and (3.4)—with  $\mu_i = \mu_*(u_i)$  for  $i = 1, 2$ —that

$$\begin{aligned} & p_*(u) - t p_*(u_1) - (1 - t) p_*(u_2) \\ &= t(p_*(u) - p_*(u_1)) + (1 - t)(p_*(u) - p_*(u_2)) \\ &< t \left( (\mu - \mu_1) \rho_* - \frac{1 - t}{2} \int_{\mathbb{R}^d} (u_2 - u_1)(x) \rho^{(2)}(x) dx \right) \\ &\quad + (1 - t) \left( (\mu - \mu_2) \rho_* - \frac{t}{2} \int_{\mathbb{R}^d} (u_1 - u_2)(x) \rho^{(2)}(x) dx \right) \\ &= \mu \rho_* - t \mu_1 \rho_* - (1 - t) \mu_2 \rho_* . \end{aligned}$$

Note that this inequality is strict because  $u$  is different from  $u_1$  and  $u_2$  by construction. Reordering terms, we thus arrive at

$$p_*(u) - \mu \rho_* < t(p_*(u_1) - \mu_1 \rho_*) + (1 - t)(p_*(u_2) - \mu_2 \rho_*). \tag{4.7}$$

It thus follows from (4.5) and (4.7) that if the specific interaction energy  $E(\cdot, \mathbb{P}_*)$  stays finite, then  $\Phi$  is strictly convex, because  $E(\cdot, \mathbb{P}_*)$  is linear in the first argument. □

Theorem 4.2 implies that the relative entropy functional has at most one local minimizer, which is then also a global one. Concerning this minimizer, we have the following result.

**Theorem 4.3** *Let  $\mathbb{P}_*$  be a  $(\mu, u_*)$ -Gibbs measure for some  $u_* \in \mathcal{U}_0$  and  $\mu \in \mathbb{R}$ , and let  $\rho_*$  be its density. Then, the relative entropy function (4.5) attains its minimum for  $u = u_*$ .*

**Proof** The given Gibbs measure  $\mathbb{P}_*$  has finite specific entropy  $S_*$ , and hence, (2.6) implies that

$$\Phi(u) = \frac{1}{\beta} S_{\text{rel}}(\mu_*(u), u) - \frac{1}{\beta} S_* \tag{4.8}$$

for every  $u \in \mathcal{U}_0$ . Since  $\mathbb{P}_*$  has density  $\rho_*$  the chemical potential  $\mu$  associated with  $\mathbb{P}_*$  must be given by  $\mu = \mu_*(u_*)$  according to Theorem 4.1. From (4.8) and the Gibbs variational principle (Theorems A and 2.2), it follows that

$$\Phi(u_*) = -\frac{1}{\beta} S_* < \frac{1}{\beta} S_{\text{rel}}(\mu_*(u), u) - \frac{1}{\beta} S_* = \Phi(u) \tag{4.9}$$

for every  $u \in \mathcal{U}_0 \setminus \{u_*\}$ , and this was to be shown. □

Note from (4.9) that the minimal value of  $\Phi$  depends on the specific entropy of the target Gibbs measure, and is therefore unknown in general.

**Remark 4.4** For a hard-core potential  $u \in \mathcal{U}$  the density  $\rho$  of any associated  $(\mu, u)$ -Gibbs measure is bounded from above by a finite *closest-packing density*  $\rho_{cp} = \rho_{cp}(u)$ : This bound only depends on the hard-core radius  $r_0$  of (2.2) and is a decreasing function of  $r_0$ , cf. [9, Sect. 7]. We formally set  $\rho_{cp}(u) = +\infty$  for  $u \in \mathcal{U}_0$ .

If the given density  $\rho_*$  of the target happens to be below  $\rho_{cp}(u)$  for a given  $u \in \mathcal{U}$ , then Theorem 4.1 is still valid for this pair potential; to establish (4.3) in its proof,  $\varepsilon > 0$  must be so small that  $(1 + \varepsilon)\rho_*$  is still below  $\rho_{cp}(u)$ , because this is needed for [9, Lemma 7.1], and hence, for (3.3). The relative entropy functional (4.5), however, is only well-defined on the domain

$$\mathcal{D}(\Phi) = \{u \in \mathcal{U} : \rho_{cp}(u) > \rho_*\},$$

but this is again a convex set, compare Remark 2.1, and  $\Phi$  happens to be strictly convex on  $\mathcal{D}(\Phi)$ . Theorem 4.3 extends literally to every  $u_* \in \mathcal{U}$  with this understanding of the domain of  $\Phi$ . ◇

Theorem 4.3 is the basis for a variational setting of the inverse Henderson problem: If the density and the pair correlation function of a  $(\mu_*, u_*)$ -Gibbs measure target  $\mathbb{P}_*$  are given, then the unique minimizer of the strictly convex relative entropy functional (4.5) with  $E(u, \mathbb{P}_*)$  of (2.7) yields the corresponding pair potential  $u_*$ , i.e., the solution of the inverse Henderson problem. Nevertheless, this approach also has some pitfalls as discussed in the following remark.

**Remark 4.5** If the target  $\mathbb{P}_*$  fails to be some  $(\mu, u)$ -Gibbs measure, then it is not clear to us whether the relative entropy functional  $\Phi$  will still be bounded from below on  $\mathcal{U}$ , and even if it may, its infimum need not be attained on  $\mathcal{U}$ .

Vice versa, if  $u_* \in \mathcal{U}$  happens to be the minimizer of  $\Phi$ , then this does not imply that  $\mathbb{P}_*$  is a  $(\mu_*(u_*), u_*)$ -Gibbs measure. To see the latter, consider the following example: Let  $u_*$  be any hard-core potential in  $\mathcal{U}$  and  $\rho_* < \rho_{cp}(u_*)$ ; compare Remark 4.4. Then, provided  $\rho_*$  is sufficiently small, a result by Kuna, Lebowitz, and Speer [24, Corollary 4.3] states that there exist uncountably many distinct translation invariant probability measures with finite specific entropy and density  $\rho_*$ , which share the pair correlation function  $\rho_*^{(2)}$  with the  $(\mu_*(u_*), u_*)$ -Gibbs measure  $\mathbb{P}$ , but  $\mathbb{P}$  is the only  $(\mu, u)$ -Gibbs measure among them according to the Henderson theorem. Because the specific interaction energy is given by (2.7) for all of these probability measures, the relative entropy functional does not differ, and hence,  $u_*$  is its unique minimizer, regardless which of them has been the target  $\mathbb{P}_*$ .

Finally, we mention that even if the target  $\mathbb{P}_*$  is a  $(\mu, u_*)$ -Gibbs measure, and hence, the functional (4.5) is minimized by  $u_*$  according to Theorem 4.3, then this does not seem to imply that the model and the target have the same pair correlation function. To illustrate this, imagine that a fluid corresponding to a pair potential  $u_* \in \mathcal{U}$  exhibits a so-called *triple point* (compare, e.g., [17]), where three different phases coexist at the same thermodynamical state point, i.e., for the same values of pressure (or chemical potential  $\mu_*$ , say) and temperature (or inverse temperature  $\beta$ ). It can be expected that the different phases have linearly independent pair correlation functions; taking convex combinations of the corresponding Gibbs measures, one can thus determine

two  $(\mu_*, u_*)$ -Gibbs measures which have the same density  $\rho_*$ , but have different pair correlation functions. One of them could be the target and the other one the minimizing model. We will see in Sect. 5 that if the density  $\rho_*$  belongs to the gas phase of the minimizer of  $\Phi$ , then the pair correlation function of the minimizing Gibbs measure always coincides with the given  $\rho_*^{(2)}$ ; see Remark 5.5.  $\diamond$

### 5 Differentiability of the relative entropy functional

In the remainder of this paper, we further analyze the case when  $\Phi$  is a differentiable function of  $u$ . For this, we build upon our earlier work [14, 15].<sup>1</sup> A conceptual difficulty in this context is the fact that  $\mathcal{U}$  lacks a universal topology. Therefore, following [14], we consider a tailored neighborhood for any given  $u_0 \in \mathcal{U}$  by introducing a corresponding Banach space  $\mathcal{V}$  of perturbations, consisting of all even measurable functions  $v: \mathbb{R}^d \rightarrow \mathbb{R}$ , for which the associated norm

$$\|v\|_{\mathcal{V}} = \sup_{x \in \mathbb{R}^d} \frac{|v(x)|}{\psi_0(|x|)} \tag{5.1}$$

is finite. Here,  $\psi_0$  is the majorant  $\psi$  of (2.1) associated with  $u_0$ . Note that the norm (5.1) is somewhat stronger than the one employed in [14] because it is more restrictive in the core region  $0 < r < r_0$ , and hence, the resulting space of perturbations is smaller. We believe that this restriction allows a more natural formulation of our results.

