Determining pair interactions from structural data: An inverse problem in statistical mechanics

Martin Hanke

Johannes Gutenberg-Universität
Institut für Mathematik

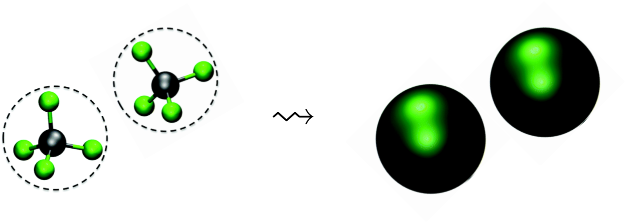
Berlin, Oct. 2018 ; MOANSI meeting

supported by TRR 146 (DFG)

The application

Atomistic numerical simulation techniques of complex molecules in material science require advanced multilevel techniques.

One such technique, called *coarse graining* (CG), replaces (sub)molecular structures by single *beads*:

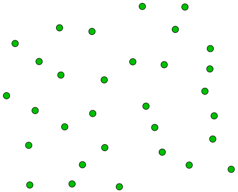


The simulation of the beads then requires the determination of *effective pair potentials* for the interaction of these beads.

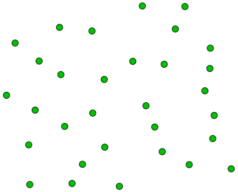
Outline

- Setting of the problem
- The Henderson problem
- Iterative solution methods
- Newton-type iterative schemes

Setting of the problem



Consider a huge ensemble of particles with (counting) density ρ_0 in thermodynamical equilibrium



whose potential energy (structural Hamiltonian) is determined by a pair potential

$$u : \mathbb{R}^+ \to \mathbb{R}$$

depending only on the distance of the interacting particles.

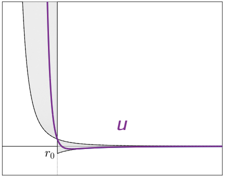
Standing assumptions

- the temperature T is sufficiently large and the (counting) density ρ_0 is sufficiently small
- the pair potential is of *Lennard-Jones type*, i.e.,
  - u decays fast enough as $r \to \infty$:

$$|u(r)| \leq Cr^{-\alpha}, \quad r \geq r_0, \qquad C > 0,\ \alpha > 3$$

  - u diverges fast enough to $+\infty$ as $r \to 0$:

$$u(r) > cr^{-\alpha}, \quad r \leq r_0, \qquad c > 0$$



$$u(r) = 4\epsilon\Big(\Big(\frac{\sigma}{r}\Big)^{12} - \Big(\frac{\sigma}{r}\Big)^{6}\Big)$$

Radial distribution function

The statistical distribution of the particles in the full space $\mathbb{R}^3$ (thermodynamical limit) is determined by the so-called *Gibbs' measure*.

It states that there exists a (translation and rotation invariant) pair-distribution function $\rho^{(2)}(x, y)$ and an associated *radial distribution function* (RDF)

$$g(r) = \frac{1}{\rho_0^2}\,\rho^{(2)}(x, x')\,, \qquad |x - x'| = r\,,$$

such that

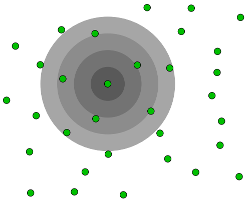
$$N_R = \int_{|x|<R} \rho^{(2)}(0, x)\,\mathrm{d}x = 4\pi\rho_0^2 \int_0^R g(r)\, r^2\,\mathrm{d}r$$

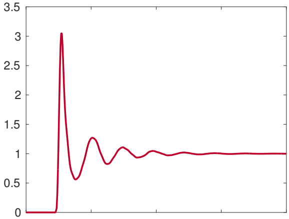
is the expected number of particles in a sphere of radius $R > 0$ around a given particle.

