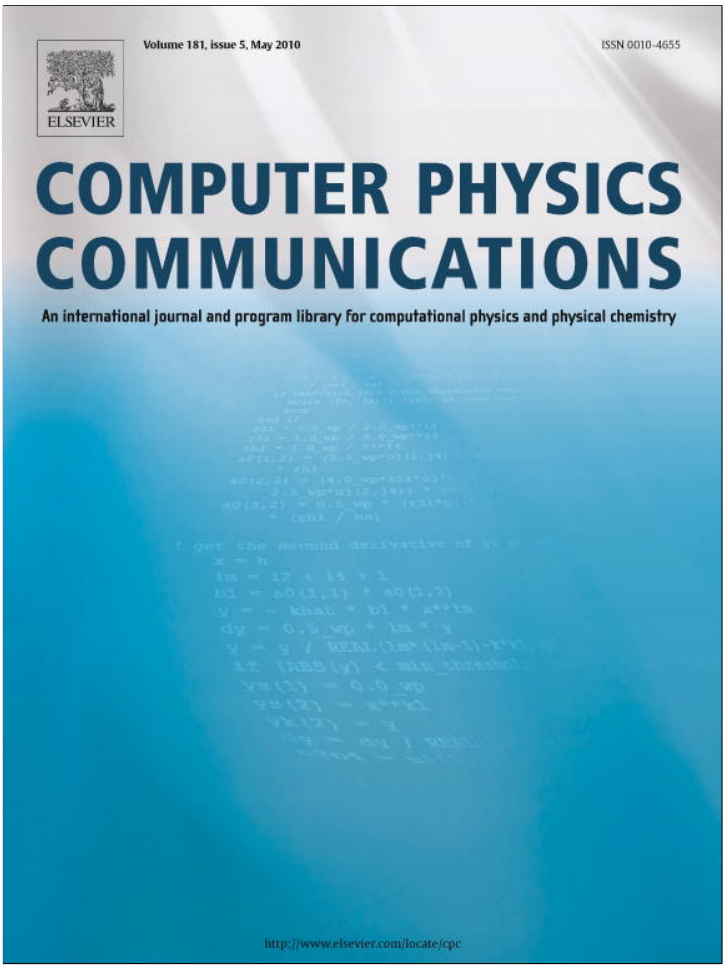


Computer Physics Communications 181 (2010) 887–905

Contents lists available at ScienceDirect

Computer Physics Communications

www.elsevier.com/locate/cpc

Assessment of numerical optimization algorithms for the development of molecular models

Marco Hülsmann [a], Jadran Vrabec [b], Astrid Maaß [a], Dirk Reith [a,*]

[a] Fraunhofer-Institute for Algorithms and Scientific Computing (SCAI), Schloß Birlinghoven, 53754 Sankt Augustin, Germany
[b] Lehrstuhl für Thermodynamik und Energietechnik (ThEt), Universität Paderborn, Warburger Str. 100, 33098 Paderborn, Germany

ARTICLE INFO

Article history:
Received 2 September 2009
Received in revised form 6 November 2009
Accepted 5 January 2010
Available online 7 January 2010

Keywords:
Molecular models
Numerical optimization
Gradient-based algorithms
Vapor–liquid equilibrium
Lennard-Jones potential

ABSTRACT

In the pursuit to study the parameterization problem of molecular models with a broad perspective, this paper is focused on an isolated aspect: It is investigated, by which algorithms parameters can be best optimized simultaneously to different types of target data (experimental or theoretical) over a range of temperatures with the lowest number of iteration steps. As an example, nitrogen is regarded, where the intermolecular interactions are well described by the quadrupolar two-center Lennard-Jones model that has four state-independent parameters. The target data comprise experimental values for saturated liquid density, enthalpy of vaporization, and vapor pressure. For the purpose of testing algorithms, molecular simulations are entirely replaced by fit functions of vapor–liquid equilibrium (VLE) properties from the literature to assess efficiently the diverse numerical optimization algorithms investigated, being state-of-the-art gradient-based methods with very good convergency qualities. Additionally, artificial noise was superimposed onto the VLE fit results to evaluate the numerical optimization algorithms so that the calculation of molecular simulation data was mimicked. Large differences in the behavior of the individual optimization algorithms are found and some are identified to be capable to handle noisy function values.