**Remark 5.1** It can easily be seen that for any  $v \in \mathcal{V}$  the sum  $u_0 + v$  belongs to  $\mathcal{U}$ : If  $u_0$  is a hard-core potential, then  $r_0$  is its hard-core radius and the corresponding function  $\varphi$  from (2.1) equals  $+\infty$ , and we have

$$\begin{aligned} u_0(x) + v(x) &\geq \varphi(|x|) && \text{for } 0 < |x| < r_0, \\ |u_0(x) + v(x)| &\leq (1 + \|v\|_{\mathcal{V}})\psi_0(|x|) && \text{for } |x| \geq r_0. \end{aligned}$$

Otherwise, there exists  $r_1 \in (0, r_0]$  such that  $\varphi(r) \geq \|v\|_{\mathcal{V}}\psi_0(0)$  for  $0 < r \leq r_1$ , and therefore

$$\begin{aligned} u_0(x) + v(x) &\geq \varphi(|x|) - \|v\|_{\mathcal{V}}\psi_0(0) && \text{for } 0 < |x| < r_1, \\ |u_0(x) + v(x)| &\leq \left( \sup_{|x'| \geq r_1} \frac{|u(x')|}{\psi_0(|x'|)} + \|v\|_{\mathcal{V}} \right) \psi_0(|x|) && \text{for } |x| \geq r_1. \end{aligned}$$

In either case, this shows that  $u_0 + v \in \mathcal{U}$ .  $\diamond$

We now consider the ball

$$\mathcal{B}_{\mathcal{V}}(u_0) = \{u = u_0 + v : \|v\|_{\mathcal{V}} < \delta_0\} \subset \mathcal{U} \tag{5.2}$$

<sup>1</sup> The results in [14, 15] are formulated in three-dimensional space (i.e., for  $d = 3$ ) and for pair potentials, which are rotation invariant; also, hard-core potentials had not been considered. All results extend verbatim to the setting considered here.

around  $u_0$  for a suitable radius  $\delta_0$  with  $0 < \delta_0 < 1$ ; we further specify  $\delta_0$  in the context of (5.7) below. As has been verified in [14, Proposition 2.1], there are positive constants  $c_\beta$  and  $B$ , such that every  $u \in \mathcal{B}_\gamma(u_0)$  satisfies

$$\int_{\mathbb{R}^d} |e^{-\beta u(x)} - 1| dx \leq c_\beta, \tag{5.3}$$

and the interaction energy of every configuration  $\gamma \subset \mathbb{R}^d$  with  $N(\gamma)$  particles is bounded from below by

$$U(\gamma) \geq -BN(\gamma). \tag{5.4}$$

Moreover, for every  $u \in \mathcal{B}_\gamma(u_0)$  and  $\mu \in \mathbb{R}$  and every associated  $(\mu, u)$ -Gibbs measure, the sequence

$$\rho(\mu, u) = [\rho, \rho^{(2)}, \dots]^T$$

of its  $m$ -particle correlation functions is a solution of the so-called *Kirkwood-Salsburg equations* (cf., e.g., [39, Corollary 5.3], or compare (A.3) in the appendix), which can be written in the form

$$(I - e^{\beta\mu} A(u))\rho(\mu, u) = e^{\beta\mu} \mathbf{e}_1, \tag{5.5}$$

where  $\mathbf{e}_1 = [1, 0, \dots]^T$ ,  $I$  is the identity operator, and  $A(u)$  is a bounded operator in an associated Banach space  $\mathcal{X}$  of sequences of  $L^\infty$  functions. This operator  $A(u)$  is given by a certain infinite-dimensional matrix of integral operators, compare [14, 38], and is Fréchet differentiable with respect to  $u \in \mathcal{B}_\gamma(u_0)$ .

We fix

$$\mu < \frac{1}{\beta} \log \frac{1}{c_\beta e^{2\beta B+1}}, \tag{5.6}$$

and call the range  $\mu \leq \mu_0$  of chemical potentials the *gas phase* associated with the pair potentials in  $\mathcal{B}_\gamma(u_0)$ . In this gas phase, the Kirkwood-Salsburg equations (5.5) have a unique solution  $\rho(\mu, u)$ , which can be developed in a converging Neumann series in  $\mathcal{X}$ ; in particular, this means that for  $u \in \mathcal{B}_\gamma(u_0)$ , there exists only one  $(\mu, u)$ -Gibbs measure for each chemical potential  $\mu$  within the gas phase. The individual components  $\rho^{(m)}(\mu, u)$  of  $\rho(\mu, u)$  are Fréchet differentiable (in  $L^\infty$ ) with respect to  $u$ , analytic with respect to  $\mu < \mu_0$  and continuous in  $\mu \leq \mu_0$ . It is also easy to see that  $\rho$  is a  $C^1$  function of  $\mu \leq \mu_0$  and  $u \in \mathcal{B}_\gamma(u_0)$ .

Further, for fixed  $u \in \mathcal{B}_\gamma(u_0)$ , the density  $\rho = \rho(\mu, u)$  of the associated Gibbs measure, i.e., the first entry of  $\rho(\mu, u)$ , is differentiable and strictly increasing in  $\mu$  up to  $\mu = \mu_0$  by virtue of Corollary 3.2. Since it is also continuous in  $u \in \mathcal{B}_\gamma(u_0)$ , we can choose  $\delta_0$  in (5.2) so small that

$$\rho_0 = \inf\{\rho(\mu_0, u) : u \in \mathcal{B}_\gamma(u_0)\} \geq (1 - \varepsilon) \rho(\mu_0, u_0) \tag{5.7}$$

for any given  $\varepsilon > 0$ , and then every density  $\rho \in (0, \rho_0)$  belongs to the gas phase of all pair potentials in  $\mathcal{B}_\gamma(u_0)$ . Finally, for fixed  $u \in \mathcal{B}_\gamma(u_0)$ , the pressure  $p(\mu, u)$  is also

differentiable and strictly increasing in  $\mu$  up to  $\mu = \mu_0$ , and its derivative is given by the density, cf. (4.1).

**Proposition 5.2** *Let  $0 < \rho_* < \rho_0$  with  $\rho_0$  as in (5.7). Then, the chemical potential  $\mu_* = \mu_*(u)$  defined in Theorem 4.1 is differentiable with respect to  $u \in \mathcal{B}_{\mathcal{V}}(u_0)$  with derivative  $\mu'_*(u) \in \mathcal{V}'$  given by*

$$\mu'_*(u) v = -\frac{\partial_u \rho(\mu_*(u), u) v}{\partial_\mu \rho(\mu_*(u), u)} \quad \text{for } v \in \mathcal{V}, \tag{5.8}$$

where  $\partial_\mu \rho$  and  $\partial_u \rho$  denote the partial derivatives of  $\rho(\mu, u)$ , and  $\mathcal{V}'$  is the dual space of  $\mathcal{V}$ .

**Proof** Since the pressure and the density of a fixed pair potential  $u \in \mathcal{B}_{\mathcal{V}}(u_0)$  are strictly increasing and differentiable functions of the chemical potential in the gas phase, we can rewrite each of these functions as a strictly increasing function of any of the other ones. The pressure, for example, can be written as a function of density, which we call  $\pi$  to avoid any confusion, i.e.,

$$p(\mu, u) = \pi(\rho(\mu, u), u).$$

Then, the chain rule can be applied to obtain

$$\partial_\mu p(\mu, u) = \partial_\rho \pi(\rho(\mu, u), u) \partial_\mu \rho(\mu, u), \tag{5.9}$$

because it has been shown in [39, Theorem 4.3] that  $\pi$  is a differentiable function of the density in the gas phase. For  $\mu = \mu_*(u)$ , we thus conclude from (4.1) and (5.9) that

$$\rho_* = \rho(\mu_*(u), u) = \partial_\mu p(\mu_*(u), u) = \alpha \partial_\mu \rho(\mu_*(u), u) \tag{5.10}$$

with

$$\alpha = \partial_\rho \pi(\rho_*, u) > 0.$$

It thus follows from (5.10) that  $\partial_\mu \rho(\mu_*(u), u) > 0$ , so that the implicit function theorem is applicable to the equation

$$\rho(\mu_*(u + v), u + v) = \rho_*$$

near  $v = 0$ . From this, we readily obtain (5.8); moreover,  $\mu'_*(u)$  belongs to  $\mathcal{V}'$  because  $\partial_u \rho \in \mathcal{V}'$ . □

Now, we return to our analysis of the relative entropy functional (4.5).

**Theorem 5.3** *Let  $u_0 \in \mathcal{U}$  be a fixed pair potential and  $\mathcal{B}_{\mathcal{V}}(u_0) \subset \mathcal{U}$  be defined as in (5.2) with  $\delta_0$  so small that (5.7) holds. Furthermore, let the target  $\mathbb{P}_*$  of the relative entropy functional have density  $\rho_* < \rho_0$ , and assume that the associated specific*



interaction energy  $E(\cdot, \mathbb{P}_*)$  is bounded on  $\mathcal{B}_{\mathcal{V}}(u_0)$ . Then,  $\Phi$  is Fréchet-differentiable in every  $u \in \mathcal{B}_{\mathcal{V}}(u_0)$  with derivative  $\Phi'(u) \in \mathcal{V}'$ , given by

$$\Phi'(u)v = E(v, \mathbb{P}_*) - \frac{1}{2} \int_{\mathbb{R}^d} v(x)\rho^{(2)}(x) \, dx \tag{5.11}$$

for  $v \in \mathcal{V}$ , where  $\rho^{(2)}$  is the pair correlation function of the associated  $(\mu_*(u), u)$ -Gibbs measure.

**Proof** Let  $u \in \mathcal{B}_{\mathcal{V}}(u_0)$  and  $v \in \mathcal{V}$ . We apply Corollary 3.2 to  $u_1 = u$  and  $u_2 = u + v$  with  $\mu_1 = \mu_*(u)$  and  $\mu_2 = \mu_*(u + v)$ . This yields

$$\begin{aligned} -\frac{1}{2} \int_{\mathbb{R}^d} v(x)\rho^{(2)}(x) \, dx &\leq p_*(u + v) - \mu_*(u + v)\rho_* - (p_*(u) - \mu_*(u)\rho_*) \\ &\leq -\frac{1}{2} \int_{\mathbb{R}^d} v(x)\tilde{\rho}^{(2)}(x) \, dx, \end{aligned}$$

where  $\tilde{\rho}^{(2)}$  denotes the pair correlation function of the corresponding  $(\mu_*(u + v), u + v)$ -Gibbs measure. It follows that

$$\begin{aligned} E(v, \mathbb{P}_*) - \frac{1}{2} \int_{\mathbb{R}^d} v(x)\rho^{(2)}(x) \, dx &\leq \Phi(u + v) - \Phi(u) \\ &\leq E(v, \mathbb{P}_*) - \frac{1}{2} \int_{\mathbb{R}^d} v(x)\tilde{\rho}^{(2)}(x) \, dx, \end{aligned}$$

i.e., that

$$\begin{aligned} 0 &\leq \Phi(u + v) - \Phi(u) - \left( E(v, \mathbb{P}_*) - \frac{1}{2} \int_{\mathbb{R}^d} v(x)\rho^{(2)}(x) \, dx \right) \\ &\leq \frac{1}{2} \int_{\mathbb{R}^d} v(x)(\rho^{(2)} - \tilde{\rho}^{(2)})(x) \, dx. \end{aligned}$$