Since

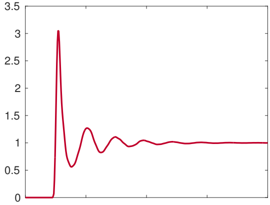
$$N_R = 4\pi\rho_0^2 \int_0^R g(r)\, r^2 \,\mathrm{d}r \quad \Leftrightarrow \quad g(r) = \frac{1}{\rho_0^2}\frac{1}{4\pi r^2}\frac{\mathrm{d}}{\mathrm{d}r} N_r\,,$$

the radial distribution function can be obtained from numerical simulations by counting particles on spherical shells:





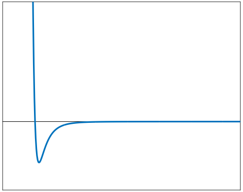
Radial distribution function



The RDF has the following properties:

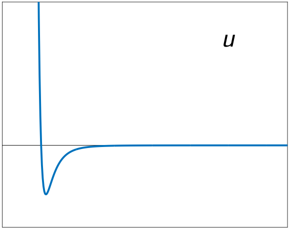
- $g(r) - 1 \in L^1(\mathbb{R}^+; r^2 \mathrm{d}r)$ RUELLE, 1969
- $g(r) - 1 \in L^\infty(\mathbb{R}^+; r^\alpha \mathrm{d}r)$ GROENEVELD, 1967; H., 2018
- $ce^{-u(r)} \leq g(r) \leq Ce^{-u(r)}$ H., 2018

The Henderson Problem

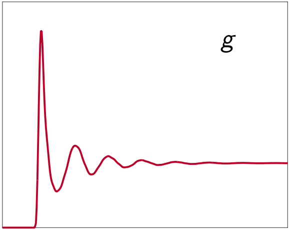


As we have seen, Lennard-Jones type pair potentials u yield a well-defined RDF g:

F :

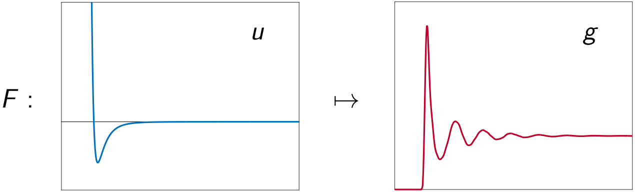


$\mapsto$



... the "*Henderson map*" F

The Henderson problem

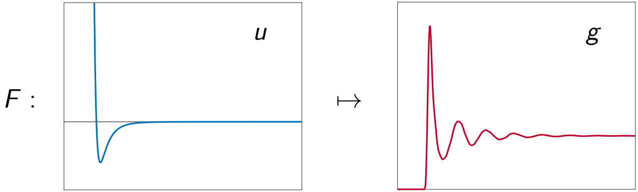


For the determination of effective potentials the *inverse problem*

- Given $g = F(u)$; determine u

is of interest

The Henderson problem



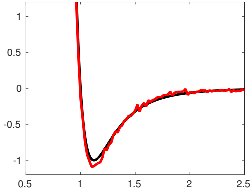
- Uniqueness: $\rightsquigarrow$ Henderson, 1974
  Frommer, H., 2018

- Existence: a *hard-core* potential solution is known to exist if

$$g(r) = 0, \quad 0 < r < r_1,$$
$$g(r) \approx 1, \qquad r > r_1$$

Koralov, 2007

Iterative solution methods



Inverse Boltzmann Iteration

To solve the inverse Henderson problem physical chemists often apply the *Inverse Boltzmann Iteration* (IBI),

$$u_{n+1} = u_n + \frac{1}{\beta} \log \frac{F(u_n)}{g}, \qquad n = 0, 1, 2, \ldots,$$

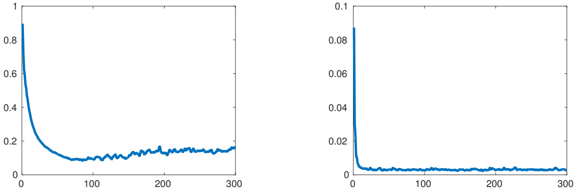
starting, e.g., with the "potential of mean force", $u_0 = -\frac{1}{\beta} \log g$.