1. Introduction

Molecular modeling and simulation has made significant progress in recent years, being applied to a variety of areas such as thermodynamic properties of fluids [1–7], mechanic properties of solids [8–10], phase change phenomena [11–13], transport processes in biologic tissue [14,15], protein folding [16–18], transport processes in liquids [19–21], polymer properties by using different length scales [22–25] or generic statistic properties of soft matter [26].

The sound physical basis of the molecular approach allows for that versatility. Simulations in most of these areas are carried out in the following fashion: First, a molecular model has to be chosen, then it is assigned to a molecular configuration, and finally, the phase space is subsequently explored under specified boundary conditions, which allows gathering a broad variety of information. The exploration of the phase space is mostly a rather technical issue which may, however, be elaborate in the sense of efficient programing and computational time.

The versatility of such simulations as a method lies in the bottom-up approach: the "external" forces due to the boundary conditions, such as thermostats or velocity gradients, are usually much weaker than the intermolecular forces which govern the simulation result. Many molecular models have shown that they may adequately represent the intra- and intermolecular interactions even in a quantitative manner in numerous applications. Therefore, molecular modeling and simulation are located somewhere between theory and experiment [2,27,28].

As the central role is played by the intermolecular interactions, there is substantial research going on over decades to define and optimize molecular models [29–39]. Such models describe to a varying extent, among others, molecular geometry, repulsion, dispersion, electrostatics and polarizability. Although these are all physical properties, they have to be parameterized. Thereby, e.g., results from quantum mechanics may be useful [40]. However, usually not all parameters can be assigned unambiguously so that some optimization to experimental data needs to be done.

Many approaches exist to optimize molecular models and to create reliable force fields [41–45]. However, most of the approaches are based on quantum mechanics and the optimization of intramolecular model parameters and charges. The intermolecular model parameters predicted by this kind of methods are not very reliable and can only be used as an initial guess for a further optimization procedure. The development of empirical and semi-empirical molecular models are accompanied by iterative simulations with varying model parameter vectors. Thereby, different simulation results are compared to experimental target values. The most time-consuming step in this process is by far molecular simulation to evaluate the response of the molecular system to a pa-

* Corresponding author.
E-mail address: dirk.reith@scai.fraunhofer.de (D. Reith).

0010-4655/$ – see front matter
doi:10.1016/j.cpc.2010.01.001

rameter variation, which may take days or more. It is thus useful to discuss optimization schemes for molecular model parameters in a systematic manner.

There are different aspects concerning the usability of automatic optimization schemes for molecular modeling, e.g. the definition of the type and the number of model parameters to be optimized simultaneously or successively, the target values to be fitted, the choice of the algorithms and the discussion of their behavior with respect to statistical noise, as well as the technical implementation, which is presented and discussed in detail in another contribution [46].

The present work entirely focuses on the best choice of the optimization algorithms. They are assessed from a physical perspective, using a simple molecular model type, which is reasonable for numerous real fluids. The reason for this is the fact that the computation of the considered physical properties is very fast because simple fit functions are evaluated and hence, no simulations have to be performed in this contribution. The objective is to identify the best and most efficient algorithm, i.e. the one that requires the lowest number of iteration steps, to save substantial research and computation time. For the assessment procedure a trick can be used to avoid molecular simulations. By using fit functions instead, a large number of optimization algorithms can be tested easily and fast. In order to mimic simulation runs as realistically as possible, artificial noise is added to the calculated physical properties so that a detailed assessment also with respect to noise can be carried out.

Molecular models for fluids have been optimized to changing properties over time: early on, the pair correlation function from neutron scattering was used [47,48], later, the second virial coefficient [49] or the liquid density at ambient conditions [50] was used, nowadays, vapor–liquid equilibrium (VLE) data over the full temperature range [51–57] is accepted as a good choice for models of the fluid state.

In this work, optimizations to experimental data of saturated liquid density, enthalpy of vaporization, and vapor pressure as functions of the temperature are taken. As a case study nitrogen is regarded, being described by the quadrupolar two-center Lennard-Jones (2CLJQ) model. This model has four parameters, i.e. Lennard-Jones size and energy parameter, elongation and quadrupolar moment, which are all optimized simultaneously to experimental data. The 2CLJQ model for nitrogen has shown its appropriateness by different authors [58–61].