This shows that

$$\begin{aligned} &\left| \Phi(u + v) - \Phi(u) - \left( E(v, \mathbb{P}_*) - \frac{1}{2} \int_{\mathbb{R}^d} v(x)\rho^{(2)}(x) \, dx \right) \right| \\ &\leq \frac{C_{\psi_0}}{2} \|v\|_{\mathcal{V}} \|\tilde{\rho}^{(2)} - \rho^{(2)}\|_{L^\infty(\mathbb{R}^d)}, \end{aligned}$$

where

$$C_{\psi_0} = |\mathbb{S}^{d-1}| \int_0^\infty r^{d-1} \psi_0(r) \, dr < \infty$$

by virtue of (2.1). We finally observe that

$$\tilde{\rho}^{(2)} = \rho^{(2)}(\mu_*(u + v), u + v)$$

converges to  $\rho^{(2)} = \rho^{(2)}(\mu_*(u), u)$  in  $L^\infty(\mathbb{R}^d)$  as  $\|v\|_{\mathcal{V}} \rightarrow 0$  because of Proposition 5.2 and the smoothness of  $\rho^{(2)}$  as a function of  $\mu$  and  $u$ . We have therefore established that

$$\left| \Phi(u + v) - \Phi(u) - \left( E(v, \mathbb{P}_*) - \frac{1}{2} \int_{\mathbb{R}^d} v(x) \rho^{(2)}(x) dx \right) \right| = o(\|v\|_{\mathcal{V}})$$

as  $\|v\|_{\mathcal{V}} \rightarrow 0$ , which yields (5.11). Since  $E(\cdot, \mathbb{P}_*)$  is bounded on  $\mathcal{B}_{\mathcal{V}}(u_0)$ , it is easy to see that  $E(\cdot, \mathbb{P}_*) \in \mathcal{V}'$ , and hence, it follows in the same way as above that  $\Phi'(u) \in \mathcal{V}'$  with

$$\|\Phi'(u)\|_{\mathcal{V}'} \leq \|E(\cdot, \mathbb{P}_*)\|_{\mathcal{V}'} + \frac{C\psi_0}{2} \|\rho^{(2)}\|_{L^\infty(\mathbb{R}^d)}.$$

This concludes the proof. □

**Remark 5.4** We emphasize that the dual space  $\mathcal{V}'$  of  $\mathcal{V}$  is a complicated Banach space. However, the representation (5.11), in combination with (2.7), reveals that if  $\mathbb{P}_*$  has a bounded and measurable pair correlation function  $\rho_*^{(2)}$ , then  $\Phi'(u)$  can be identified with

$$\nabla\Phi(u) = \frac{1}{2}(\rho_*^{(2)} - \rho^{(2)}(\mu_*(u), u)) \in L^\infty(\mathbb{R}^d), \tag{5.12}$$

when using the natural dual pairing

$$\Phi'(u)v = \langle v, \nabla\Phi(u) \rangle = \int_{\mathbb{R}^d} v(x)(\nabla\Phi(u))(x) dx \tag{5.13}$$

of  $L^2(\mathbb{R}^d)$ .

This is the case, for example, if  $\mathbb{P}_*$  is a  $(\mu, u_*)$ -Gibbs measure for some  $u_* \in \mathcal{U}$  (with  $\mu = \mu_*(u_*)$ ), and then (5.12) can further be rewritten as

$$\nabla\Phi(u) = \frac{1}{2}(\omega^{(2)}(\mu, u_*) - \omega^{(2)}(\mu_*(u), u)),$$

where

$$\omega^{(2)}(\mu, u) = \rho^{(2)}(\mu, u) - \rho(\mu, u)^2. \tag{5.14}$$

It follows from (5.17) below that  $\nabla\Phi(u)$  then also belongs to  $L^1(\mathbb{R}^d)$ , and hence, to  $L^2(\mathbb{R}^d)$  as well. Note that  $\mathcal{V}$  is embedded in  $L^2(\mathbb{R})$  for the same reason. We refer to Sect. 7 for further arguments which support the choice of the  $L^2$ -pairing. ◇

**Remark 5.5** Assume that the target  $\mathbb{P}_*$  has a bounded and measurable pair correlation function  $\rho_*^{(2)}$ , and that the relative entropy functional  $\Phi$  is minimized by some  $u_* \in \mathcal{U}$ . Assume further that the given density  $\rho_*$  belongs to the gas phase of  $u_*$ . Then, it follows from Theorem 5.3 that  $\Phi$  is differentiable at  $u_*$ , and that  $\nabla\Phi(u_*) = 0$ . From (5.12), we thus conclude that the target and the model share the same pair correlation function in this case; compare Remark 4.5. ◇

**Remark 5.6** For a  $(\mu, u)$ -Gibbs measure the *cluster functions*, sometimes also called Ursell functions (e.g., Stell [43], and [15]), or truncated correlation functions (e.g., Jansen [21]), are defined recursively by the constant function  $\omega^{(1)} = \rho$  and, for  $m \geq 2$ , by

$$\omega^{(m)}(x_1, \dots, x_m) = \rho^{(m)}(x_1, \dots, x_m) - \sum_{k=2}^m \sum_{\pi \in \Pi_k^m} \prod_{i=1}^k \omega^{(|\pi_i|)}(x_{\pi_i}), \tag{5.15}$$

cf. [38, p. 87], where  $\Pi_k^m$  denotes the set of partitions  $\pi$  of  $x_1, \dots, x_m$  into  $k$  subsets  $x_{\pi_1}, \dots, x_{\pi_k}$ . Note that it immediately follows from (5.15) that each cluster function inherits the translation and permutation invariance of the correlation functions.

For  $m = 2$ , we recover from (5.15) the definition (5.14), and for fixed  $\mu$  and  $u$ , we adopt the short-hand notation

$$\omega^{(2)}(x) = \rho^{(2)}(x) - \rho^2$$

for  $\omega^{(2)}(x, 0)$  from the pair correlation function. Note that

$$\omega^{(2)}(x) = \omega^{(2)}(-x) \quad \text{for every } x \in \mathbb{R}^d \tag{5.16}$$

because of the translation invariance.

We will make repeatedly use of [38, Theorem 4.4.8] which states that

$$\int_{(\mathbb{R}^d)^{(m-2)}} |\omega^{(m)}(\cdot, 0, x_3, \dots, x_m)| dx_3 \cdots dx_m \in L^1(\mathbb{R}^d), \tag{5.17}$$

provided that  $u \in \mathcal{U}$  and  $\mu \leq \mu_0$ ; see also Lemma A.4 in the appendix. In particular, (5.17) shows that  $\omega^{(2)} \in L^1(\mathbb{R}^d)$ .  $\diamond$

Based on Theorem 5.3, we now compute the Hessian of  $\Phi$ .

**Theorem 5.7** *Under the assumptions of Theorem 5.3, the relative entropy functional (4.5) is twice differentiable, and its Hessian at  $u \in \mathcal{B}_{\mathcal{V}}(u_0)$  is given by*

$$\Phi''(u)(v, \tilde{v}) = -\frac{1}{\partial_{\mu}\rho} (\partial_u \rho \tilde{v})(\partial_u \rho v) - \frac{1}{2} \langle v, \partial_u \rho^{(2)} \tilde{v} \rangle \tag{5.18}$$

in terms of the dual pairing (5.13) of  $L^2(\mathbb{R}^d)$ , where  $v, \tilde{v} \in \mathcal{V}$ , and the partial derivatives  $\partial_u \rho$  and  $\partial_u \rho^{(2)}$  are evaluated at the pair potential  $u$  and the chemical potential  $\mu_*(u)$ .

**Proof** We differentiate (5.11) with respect to  $u$  in the direction  $\tilde{v} \in \mathcal{V}$ . Keeping in mind that  $\rho^{(2)}$  in (5.11) stands for  $\rho^{(2)}(\mu_*(u), u)$ , the chain rule gives

$$\begin{aligned} \Phi''(u)(v, \tilde{v}) &= -\frac{1}{2} \int_{\mathbb{R}^d} v(x) \left( \partial_{\mu}\rho^{(2)} \mu'_*(u) \tilde{v} + \partial_u \rho^{(2)} \tilde{v} \right) (x) dx \\ &= -\frac{\mu'_*(u) \tilde{v}}{2} \langle v, \partial_{\mu}\rho^{(2)} \rangle - \frac{1}{2} \langle v, \partial_u \rho^{(2)} \tilde{v} \rangle, \end{aligned} \tag{5.19}$$

where we have adopted the notation (5.13), and the partial derivatives of  $\rho^{(2)}$ —which belong to  $L^\infty(\mathbb{R}^d)$ —are evaluated at the pair potential  $u$  and the chemical potential  $\mu_*(u)$ .

To determine  $\partial_\mu \rho^{(2)}(\mu, u)$  for an arbitrary chemical potential  $\mu < \mu_0$  we apply Corollary 3.2 with  $u_1 = u$  and  $u_2 = u + w$  for any  $w \in \mathcal{V}$ , and with  $\mu_1 = \mu_2 = \mu$ : As in the proof of Theorem 5.3, we conclude that the pressure is differentiable with respect to  $u$  with derivative  $\partial_u p(\mu, u) \in \mathcal{V}'$  given by

$$\partial_u p(\mu, u) w = -\frac{1}{2} \int_{\mathbb{R}^d} w(x) \rho^{(2)}(x) dx = -\frac{1}{2} \langle w, \rho^{(2)} \rangle,$$

where  $\rho^{(2)}$  is the pair correlation function of the  $(\mu, u)$ -Gibbs measure. Schwarz’s theorem and (4.1) therefore yield

$$\langle w, \partial_\mu \rho^{(2)} \rangle = \partial_\mu \langle w, \rho^{(2)} \rangle = -2 \partial_\mu ((\partial_u p) w) = -2 \partial_u (\partial_\mu p) w = -2 \partial_u \rho w. \tag{5.20}$$

Inserting this into (5.19), we obtain

$$\Phi''(u)(v, \tilde{v}) = (\mu'_*(u) \tilde{v})(\partial_u \rho v) - \frac{1}{2} \langle v, \partial_u \rho^{(2)} \tilde{v} \rangle,$$

and assertion (5.18) thus follows from Proposition 5.2. □

Note that  $\Phi''(u)$  is a continuous bilinear form: This follows readily from the result in [14] that  $\partial_u \rho \in \mathcal{V}'$  and  $\partial_u \rho^{(2)} \in \mathcal{L}(\mathcal{V}, L^\infty(\mathbb{R}^d))$ . We further observe that  $\Phi''$  is continuous in  $u$  because  $\rho$  and  $\rho^{(2)}$  are  $C^1$  functions of  $\mu$  and  $u$ . This implies that  $\Phi''(u)$  is a symmetric bilinear form. Finally, this bilinear form is positive semidefinite, because  $\Phi$  is strictly convex.