Apparently:

- if $g = F(u^\dagger)$ ("*attainability*") then $u^\dagger$ is a fixed point of this iteration
- if g fails to be attainable (due to noise, for example) then the iteration must diverge

Semiconvergence

In practice this scheme is fairly robust, but exhibits (slight) *semiconvergence* due to noise:



error history   data fit in $L^1(\mathbb{R}^3)$

- here the error (unknown in practice!) is measured as

$$\|u_n - u^\dagger\|_g^2 := \int_0^\infty g(r)\big(u_n - u^\dagger(r)\big)^2 \, \mathrm{d}r$$

Well-posedness of IBI

$$u_{n+1} = \Phi(u_n) = u_n + \frac{1}{\beta} \log \frac{F(u_n)}{g}, \quad n = 0, 1, 2, \ldots$$

Qu: Will u_{n+1} be of Lennard-Jones type, if u_n is close to $u^\dagger$, i.e., does Φ map a neighborhood of $u^\dagger$ onto some (other) neighborhood of $u^\dagger$?

Ans: There are appropriate topologies such that the Henderson map and also Φ are locally differentiable. Accordingly, if $\|u - u^\dagger\|$ is small then $\Phi(u)$ will again be of Lennard-Jones type.

H., 2018

Convergence analysis (?)

$$u_{n+1} = u_n + \frac{1}{\beta} \log \frac{F(u_n)}{g}, \quad n = 0, 1, 2, \ldots$$

Error analysis (formal) for the attainable situation:

$$\begin{aligned} \sqrt{g}\,(u_{n+1} - u^\dagger) &= \sqrt{g}\,(u_n - u^\dagger) + \frac{1}{\beta}\sqrt{g}\, \log \frac{F(u_n)}{g} \\ &\approx \sqrt{g}\,(u_n - u^\dagger) + \frac{1}{\beta}\,\sqrt{g}\,\frac{g}{F(u^\dagger)}\,\frac{F'(u^\dagger)(u_n - u^\dagger)}{g} \end{aligned}$$

Convergence analysis (?)

$$u_{n+1} = u_n + \frac{1}{\beta} \log \frac{F(u_n)}{g}, \quad n = 0, 1, 2, \dots$$

Error analysis (formal) for the attainable situation:

$$\begin{aligned}
\sqrt{g}\,(u_{n+1} - u^\dagger) &= \sqrt{g}\,(u_n - u^\dagger) + \frac{1}{\beta}\sqrt{g}\, \log \frac{F(u_n)}{g} \\
&\approx \sqrt{g}\,(u_n - u^\dagger) + \frac{1}{\beta}\,\sqrt{g}\,\frac{g}{F(u^\dagger)}\,\frac{F'(u^\dagger)(u_n - u^\dagger)}{g} \\
&= \left(I + \frac{1}{\beta}\,\frac{1}{\sqrt{g}}\,F'(u^\dagger)\frac{1}{\sqrt{g}} \right) \sqrt{g}\,(u_n - u^\dagger)
\end{aligned}$$

Convergence analysis (?)

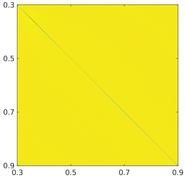
$$u_{n+1} = u_n + \frac{1}{\beta}\log\frac{F(u_n)}{g}, \quad n = 0,1,2,\dots$$

Error analysis (formal) for the attainable situation:

$$\begin{aligned}\sqrt{g}\,(u_{n+1} - u^\dagger) &= \sqrt{g}\,(u_n - u^\dagger) + \frac{1}{\beta}\sqrt{g}\,\log\frac{F(u_n)}{g}\\ &\approx \sqrt{g}\,(u_n - u^\dagger) + \frac{1}{\beta}\sqrt{g}\,\frac{g}{F(u^\dagger)}\,\frac{F'(u^\dagger)(u_n - u^\dagger)}{g}\\ &= \left(I + \frac{1}{\beta}\frac{1}{\sqrt{g}}F'(u^\dagger)\frac{1}{\sqrt{g}}\right)\sqrt{g}\,(u_n - u^\dagger)\end{aligned}$$

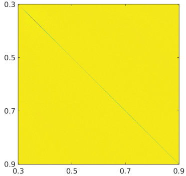
Attn: Note that $-F'$ is a positive (unbounded) operator in L^2.