The present study is facilitated by the availability of temperature-dependent fits for VLE properties as functions of the considered four molecular parameters from prior work [52]. Please note again that no simulation runs had to be made here and a large number of mathematical optimization schemes from the literature was assessed with little computational effort.

2. Problem definition

The present work considers the optimization of the four state-independent parameters of the 2CLJQ model. It consists of two identical Lennard-Jones (LJ) sites, separated by an elongation L and a point quadrupole site at the center of mass with a moment Q, which is oriented along the molecular axis. The pair potential $u_{2\mathrm{CLJQ}}$ is given by

$$\begin{aligned} &u_{2\mathrm{CLJQ}}\left(\boldsymbol{r}_{ij}, \boldsymbol{\omega}_i, \boldsymbol{\omega}_j, L, Q^2\right) \\ &= \sum_{a=1}^{2}\sum_{b=1}^{2} 4\epsilon\left[\left(\frac{\sigma}{r_{ab}}\right)^{12} - \left(\frac{\sigma}{r_{ab}}\right)^{6}\right] \\ &\quad + \frac{3}{4}\frac{Q^2}{\|\boldsymbol{r}_{ij}\|^5}\left[1 - 5\left(\cos^2\theta_i + \cos^2\theta_j\right) - 15\cos^2\theta_i\cos^2\theta_j\right. \\ &\quad \left. + 2(\sin\theta_i\sin\theta_j\cos\phi_{ij} - 4\cos\theta_i\cos\theta_j)^2\right]. \end{aligned}$$

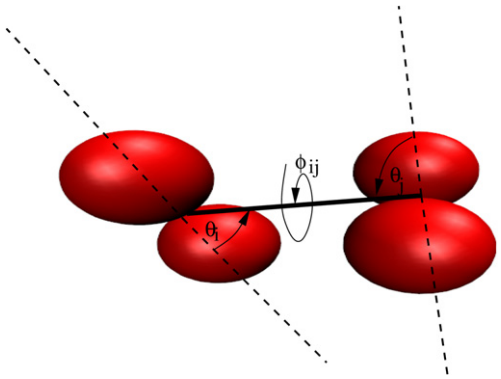


Fig. 1. Orientation angles used within the 2CLJQ potential. Note that $\boldsymbol{\omega}_i = (\phi_{ij}, \theta_i)$.

Therein, $\boldsymbol{r}_{ij}$ is the center–center distance vector of two molecules i and j, r_{ab} is one of the four LJ site–site distances, where a refers to the two LJ sites of molecule i and b to the two LJ sites of molecule j. Furthermore, $\boldsymbol{\omega}_i$ and $\boldsymbol{\omega}_j$ represent the orientations of the two molecules, where θ_i is the azimuthal angle between the axis of the molecule i and the center–center connection line, and ϕ_{ij} is the angle between the axes of molecules i and j. For more details, cf. Fig. 1 and [52].

2.1. Loss function

The following experimental VLE properties were considered for the optimization of the model parameters: the saturated liquid density ρ_l, the enthalpy of vaporization Δh_v and the vapor pressure p_σ. As the VLE properties are interrelated by the Clausius–Clapeyron equation, either ρ_l and Δh_v or ρ_l and p_σ were considered during the optimization. Note that by this equation, also Δh_v can be computed without executing any molecular simulation run.

The VLE properties mentioned above are temperature dependent so that the model is optimized at one or more temperatures and the parameters are adjusted to experimental data at those temperatures simultaneously.

In order to optimize with respect to different properties simultaneously, the following quadratic loss function between model and experimental data in a temperature range $\mathcal{T}$ was considered:

$$F(\boldsymbol{x}) = \sum_{i=1}^{n}\sum_{T\in\mathcal{T}} w_{i,T}\left(\frac{f_{i,T}^{\mathrm{exp}} - f_{i,T}^{\mathrm{sim}}(\boldsymbol{x})}{f_{i,T}^{\mathrm{exp}}}\right)^2, \tag{1}$$

where $\boldsymbol{x} = (\epsilon, \sigma, Q^2, L)^T \in \mathbb{R}^4$ is the molecular parameter vector, n is the number of physical properties involved at different temperatures $T \in \mathcal{T}$. Moreover, $f_{i,T}^{\mathrm{sim}}(\boldsymbol{x})$ is the ith property from the molecular model, depending on its parameters, at a certain temperature T and $f_{i,T}^{\mathrm{exp}}$ is the respective experimental target value. The usage of weights $w_{i,T}$ allows for the fact that the loss function includes the easier reproducibility of particular properties at particular temperatures. Please note that in this work $\forall_{i,T}\ w_{i,T} = 1$, since for the assessment of the optimization algorithms, all properties are considered as homologous. The reason for this is the fact that the first try consists in the simultaneous optimization of all properties independent from the amount of noise. Section 5 will show that this consideration led to successful results. As mentioned before, in this work, simulations are replaced by fit functions in order to obtain a well-balanced assessment. However, when simulations are used instead, the loss function should be weighted in a different way. Then, the resulting molecular model may not be optimal but the optimization task may be easier to solve.