### 6 The connection to the IMC iterative scheme

In the preceding section, we have seen that the relative entropy functional is twice differentiable at any  $u_0 \in \mathcal{U}$ , provided that the target density  $\rho_*$  belongs to its gas phase. The convex quadratic approximation

$$\Phi_2(u_0 + v) := \Phi(u_0) + \Phi'(u_0)v + \frac{1}{2} \Phi''(u_0)(v, v) \approx \Phi(u_0 + v), \tag{6.1}$$

valid for small enough  $v \in \mathcal{V}$ , attains its minimum for every solution  $v_0 \in \mathcal{V}$  of the variational problem

$$\Phi''(u_0)(v_0, w) = -\Phi'(u_0)w \quad \text{for all } w \in \mathcal{V}, \tag{6.2}$$

and the well-known Newton scheme for minimizing  $\Phi$  consists in updating  $u_0$  via

$$u_1 = u_0 + v_0 \tag{6.3}$$

to obtain a (hopefully) better approximation of the global unique minimizer of  $\Phi$ . Unfortunately, however, we only know that  $\Phi''(u_0)$  is semidefinite; it is therefore not clear whether (6.2) has a solution, and if it does, whether it is unique.

The bilinear form  $\Phi''(u_0)$  defines a linear operator  $A : \mathcal{V} \rightarrow \mathcal{V}'$  via the identity

$$\Phi''(u_0)(v, \tilde{v}) = \langle \tilde{v}, Av \rangle_{\mathcal{V}' \times \mathcal{V}} \quad \text{for all } v, \tilde{v} \in \mathcal{V}.$$

As in Remark 5.4 the dual form on the right-hand side can be replaced by the dual pairing induced by  $L^2(\mathbb{R}^d)$ , when  $A$  is identified with the (bounded) operator

$$A : \mathcal{V} \rightarrow L^\infty(\mathbb{R}^d), \quad A : v \mapsto -\frac{1}{2} \partial_u \rho^{(2)} v - \frac{1}{\partial_{\mu} \rho} (\partial_u \rho v) \nabla_u \rho, \quad (6.4)$$

where  $\nabla_u \rho \in L^\infty(\mathbb{R}^d)$  denotes the representative of the functional  $\partial_u \rho \in \mathcal{V}'$  for this pairing. This is an immediate consequence of Theorem 5.7. According to [15, Sect. 6] there holds

$$\nabla_u \rho(x) = -\beta \rho^{(2)}(x) - \frac{\beta}{2} \int_{\mathbb{R}^d} \chi^{(3)}(0, x + x', x') \, dx', \quad (6.5)$$

where

$$\chi^{(3)}(0, x + x', x') = \omega^{(3)}(0, x + x', x') + \rho \omega^{(2)}(x') + \rho \omega^{(2)}(x + x')$$

is given in terms of the second and third cluster functions associated with  $\mu$  and  $u$ ; compare Remark 5.6. Using the translation and permutation invariance of  $\omega^{(3)}$ , this can be rewritten as

$$\chi^{(3)}(0, x + x', x') = \omega^{(3)}(x, 0, -x') + \rho \omega^{(2)}(x') + \rho \omega^{(2)}(x + x'). \quad (6.6)$$

Note that it has been shown in [15, Proposition 4.2] that the integral in (6.5) is absolutely convergent and that it defines a bounded function of  $x \in \mathbb{R}^d$ . Since  $\omega^{(2)} \in L^1(\mathbb{R}^d)$ , it thus follows from (6.6) that

$$\int_{\mathbb{R}^d} |\omega^{(3)}(\cdot, 0, -x')| \, dx' \in L^\infty(\mathbb{R}^d), \quad (6.7)$$

and inserting (6.6) into (6.5), we finally arrive at the representation

$$\nabla_u \rho(x) = -\beta \rho^{(2)}(x) - \frac{\beta}{2} \int_{\mathbb{R}^d} \omega^{(3)}(x, 0, x') \, dx' - \beta \rho \int_{\mathbb{R}^d} \omega^{(2)}(x') \, dx', \quad (6.8)$$

which will be used later on.

Now, we introduce the (nonlinear) ‘‘Henderson operator’’

$$F : u \mapsto \rho^{(2)}(\mu_*(u), u), \quad (6.9)$$

which maps  $u \in \mathcal{B}_\gamma(u_0)$  to its associated pair correlation function at the given density  $\rho_*$ , and which is an injective operator according to the Henderson theorem. Since  $F(u) + 2\nabla\Phi(u)$  is independent of  $u$  according to (5.11), see also (5.12), it follows that  $F : \mathcal{B}_\gamma(u_0) \rightarrow L^\infty(\mathbb{R}^d)$  is differentiable for  $\rho_* < \rho_0$  with

$$\langle \tilde{v}, F'(u_0)v \rangle = -2 \Phi''(u_0)(v, \tilde{v}). \tag{6.10}$$

This shows that if the target  $\mathbb{P}_*$  has a pair correlation function  $\rho_*^{(2)} \in L^\infty(\mathbb{R}^d)$ , then the variational problem (6.2) is equivalent to the linear operator equation

$$F'(u_0)v_0 = \rho_*^{(2)} - F(u_0). \tag{6.11}$$

The step (6.3) for minimizing  $\Phi$  with Newton’s method is thus equivalent to one iteration of the IMC method, where the update  $v_0$  is determined by solving (6.11). This is the thermodynamic limit analog of the observation in [29, 36], referred to in the introduction.

**Remark 6.1** As follows from Theorem 5.3 and (2.7), the Henderson operator satisfies

$$\langle v, F(u_0) \rangle = 2 E(v, \mathbb{P}_0),$$

where  $\mathbb{P}_0$  denotes the corresponding  $(\mu_*(u_0), u_0)$ -Gibbs measure. In the IMC scheme, as introduced in [26], this interaction energy is approximated by the corresponding expectation value

$$E(v, \mathbb{P}_0) \approx \frac{1}{|\Lambda|} \langle V \rangle_{\rho_*, \Lambda}$$

of the *canonical ensemble* corresponding to the pair potential  $u_0$  at density  $\rho_*$  in the bounded box  $\Lambda \subset \mathbb{R}^d$ , where  $V$  is the observable

$$V(\gamma) = \sum_{1 \leq i < j \leq N} v(x_i - x_j)$$

associated with the perturbation  $v \in \mathcal{V}$ , compare (2.3). Likewise, the derivative  $F'(u_0)$  is approximated in the IMC scheme via the cross-correlation matrix

$$-\frac{2\beta}{|\Lambda|} \left( \langle V \tilde{V} \rangle_{\rho_*, \Lambda} - \langle V \rangle_{\rho_*, \Lambda} \langle \tilde{V} \rangle_{\rho_*, \Lambda} \right) \approx \langle \tilde{v}, F'(u_0)v \rangle.$$

It has to be emphasized that the density constraint  $\rho(u) = \rho_*$  is inherently built into the canonical ensemble. Accordingly, this approximation of  $F'(u_0)$  does respect this constraint. When working in the grand canonical ensemble instead (or in the thermodynamic limit), this constraint necessitates a correction term to project the unconstrained derivative into the tangent plane of the manifold of pair correlation functions corresponding to particle systems with the correct density. This is the reason

for the first term on the right-hand side of (5.18) which does not (and need not) occur in the original IMC scheme.  $\diamond$

### 7 Extension of the Hessian to a semidefinite bilinear form on $L^2(\mathbb{R}^d)$

In the remainder of this paper, we study the Hessian of the relative entropy functional somewhat further. More precisely, we examine the mapping properties of the Jacobian of the Henderson map, which is connected to the Hessian  $\Phi''$  via (6.10).

To begin with, recall from Remark 5.4 that the right-hand side of (6.11) belongs to  $L^\infty(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$ , when the target  $\mathbb{P}_*$  is a  $(\mu, u_*)$ -Gibbs measure. As our first result of this section, we show that  $F'(u_0)$  has matching mapping properties.

**Theorem 7.1** *Under the assumptions of Theorem 5.3, the Jacobian  $F'(u_0)$  of the Henderson operator (6.9) belongs to  $\mathcal{L}(\mathcal{V}, L^\infty(\mathbb{R}^d)) \cap \mathcal{L}(\mathcal{V}, L^1(\mathbb{R}^d))$ .*

**Proof** We already know from Sect. 6 that  $F'(u_0) = -2A \in \mathcal{L}(\mathcal{V}, L^\infty(\mathbb{R}^d))$ . It thus remains to establish that  $F'(u_0) \in \mathcal{L}(\mathcal{V}, L^1(\mathbb{R}^d))$ . To this end, we rewrite the Henderson operator in terms of the cluster function (5.14), i.e.,

$$F(u_0) = \omega^{(2)}(\mu_*(u_0), u_0) + \rho_*^2,$$

which yields

$$F'(u_0)v = \partial_\mu \omega^{(2)}(\mu_*, u_0) \mu'_*(u_0)v + \partial_u \omega^{(2)}(\mu_*, u_0)v, \tag{7.1}$$

where  $\mu_* = \mu_*(u_0)$ . The partial derivative of the cluster function with respect to the chemical potential is given by

$$\begin{aligned} \partial_\mu \omega^{(2)}(\mu, u) &= \partial_\mu (\rho^{(2)}(\mu, u) - \rho(\mu, u)^2) \\ &= \partial_\mu \rho^{(2)}(\mu, u) - 2\rho(\mu, u)\partial_\mu \rho(\mu, u). \end{aligned} \tag{7.2}$$