Newton-type iterative schemes



Inverse Monte Carlo

The derivative $F'(u)$ of the Henderson map can be assembled from the joint 3- and 4-particle distributions of the ensemble.



The corresponding Newton scheme is known as *Inverse Monte Carlo*:

$$u_{k+1} = u_k + F'(u_k)^{-1}\big(g - F(u_k)\big)$$

Generalized Newton iteration

We propose a generalized Newton scheme, where the inverse of the Henderson map is approximated by the *hypernetted chain* integral equation

$$u \approx U(g) = -\frac{1}{\beta} \log g + \frac{1}{\beta}(h - c),$$

Here,

$$h = g - 1 \in L^\infty(\mathbb{R}^+; r^\alpha \mathrm{d}r),$$

and c is defined by the convolution integral†

$$c + \rho_0 h * c = h.$$

It can be shown that the convolution defines a Banach algebra in $L^\infty(\mathbb{R}^+; r^\alpha \mathrm{d}r)$, and hence $c \in L^\infty(\mathbb{R}^+; r^\alpha \mathrm{d}r)$ is well-defined provided that the *structure factor*

$$S(\omega) = 1 + \rho_0 \widehat{h}(\omega)$$

is positive (Wiener Lemma).

Generalized Newton iteration

$$u \approx U(g) = -\frac{1}{\beta} \log g + \frac{1}{\beta}(h - c),$$

It follows that

$$F'(u_k)^{-1} g' \approx U'(g) g' = -\frac{1}{\beta} \frac{g'}{g} + \frac{1}{\beta}(g' - c')$$

where $\varphi = g' - c'$ is given in Fourier space by

$$\widehat{\varphi} = \rho_0^2 \frac{2 + \rho_0 \widehat{h}}{\left(1 + \rho_0 \widehat{h}\right)^2} \widehat{h}\,\widehat{g'},$$

The corresponding *inverse hypernetted chain iteration* is defined as

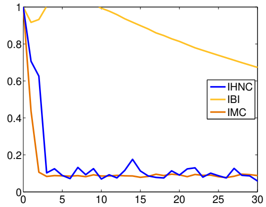
$$u_{k+1} = u_k + \frac{1}{\beta} \log \frac{g_k}{g} + \frac{\rho_0}{\beta} \varphi_k$$

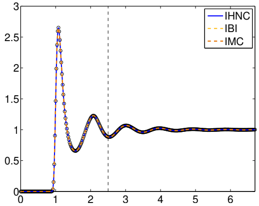
Lennard-Jones potential

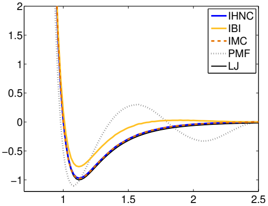
$$u = 4\varepsilon\big((\sigma/r)^{12} - (\sigma/r)^{6}\big)$$

near the "triple point" (phase transition)

error history:







Open problems

- Uniqueness of potential ✓
- Existence of potential ?
- Well-posedness of IBI ✓
- Convergence of IBI ?
- Stability/Regularization properties ?

References

F. Delbary, M. Hanke, and D. Ivanizki
A generalized Newton iteration for computing the solution of the inverse Henderson problem.
arXiv:1806.11135

F. Frommer, M. Hanke
A note on the uniqueness result for the inverse Henderson problem.
In preparation

M. Hanke
Fréchet differentiability of molecular distribution functions I. L^∞ analysis.
Lett. Math. Phys. **108** (2018), 285–306

M. Hanke
Fréchet differentiability of molecular distribution functions II. The Ursell function.
Lett. Math. Phys. **108** (2018), 307–329

D. Rosenberger, M. Hanke, N.F.A. van der Vegt
Comparison of iterative inverse coarse-graining methods.
Eur. Phys. J. Special Topics **225** (2016), 1323–1345.

M. Hanke
Well-posedness of the Iterative Boltzmann Iteration.
J. Stat. Phys. **170** (2018), 536–553