From (5.20) and (6.8), we conclude that

$$\begin{aligned} \partial_\mu \rho^{(2)}(x) &= -2 \nabla_u \rho(x) \\ &= 2\beta \rho^{(2)}(x) + \beta \int_{\mathbb{R}^d} \omega^{(3)}(x, 0, x') dx' + 2\beta \rho \int_{\mathbb{R}^d} \omega^{(2)}(x') dx', \end{aligned}$$

and hence, together with Lemma A.2, it follows that

$$\partial_\mu \omega^{(2)}(x) = 2\beta \omega^{(2)}(x) + \beta \int_{\mathbb{R}^d} \omega^{(3)}(x, 0, x') dx'. \tag{7.3}$$

In particular, this shows that  $\partial_\mu \omega^{(2)} \in L^1(\mathbb{R}^d)$  by virtue of (5.17). Together with Proposition 5.2 and (7.1), we thus obtain

$$\|F'(u_0)v\|_{L^1(\mathbb{R}^d)} \leq \left( \|\partial_\mu \omega^{(2)}(\mu_*, u_0)\|_{L^1(\mathbb{R}^d)} \|\mu'_*(u_0)\|_{\mathcal{V}'} + \|\partial_u \omega^{(2)}(\mu_*, u_0)\|_{\mathcal{L}(\mathcal{V}, L^1(\mathbb{R}^d))} \right) \|v\|_{\mathcal{V}},$$

because it has been proved in [15] that  $\partial_u \omega^{(2)} \in \mathcal{L}(\mathcal{V}, L^1(\mathbb{R}^d))$ . This shows that  $F'(u_0) \in \mathcal{L}(\mathcal{V}, L^1(\mathbb{R}^d))$ .  $\square$

Since  $L^\infty(\mathbb{R}^d) \cap L^1(\mathbb{R}^d) \subset L^2(\mathbb{R}^d)$ , Theorem 7.1 shows that  $F'(u_0)$  is a bounded operator from  $\mathcal{V}$  to  $L^2(\mathbb{R}^d)$ . As  $\mathcal{V}$  is a dense subspace of  $L^2(\mathbb{R}^d)$ , and because we have repeatedly identified the dual pairing  $\mathcal{V} \times \mathcal{V}'$  with the dual pairing of  $L^2(\mathbb{R}^d)$ , this raises the question whether  $F'(u_0)$  extends to a bounded operator in  $L^2(\mathbb{R}^d)$ , or, in terms of the relative entropy functional, whether  $\Phi''(u_0)$  extends to a quadratic form on  $L^2(\mathbb{R}^d)$ . This will be answered to the affirmative in the remainder of this section.

**Lemma 7.2** *Under the assumptions of Theorem 5.3, the Jacobian of the Henderson map is given by*

$$(F'(u_0)v)(x) = -\beta\rho^{(2)}(x)v(x) \tag{7.4a}$$

$$- 2\beta\rho(\omega^{(2)} * v)(x) - \beta(\omega^{(2)} * \omega^{(2)} * v)(x) \tag{7.4b}$$

$$+ \frac{1}{2\partial_\mu \rho} \partial_\mu \omega^{(2)}(x) \int_{\mathbb{R}^d} \partial_\mu \omega^{(2)}(x') v(x') dx' \tag{7.4c}$$

$$- 2\beta \int_{\mathbb{R}^d} \omega^{(3)}(x, 0, x') v(x') dx' \tag{7.4d}$$

$$- \frac{\beta}{2} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \omega^{(4)}(x, 0, x'', x' + x'') dx'' v(x') dx' \tag{7.4e}$$

for  $v \in \mathcal{V}$  and  $x \in \mathbb{R}^d$ .

**Proof** According to (6.4), we have

$$F'(u_0)v = -2Av = \partial_u \rho^{(2)}v + \frac{2}{\partial_\mu \rho} (\partial_u \rho v) \nabla_u \rho,$$

where

$$\begin{aligned} (\partial_u \rho^{(2)}v)(x) &= -\beta\rho^{(2)}(x)v(x) - 2\beta \int_{\mathbb{R}^d} \rho^{(3)}(x, 0, x')v(x') dx' \\ &\quad - \frac{\beta}{2} \int_{\mathbb{R}^d} v(x') \int_{\mathbb{R}^d} (\rho^{(4)}(x, 0, x'', x'' + x') - \rho^{(2)}(x)\rho^{(2)}(x')) dx'' dx' \end{aligned} \tag{7.5}$$

by virtue of [15, Eq. (6.9)]. Concerning the nested integral in the second line of (7.5), it has further been shown in [15, Proposition 4.3] that the inner integral converges absolutely and defines an  $L^\infty$  function of  $x$  and  $x'$ . It follows that

$$(F'(u_0)v)(x) = -\beta\rho^{(2)}(x)v(x) + \int_{\mathbb{R}^d} k(x, x')v(x') dx' \tag{7.6}$$



with integral kernel

$$\begin{aligned}
 k(x, x') &= \frac{2}{\partial_\mu \rho} \nabla_u \rho(x) \nabla_u \rho(x') - 2\beta \rho^{(3)}(x, 0, x') \\
 &\quad - \frac{\beta}{2} \int_{\mathbb{R}^d} (\rho^{(4)}(x, 0, x'', x'' + x') - \rho^{(2)}(x) \rho^{(2)}(x')) dx'' .
 \end{aligned}
 \tag{7.7}$$

We now rewrite the individual terms of this kernel function. To begin with, we recall from (5.20) and (7.2) that

$$\nabla_u \rho = -\frac{1}{2} \partial_\mu \rho^{(2)} = -\frac{1}{2} \partial_\mu \omega^{(2)} - \rho \partial_\mu \rho ,$$

so that

$$\begin{aligned}
 \frac{2}{\partial_\mu \rho} \nabla_u \rho(x) \nabla_u \rho(x') &= \frac{1}{2\partial_\mu \rho} \partial_\mu \omega^{(2)}(x) \partial_\mu \omega^{(2)}(x') \\
 &\quad + \rho \partial_\mu \omega^{(2)}(x) + \rho \partial_\mu \omega^{(2)}(x') + 2\rho^2 \partial_\mu \rho .
 \end{aligned}$$

The partial derivatives with respect to  $\mu$  in the final three terms of the right-hand side can be eliminated with the help of (7.3) and Lemma A.2: This yields

$$\begin{aligned}
 \frac{2}{\partial_\mu \rho} \nabla_u \rho(x) \nabla_u \rho(x') &= \frac{1}{2\partial_\mu \rho} \partial_\mu \omega^{(2)}(x) \partial_\mu \omega^{(2)}(x') \\
 &\quad + 2\beta \rho \omega^{(2)}(x) + \beta \rho \int_{\mathbb{R}^d} \omega^{(3)}(x, 0, x'') dx'' + 2\beta \rho \omega^{(2)}(x') \\
 &\quad + \beta \rho \int_{\mathbb{R}^d} \omega^{(3)}(x', 0, x'') dx'' + 2\beta \rho^3 + 2\beta \rho^2 \int_{\mathbb{R}^d} \omega^{(2)}(x'') dx'' .
 \end{aligned}
 \tag{7.8}$$

Concerning the second term of the integral kernel (7.7), we apply definition (5.15) of the cluster functions to rewrite

$$\rho^{(3)}(x, 0, x') = \omega^{(3)}(x, 0, x') + \rho \omega^{(2)}(x) + \rho \omega^{(2)}(x - x') + \rho \omega^{(2)}(x') + \rho^3 ,
 \tag{7.9}$$

where we have also used (5.16).

In the same way, we can rewrite the integrand of the integral in (7.7) in terms of the cluster functions:

$$\begin{aligned}
 &\rho^{(4)}(x, 0, x'', x'' + x') - \rho^{(2)}(x) \rho^{(2)}(x') \\
 &= \omega^{(4)}(x, 0, x'', x'' + x') + \omega^{(3)}(x, 0, x'') \rho + \omega^{(3)}(x, 0, x'' + x') \rho \\
 &\quad + \omega^{(3)}(x, x'', x'' + x') \rho + \omega^{(3)}(0, x'', x'' + x') \rho \\
 &\quad + \omega^{(2)}(x - x'') \omega^{(2)}(x'' + x') + \omega^{(2)}(x - x' - x'') \omega^{(2)}(x'') \\
 &\quad + \omega^{(2)}(x'' - x) \rho^2 + \omega^{(2)}(x'' + x' - x) \rho^2 \\
 &\quad + \omega^{(2)}(x'') \rho^2 + \omega^{(2)}(x'' + x') \rho^2
 \end{aligned}$$

$$\begin{aligned}
 &= \omega^{(4)}(x, 0, x'', x'' + x') + \omega^{(3)}(x, 0, x'') \rho + \omega^{(3)}(x, 0, x'' + x') \rho \\
 &\quad + \omega^{(3)}(x', 0, x - x'') \rho + \omega^{(3)}(x', 0, -x'') \rho \\
 &\quad + \omega^{(2)}(x - x'') \omega^{(2)}(x'' + x') + \omega^{(2)}(x - x' - x'') \omega^{(2)}(x'') \\
 &\quad + \omega^{(2)}(x'' - x) \rho^2 + \omega^{(2)}(x'' + x' - x) \rho^2 \\
 &\quad + \omega^{(2)}(x'') \rho^2 + \omega^{(2)}(x'' + x') \rho^2,
 \end{aligned}$$

where we have used in the second step that the cluster functions are translation and permutation invariant. Integrating over  $x'' \in \mathbb{R}^d$ , we thus obtain

$$\begin{aligned}
 &\int_{\mathbb{R}^d} (\rho^{(4)}(x, 0, x'', x'' + x') - \rho^{(2)}(x) \rho^{(2)}(x')) dx'' \\
 &= \int_{\mathbb{R}^d} \omega^{(4)}(x, 0, x'', x'' + x') dx'' + 4\rho^2 \int_{\mathbb{R}^d} \omega^{(2)}(x'') dx'' \\
 &\quad + 2\rho \int_{\mathbb{R}^d} \omega^{(3)}(x, 0, x'') dx'' + 2\rho \int_{\mathbb{R}^d} \omega^{(3)}(x', 0, x'') dx'' \\
 &\quad + (\omega^{(2)} * \omega^{(2)})(x + x') + (\omega^{(2)} * \omega^{(2)})(x - x'),
 \end{aligned} \tag{7.10}$$

where all the terms on the right-hand side are bounded functions of  $x$  and  $x'$ : This follows from (6.7) and the fact that  $\omega^{(2)} \in L^1(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d)$ .

Inserting (7.8), (7.9), and (7.10) into (7.7), we conclude that

$$\begin{aligned}
 k(x, x') &= \frac{1}{2\partial_\mu \rho} \partial_\mu \omega^{(2)}(x) \partial_\mu \omega^{(2)}(x') - 2\beta \omega^{(3)}(x, 0, x') - 2\beta \rho \omega^{(2)}(x - x') \\
 &\quad - \frac{\beta}{2} \int_{\mathbb{R}^d} \omega^{(4)}(x, 0, x'', x'' + x') dx'' \\
 &\quad - \frac{\beta}{2} (\omega^{(2)} * \omega^{(2)})(x + x') - \frac{\beta}{2} (\omega^{(2)} * \omega^{(2)})(x - x').
 \end{aligned}$$

Note that the final two terms of this integral kernel representation define the same convolution operator

$$v \mapsto -\frac{\beta}{2} \int_{\mathbb{R}^d} (\omega^{(2)} * \omega^{(2)})(\cdot - x') v(x') dx',$$

because every  $v \in \mathcal{V}$  is an even function. The assertion (7.4) thus follows readily from (7.6). □

Now, we can prove the main result of this section.

**Theorem 7.3** *Under the assumptions of Theorem 5.3, the operator  $F'(u_0)$  extends to a self-adjoint negative semidefinite operator on  $L^2(\mathbb{R}^d)$ .*

**Proof** We already know from Theorem 7.1 that  $F'(u_0) \in \mathcal{L}(\mathcal{V}, L^2(\mathbb{R}^d))$ . Since  $\mathcal{V}$  is a dense subset of  $L^2(\mathbb{R}^d)$ , it remains to discuss the continuity from  $L^2(\mathbb{R}^d)$  to  $L^2(\mathbb{R}^d)$  of each of the terms defined in lines (7.4a)–(7.4e) of Lemma 7.2.

For the multiplication operator in (7.4a), this is true because the pair correlation function  $\rho^{(2)}$  is bounded. The autoconvolution  $\omega^{(2)} * \omega^{(2)}$  of the cluster function  $\omega^{(2)} \in L^1(\mathbb{R}^d)$  belongs to  $L^1(\mathbb{R}^d)$  again, hence the two convolution operators in (7.4b) are bounded operators in  $L^2(\mathbb{R}^d)$ . From (7.3), it follows that  $\partial_\mu \omega^{(2)}$  belongs to  $L^1(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d)$ : the  $L^1$  property has been verified in the argument following (7.3); the  $L^\infty$  property of the second term in (7.3) has been established in (6.7). This shows that  $\partial_\mu \omega^{(2)} \in L^2(\mathbb{R}^d)$ , and the continuity of the operator in (7.4c) is therefore a consequence of the Cauchy–Schwarz inequality. By virtue of (5.17), the cluster function  $\omega^{(3)}(\cdot, 0, \cdot)$  belongs to  $L^1(\mathbb{R}^d \times \mathbb{R}^d)$ . Since it is bounded, it also belongs to  $L^2(\mathbb{R}^d \times \mathbb{R}^d)$ . Therefore, (7.4d) defines a compact operator in  $\mathcal{L}(L^2(\mathbb{R}^d))$ . Concerning (7.4e), we have already pointed out after (7.10) that the inner integral of  $\omega^{(4)}$  is a bounded function of  $x$  and  $x'$ . It also belongs to  $L^1(\mathbb{R}^d \times \mathbb{R}^d)$  by virtue of (5.17), again. The continuity of the operator defined in (7.4e) therefore follows in the same way as in the previous case.

In summary, this shows that  $F'(u_0)$  extends to an operator in  $\mathcal{L}(L^2(\mathbb{R}^d))$ . By virtue of (6.10), this extension is self-adjoint and negative semidefinite.  $\square$

Concerning the Hessian of the relative entropy functional, we readily conclude from (6.10) the following corollary.

**Corollary 7.4** *Under the assumptions of Theorem 5.3 the Hessian  $\Phi''(u_0)$  of the relative entropy functional (4.5) defines a symmetric and positive semidefinite bilinear form on  $L^2(\mathbb{R}^d)$ .*

As indicated before, we leave it as an open problem whether  $F'(u_0)$  is injective, i.e., whether the local quadratic approximation (6.1) of the relative entropy functional is strictly convex.

## 8 Lennard-Jones-type pair potentials

In this final section, we establish a stronger regularity result for the derivative of the Henderson map, respectively, the Hessian of the relative entropy functional, which is valid when the potential  $u_0 \in \mathcal{U}$  satisfies (2.2) with a majorant of the form

$$\psi_0(r) = C(1 + r^2)^{-\alpha/2} \tag{8.1}$$

for some  $C > 0$  and  $\alpha > d$ . This class of potentials includes the so-called potentials of Lennard-Jones type, cf. [38].

It is known that in the gas phase, the cluster function  $\omega^{(2)}$  corresponding to such a pair potential satisfies the same rate of decay near infinity as  $\psi_0$  does. This entails the following result.

**Theorem 8.1** *Let  $u_0$  satisfy (2.2) for some  $\varphi$  as in (2.1) and  $\psi = \psi_0$  of (8.1). Let the space  $\mathcal{V}$  of perturbations of  $u_0$  be defined as before, cf. (5.1), and let  $\rho_*$  belong to the gas phase of  $u_0$ . Then,  $F'(u_0) \in \mathcal{L}(\mathcal{V})$ .*

**Proof** The aforementioned result about the cluster function states that  $\omega^{(2)} \in \mathcal{V}$  under the given assumptions; compare Lemma A.4 in the appendix.

Now, we discuss the continuity of each of the operators corresponding to the individual terms in (7.4). Continuity obviously holds for the multiplication operator in (7.4a), because  $\rho^{(2)}$  is bounded. Concerning the convolutions in (7.4b), we refer to [16, Proposition 4.1] for the fact that the convolution is a continuous bilinear mapping from  $\mathcal{V} \times \mathcal{V}$  to  $\mathcal{V}$ . Accordingly, the two terms in (7.4b) also define operators in  $\mathcal{L}(\mathcal{V})$ .

For the remaining terms, we make use of Lemma A.4 again. For  $m = 3$ , this auxiliary result gives

$$\int_{\mathbb{R}^d} |\omega^{(3)}(\cdot, 0, x')| dx' \in \mathcal{V}, \tag{8.2}$$

and since

$$\left| \int_{\mathbb{R}^d} \omega^{(3)}(x, 0, x') v(x') dx' \right| \leq \|v\|_{L^\infty(\mathbb{R}^d)} \int_{\mathbb{R}^d} |\omega^{(3)}(x, 0, x')| dx'$$

for every  $x \in \mathbb{R}^d$ , it follows that

$$\left\| \int_{\mathbb{R}^d} \omega^{(3)}(\cdot, 0, x') v(x') dx' \right\|_{\mathcal{V}} \leq \|v\|_{L^\infty(\mathbb{R}^d)} \left\| \int_{\mathbb{R}^d} |\omega^{(3)}(\cdot, 0, x')| dx' \right\|_{\mathcal{V}}.$$

Much the same argument applies to the term in (7.4e):

$$\begin{aligned} & \left| \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \omega^{(4)}(x, 0, x'', x' + x'') dx'' v(x') dx' \right| \\ & \leq \|v\|_{L^\infty(\mathbb{R}^d)} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |\omega^{(4)}(x, 0, x'', x')| dx' dx'' \end{aligned}$$

for every  $x \in \mathbb{R}^d$ , and the right-hand side again belongs to  $\mathcal{V}$  as a function of  $x$  by virtue of Lemma A.4 for  $m = 4$ . Since  $\mathcal{V}$  is continuously embedded in  $L^\infty(\mathbb{R}^d)$  this shows that the two terms in (7.4d) and (7.4e) correspond to operators in  $\mathcal{L}(\mathcal{V})$ .

Finally, concerning the term in (7.4c), we conclude from (7.3), (8.2), and the fact that  $\omega^{(2)} \in \mathcal{V}$  that  $\partial_\mu \omega^{(2)}$  belongs to  $\mathcal{V}$  as well. Accordingly, this term also represents a bounded operator from  $\mathcal{V}$  to  $\mathcal{V}$ , and the proof is done.  $\square$

Note that if the target  $\mathbb{P}_*$  and the model  $\mathbb{P}_0$  are Gibbs measures corresponding to pair potentials  $u_*$  and  $u_0$  in  $\mathcal{U}$  and their associated chemical potentials, where  $u_*$  and  $u$  satisfy (2.2) for the same majorant  $\psi = \psi_0$  given by (8.1), and if the density  $\rho_*$  belongs to the gas phase of both pair potentials, then the cluster functions  $\omega_*^{(2)}$  and  $\omega_0^{(2)}$  of  $\mathbb{P}_*$  and  $\mathbb{P}_0$ , respectively, both belong to  $\mathcal{V}$ . It thus follows from Remark 5.4 that the gradient  $\nabla \Phi(u_0)$  of the relative entropy functional belongs to  $\mathcal{V}$  as well, and therefore the Newton equation (6.11) of the IMC iterative scheme is an operator equation in  $\mathcal{V}$  by virtue of Theorem 8.1. As shown in [16, Remark 6.5], a valid choice  $u_0 \in \mathcal{U}$ , which satisfies (2.2) for the same majorant (8.1), is given by the potential of mean

force,

$$u_0 = -\frac{1}{\beta} \log(\rho_*^{(2)}/\rho_*^2) = -\frac{1}{\beta} \log(1 + \omega_*^{(2)}/\rho_*^2),$$

provided the density  $\rho_*$  is sufficiently small. Moreover,  $u_* - u_0$  belongs to  $\mathcal{V}$  for this initial guess.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Data Availability Statement** Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

### Declarations

**Conflict of interest** On behalf of both authors, the corresponding author states that there is no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

### Appendix

In this appendix, we provide some auxiliary results. In doing so, we will repeatedly make use of the short-hand notation

$$\mathbf{x}_n = (x_1, \dots, x_n) \in (\mathbb{R}^d)^n.$$

The first result is needed for the proof of Theorem 4.1, where we construct a  $(\mu, u)$ -Gibbs measure for a pair potential  $u \in \mathcal{U}$  with prescribed density  $\rho_* < \rho_{\text{cp}}(u)$ .

**Lemma A.1** *Let  $u \in \mathcal{U}$  and assume that the sequence  $(\mu_k)_k \subset \mathbb{R}$  satisfies  $\mu_k \rightarrow \mu \in \mathbb{R}$  as  $k \rightarrow \infty$ . Furthermore, for every  $k \in \mathbb{N}$  let  $\mathbb{P}_k$  be a  $(\mu_k, u)$ -Gibbs measure. Then,  $(\mathbb{P}_k)_k$  has an accumulation point with respect to the local topology, and every such accumulation point is a  $(\mu, u)$ -Gibbs measure.*

**Proof** Let  $(\rho_k^{(m)})_m$  be the sequence of correlation functions of  $\mathbb{P}_k$ . Since the sequence  $(\mu_k)_k$  is bounded, the measures  $\mathbb{P}_k$  satisfy a uniform Ruelle condition by virtue of [39, Corollary 5.3], i.e., there exists  $\xi > 0$ , such that

$$\|\rho_k^{(m)}\|_{L^\infty((\mathbb{R}^d)^m)} \leq \xi^m \quad \text{for every } k, m \in \mathbb{N}. \tag{A.1}$$

This implies that for fixed  $m \in \mathbb{N}$ , the sequence  $(\rho_k^{(m)})_k$  of correlation functions has a weak\* convergent subsequence in  $L^\infty((\mathbb{R}^d)^m)$  with limit  $\rho^{(m)}$ , say. With a diagonalization argument, we can thus find a subsequence of indices  $k \in \mathbb{N}$ , for which

$$\rho_k^{(m)} \rightharpoonup^* \rho^{(m)}, \quad k \rightarrow \infty, \tag{A.2}$$

holds for every  $m \in \mathbb{N}$ . This is equivalent to saying that  $(\mathbb{P}_k)_k$  has a subsequence which converges to some probability measure  $\mathbb{P}$  in the local topology. To simplify the notation, we assume in the sequel that the entire sequence  $(\mathbb{P}_k)_k$  is convergent.

Since every  $\mathbb{P}_k$  is a Gibbs measure, the respective correlation functions satisfy the Kirkwood-Salsburg equations, cf., e.g., [39, Corollary 5.3]: For each  $k \in \mathbb{N}$  and every  $m \in \mathbb{N}_0$ , there holds

$$\rho_k^{(m+1)}(x_0, \mathbf{x}_m) = e^{\beta\mu_k} e^{-\beta W(x_0; \mathbf{x}_m)} \cdot \left( \rho_k^{(m)}(\mathbf{x}_m) + \sum_{n=1}^\infty \frac{1}{n!} \int_{(\mathbb{R}^d)^n} \prod_{j=1}^n f(x_0 - y_j) \rho_k^{(m+n)}(\mathbf{x}_m, \mathbf{y}_n) \, d\mathbf{y}_n \right), \tag{A.3}$$

where  $\rho_k^{(0)} = 1, \mathbf{y}_n = (y_1, \dots, y_n) \in (\mathbb{R}^d)^n$ ,

$$f(x) = e^{-\beta u(x)} - 1 \tag{A.4}$$

is the well-known Mayer function, and

$$W(x_0; \mathbf{x}_m) = \sum_{i=1}^m u(x_0 - x_i),$$

the latter being taken to be zero for  $m = 0$ . Note that the right-hand side of (A.3) converges in  $L^\infty((\mathbb{R}^d)^{m+1})$  by virtue of the Ruelle condition (A.1) and the fact that the Mayer function (A.4) belongs to  $L^1(\mathbb{R}^d)$ ; compare (5.3). Moreover, since the Ruelle condition is uniform in  $k$ , this convergence is also uniform with respect to  $k$ , and hence, the weak\* convergence (A.2) and the convergence  $\mu_k \rightarrow \mu$  imply that the Kirkwood-Salsburg equations (A.3) also hold true for the limiting correlation functions  $(\rho^{(m)})_m$ . It thus follows from [39, Corollary 5.3] again that  $\mathbb{P}$  is a tempered  $(\mu, u)$ -Gibbs measure. Finally, since  $\rho_k^{(m)}$  is translation invariant for every  $k \in \mathbb{N}$  and every  $m \in \mathbb{N}$ , its weak\* limits  $\rho^{(m)}$  must also be translation invariant, and hence  $\mathbb{P}$  is translation invariant.  $\square$

The remaining two results of the appendix concern properties of the cluster functions (see Remark 5.6) within the gas phase. The proofs make use of their *cluster expansions* (compare [38, Sect. 4.4] and [43])

$$\omega^{(m)}(\mathbf{x}_m) = \sum_{n=m}^\infty \frac{e^{n\beta\mu}}{(n-m)!} \int_{(\mathbb{R}^d)^{n-m}} \phi^{(n)}(\mathbf{x}_n) \, dx_{m+1} \cdots dx_n \tag{A.5}$$

for  $\mathbf{x}_m \in (\mathbb{R}^d)^m$  with

$$\phi^{(n)}(\mathbf{x}_n) = \sum_{\mathcal{C} \in \mathfrak{C}_n} \prod_{(i,j) \in \mathcal{C}} f(x_i - x_j), \tag{A.6}$$

where  $f$  is, again, the Mayer function (A.4), and  $\mathfrak{C}_n$  denotes the set of undirected connected graphs with vertices  $\{1, \dots, n\}$  and edges  $(i, j) \in \mathcal{C}$ , which connect vertices  $i$  and  $j$ ; for  $n = m$  the integral in (A.5) has to be replaced by the value  $\phi^{(m)}(\mathbf{x}_m)$  of its integrand. These cluster expansions are absolutely convergent for  $u_0 \in \mathcal{U}$  and  $\mu \leq \mu_0$  with  $\mu_0$  of (5.6).

**Lemma A.2** *Let  $u_0 \in \mathcal{U}$  and  $\mu < \mu_0$  with  $\mu_0$  as in (5.6). Then, the density  $\rho$  of the corresponding Gibbs measure is differentiable with respect to  $\mu$  and the derivative is given by*

$$\partial_\mu \rho = \beta \rho + \beta \int_{\mathbb{R}^d} \omega^{(2)}(x) dx. \tag{A.7}$$

**Proof** By (A.5) and (5.15), there holds (with  $x_1 = 0$ )

$$\rho = \omega^{(1)}(0) = \sum_{n=1}^{\infty} \frac{e^{n\beta\mu}}{(n-1)!} \sum_{\mathcal{C} \in \mathfrak{C}_n} \int_{(\mathbb{R}^d)^{n-1}} \prod_{(i,j) \in \mathcal{C}} f(x_i - x_j) dx_2 \cdots dx_n.$$

Differentiating with respect to  $\mu$ , we obtain

$$\begin{aligned} \partial_\mu \rho &= \sum_{n=1}^{\infty} \frac{n\beta e^{n\beta\mu}}{(n-1)!} \sum_{\mathcal{C} \in \mathfrak{C}_n} \int_{(\mathbb{R}^d)^{n-1}} \prod_{(i,j) \in \mathcal{C}} f(x_i - x_j) dx_2 \cdots dx_n \\ &= \beta \sum_{n=1}^{\infty} \frac{e^{n\beta\mu}}{(n-1)!} \sum_{\mathcal{C} \in \mathfrak{C}_n} \int_{(\mathbb{R}^d)^{n-1}} \prod_{(i,j) \in \mathcal{C}} f(x_i - x_j) dx_2 \cdots dx_n \\ &\quad + \beta \int_{\mathbb{R}^d} \left( \sum_{n=2}^{\infty} \frac{e^{n\beta\mu}}{(n-2)!} \sum_{\mathcal{C} \in \mathfrak{C}_n} \int_{(\mathbb{R}^d)^{n-2}} \prod_{(i,j) \in \mathcal{C}} f(x_i - x_j) dx_3 \cdots dx_n \right) dx_2 \\ &= \beta \rho + \beta \int_{\mathbb{R}^d} \omega^{(2)}(0, x_2) dx_2. \end{aligned}$$

Using the short-hand notation for  $\omega^{(2)}$  and (5.16), the assertion follows. □

**Remark A.3** Lemma A.2 is general folklore, at least its “grand canonical version”

$$\partial_\mu \rho_\Lambda(0) = \beta \rho_\Lambda(0) + \beta \int_\Lambda \left( \rho_\Lambda^{(2)}(0, x) - \rho_\Lambda(0) \rho_\Lambda(x) \right) dx \tag{A.8}$$

for a bounded domain  $\Lambda \subset \mathbb{R}^d$  is well-known, where  $\rho_\Lambda$  and  $\rho_\Lambda^{(2)}$  denote the corresponding grand canonical correlation functions; compare, e.g., [43, Eq. (2-9)], or Ben-Naim [1, Eq. (1.53)]. An alternative proof of (A.7) is thus obtained by passing in (A.8) to the thermodynamic limit. ◇

Our final lemma strenghtens Ruelle’s result (5.17) for Lennard-Jones-type pair potentials, and extends [16, Corollary 5.2] to higher-order cluster functions.<sup>2</sup>

**Lemma A.4** *Let  $u \in \mathcal{U}$  be a pair potential which satisfies (2.1), (2.2) with majorant  $\psi = \psi_0$  as in (8.1), and let  $\omega^{(m)}$ ,  $m \in \mathbb{N}$ , be the associated cluster functions. Further, let the chemical potentials satisfy  $\mu \leq \mu_0$ , cf. (5.6). Then,  $\omega^{(2)} \in \mathcal{V}$ , and for  $m \geq 3$ , there holds*

$$\int_{(\mathbb{R}^d)^{m-2}} |\omega^{(m)}(\cdot, 0, x_3, \dots, x_m)| dx_2 \cdots dx_m \in \mathcal{V}.$$

**Proof** For  $\mu < \mu_0$ , it follows from [16, Lemma 5.1] that  $\phi^{(n)}$  of (A.6) satisfies

$$\int_{(\mathbb{R}^d)^{n-2}} |\phi^{(n)}(\mathbf{x}_n)| dx_3 \cdots dx_n \leq C_\Sigma (c_\beta e^{\beta B})^n n^{n-2} \psi_0(|x_1 - x_2|)$$

for some constant  $C_\Sigma > 0$ , every  $x_1, x_2 \in \mathbb{R}^d$  and every  $n \geq 2$ , where  $c_\beta$  and  $B$  are the constants defined in (5.3), (5.4); for  $n = 2$  the left-hand side of this inequality has to be replaced by

$$|\phi^{(2)}(x_1, x_2)| = |f(x_1 - x_2)|.$$

Plugging this into the cluster expansion (A.5) for  $m \geq 3$ , it follows that

$$\begin{aligned} \int_{(\mathbb{R}^d)^{m-2}} |\omega^{(m)}(\mathbf{x}_m)| dx_3 \cdots dx_m &\leq C_\Sigma \psi_0(|x_1 - x_2|) \sum_{n=m}^\infty \frac{n^{n-2}}{(n-m)!} (c_\beta e^{\beta(B+\mu)})^n \\ &\leq C_\Sigma \psi_0(|x_1 - x_2|) \sum_{n=m}^\infty n^{m-2} (c_\beta e^{\beta(B+\mu)+1})^n, \end{aligned}$$

where we have used the inequality  $n^{n-m}/(n-m)! \leq e^n$  in the final step. Since  $\mu \leq \mu_0$ , the positive number  $q = c_\beta e^{\beta(B+\mu)+1}$  is below one, and hence the series on the right-hand side converges, its value being  $C_{\mu,m}$ , say. This estimate extends to  $m = 2$  when the left-hand side is replaced by  $|\omega^{(2)}(\mathbf{x}_2)|$ . The assertion therefore follows from (5.1). □

We refer to [3, 4, 27] for further results on the decay of the higher-order cluster functions.

### References

1. Ben-Naim, A.: Molecular Theory of Solutions. Oxford University Press, New York (2006)
2. Chayes, J.T., Chayes, L.: On the validity of the inverse conjecture in classical density functional theory. J. Stat. Phys. **36**, 471–488 (1984)

<sup>2</sup> The results in [16], like those in [14, 15], extend to the present setting.



3. Dorlas, T.C., Rebenko, A.L., Savoie, B.: Correlation of clusters: partially truncated correlation functions and their decay. *J. Math. Phys.* **61**, 033303 (2020)
4. Duneau, M., Iagolnitzer, D., Souillard, B.: Decay of correlations for infinite-range interactions. *J. Math. Phys.* **16**, 1662–1666 (1975)
5. Ercolessi, F., Adams, J.B.: Interatomic potentials from first-principles calculations: the force-matching method. *Europhys. Lett.* **26**, 583–588 (1994)
6. Fritsch, S., Junghans, C., Kremer, K.: Structure formation of Toluene around C60: implementation of the adaptive resolution scheme (AdResS) into GROMACS. *J. Chem. Theor. Comput.* **8**, 398–403 (2012)
7. Frommer, F., Hanke, M., Jansen, S.: A note on the uniqueness result for the inverse Henderson problem. *J. Math. Phys.* **60**, 093303 (2019)
8. Gallavotti, G., Miracle-Sole, S.: A variational principle for the equilibrium of hard sphere systems. *Ann. IHP Phys. théor.* **8**, 287–299 (1968)
9. Georgii, H.-O.: Large deviations and the equivalence of ensembles for Gibbsian particle systems with superstable interaction. *Probab. Theory Related Fields* **99**, 171–195 (1994)
10. Georgii, H.-O.: The equivalence of ensembles for classical systems of particles. *J. Stat. Phys.* **80**, 1341–1378 (1995)
11. Georgii, H.-O.: *Gibbs Measures and Phase Transitions*. de Gruyter, Berlin (2011)
12. Georgii, H.-O.: *Stochastics: Introduction to Probability and Statistics*, 2nd edn. de Gruyter, Berlin (2013)
13. Georgii, H.-O., Zessin, H.: Large deviations and the maximum entropy principle for marked point random fields. *Probab. Theory Related Fields* **96**, 177–204 (1993)
14. Hanke, M.: Fréchet differentiability of molecular distribution functions I:  $L^\infty$  analysis. *Lett. Math. Phys.* **108**, 285–306 (2018)
15. Hanke, M.: Fréchet differentiability of molecular distribution functions II: the Ursell function. *Lett. Math. Phys.* **108**, 307–329 (2018)
16. Hanke, M.: Well-Posedness of the iterative Boltzmann inversion. *J. Stat. Phys.* **170**, 536–553 (2018)
17. Hansen, J.-P., McDonald, I.R.: *Theory of Simple Liquids: With Applications to Soft Matter*, 4th edn. Academic Press, Oxford (2013)
18. Henderson, R.L.: A uniqueness theorem for fluid pair correlation functions. *Phys. Lett. A* **49**, 197–198 (1974)
19. Hughes, A.M.: *Entropy minimization, convergence and Gibbs ensembles (local and global)*. University of Missouri, PhD-Thesis (2021)
20. Izvekov, S., Voth, G.A.: Multiscale coarse graining of liquid-state systems. *J. Chem. Phys.* **123**, 134105 (2005)
21. Jansen, S.: Cluster expansions for Gibbs point processes. *Adv. Appl. Probab.* **51**, 1129–1178 (2019)
22. Jansen, S., Kuna, T., Tsagarogiannis, D.: Virial inversion and density functionals. [arXiv:1906.02322 \[math-ph\]](https://arxiv.org/abs/1906.02322) (2019)
23. Korolov, L.: An inverse problem for Gibbs fields with hard core potential. *J. Math. Phys.* **48**, 053301 (2007)
24. Kuna, T., Lebowitz, J., Speer, E.: Realizability of point processes. *J. Stat. Phys.* **129**, 417–439 (2007)
25. Larini, L., Lu, L., Voth, G.A.: The multiscale coarse-graining method. VI. Implementation of three-body coarse-grained potentials. *J. Chem. Phys.* **132**, 164107 (2010)
26. Lyubartsev, A.P., Laaksonen, A.: Calculation of effective interaction potentials from radial distribution functions: a reverse Monte Carlo approach. *Phys. Rev. E* **52**, 3730–3737 (1995)
27. Minlos, R.A., Poghosyan, S.: Estimates of Ursell functions, group functions and their derivatives. *Theor. Math. Phys.* **31**, 408–418 (1977)
28. Monticelli, L., Salonen, E. (eds.): *Biomolecular Simulations. Methods and Protocols*. Springer, New York (2013)
29. Murtola, T., Karttunen, M., Vattulainen, I.: Systematic coarse graining from structure using internal states: application to phospholipid/cholesterol bilayer. *J. Chem. Phys.* **131**, 055101 (2009)
30. Navrotskaya, I.: *Inverse problem in classical statistical mechanics*. PhD thesis, University of Pittsburgh (2016)
31. Noid, W.G.: Perspective: coarse-grained models for biomolecular systems. *J. Chem. Phys.* **139**, 090901 (2013)
32. Noid, W.G.: Systematic methods for structurally consistent coarse-grained models. In [28], pp. 487–531 (2013)

33. Peter, C., Kremer, K.: Multiscale simulation of soft matter systems - from the atomistic to the coarse-grained level and back. *Soft Matter* **5**, 4357–4366 (2009)
34. Praprotnik, M., Delle Site, L., Kremer, K.: Adaptive resolution molecular-dynamics simulation: changing the degrees of freedom on the fly. *J. Chem. Phys.* **123**, 224106 (2005)
35. Reith, D., Pütz, M., Müller-Plathe, F.: Deriving effective mesoscale potentials from atomistic simulations. *J. Comput. Chem.* **24**, 1624–1636 (2003)
36. Rosenberger, D., Sanyal, T., Shell, M.S., van der Vegt, N.F.A.: Transferability of local density-assisted implicit solvation models for homogeneous fluid mixtures. *J. Chem. Theor. Comput.* **15**, 2881–2895 (2019)
37. Rühle, V., Junghans, C., Lukyanov, A., Kremer, K., Andrienko, D.: Versatile object-oriented toolkit for coarse-graining applications. *J. Chem. Theor. Comput.* **5**, 3211–3223 (2009)
38. Ruelle, D.: *Statistical Mechanics: Rigorous Results*. W.A. Benjamin Publ, New York (1969)
39. Ruelle, D.: Superstable interactions in classical statistical mechanics. *Commun. Math. Phys.* **18**, 127–159 (1970)
40. Schmid, F.: Editorial: Multiscale simulation methods for soft matter systems. *J. Phys.: Condens. Matter* **34**, 160401 (2022)
41. Shell, M.S.: The relative entropy is fundamental to multiscale and inverse thermodynamic problems. *J. Chem. Phys.* **129**, 144108 (2008)
42. Soper, A.K.: Empirical potential Monte Carlo simulation of fluid structure. *Chem. Phys.* **202**, 295–306 (1996)
43. Stell, G.: Cluster expansions for classical systems in equilibrium. In: Fritsch, H.L., Lebowitz, J.L. (eds.) *The Equilibrium Theory of Classical Fluids*, pp. II-171–II-266. W.A. Benjamin Publ., New York (1964)
44. Tóth, G.: Interactions from diffraction data: historical and comprehensive overview of simulation assisted methods. *J. Phys.: Condens. Matter* **19**, 335220 (2007)
45. Tsourtis, A., Harmandaris, V., Tsagkarogiannis, D.: Parameterization of coarse-grained molecular interactions through potential of mean force calculations and cluster expansion techniques. *Entropy* **19**, 395 (2017)
46. Wang, Y., Noid, W.G., Liu, P., Voth, G.A.: Effective force coarse-graining. *Phys. Chem. Chem. Phys.* **11**, 2002–2015 (2009)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.