This loss function has to be minimized with respect to $\boldsymbol{x}$. In an ideal situation, some $\boldsymbol{x}^{\mathrm{opt}}$ can be found for which $F(\boldsymbol{x}^{\mathrm{opt}}) = 0$. However, the primary goal is to find a local minimum in an

admissible compact parameter domain. Hence, an admissible range for each of the molecular model parameters has to be defined.

The fit functions [52] employed in this work guarantee a smooth dependency of F on the parameters. Hence, it is aimed to find a minimum at $\boldsymbol{x}^{\text{opt}}$ for which

$$\nabla F\left(\boldsymbol{x}^{\text{opt}}\right) = 0. \tag{2}$$

This goal can be achieved by gradient-based numerical optimization algorithms, which are enumerated in Section 4.3 and discussed in detail in [46]. The gradient can be calculated by the partial derivatives

$$\frac{\partial F}{\partial x_j}(\boldsymbol{x}) = -2\sum_{i=1}^{n}\sum_{T\in\mathcal{T}} w_{i,T}\,\frac{f_{i,T}^{\text{exp}} - f_{i,T}^{\text{sim}}(\boldsymbol{x})}{(f_{i,T}^{\text{exp}})^2}\,\frac{\partial f_{i,T}^{\text{sim}}}{\partial x_j}(\boldsymbol{x}), \tag{3}$$

where $j = 1,\ldots,N$ and N is the number of model parameters. For 2CLJQ models, $N = 4$, due to the fact that four parameters have to be optimized. Finally, the partial derivatives of the properties can be approximated numerically by

$$\frac{\partial f_{i,T}^{\text{sim}}}{\partial x_j}(\boldsymbol{x}) = \frac{f_{i,T}^{\text{sim}}(x_1,\ldots,x_j+h,\ldots,x_N) - f_{i,T}^{\text{sim}}(\boldsymbol{x})}{h}, \tag{4}$$

with $h > 0$.

Please note that also central differences or other finite differences could be used in order to achieve a higher accuracy. However, Eq. (4) is a reliable method to approximate gradients and was used here to avoid unnecessary function evaluations, i.e. simulations.

Furthermore, note that the loss function itself does not have any unit but its gradient has units on its axes, namely the reciprocal value of the units of the respective molecular model parameter. Differences in the magnitudes of the gradient components can lead to deformations of the loss function. This, in turn, can cause numerical problems in finding the minimum and the norm of the gradient is not defined anymore. This problem can be avoided by the division of each model parameter by its unit, or by a certain physically meaningful reference value, which leads to a physically better defined molecular model. In this work, the former solution is used, as the units are kJ/mol, nm, and $(\text{Dnm})^2$, which do not lead to any numerical problems. Defining physically meaningful reference values is not important for an assessment of the optimization algorithms. In real applications, the user always has to care for the choice of reasonable units.

2.2. Shape of loss function in 2D

The loss function (1) is a non-negative real-valued function

$$F : \mathbb{R}^4 \to \mathbb{R}_0^+,$$
$$\left(\epsilon, \sigma, Q^2, L\right)^T \mapsto F\left(\left(\epsilon, \sigma, Q^2, L\right)^T\right) \geqslant 0.$$

The summation of the squares of relative errors on the physical properties smoothens the loss function F, which facilitates the use of gradient-based numerical optimization algorithms.

Altogether, eight distinct optimization scenarios were considered:

1. ρ_l and Δh_v at one temperature,
2. ρ_l and Δh_v at one temperature with artificial noise,
3. ρ_l and Δh_v at six different temperatures,
4. ρ_l and Δh_v at six different temperatures with artificial noise,
5. ρ_l and p_σ at one temperature,
6. ρ_l and p_σ at one temperature with artificial noise,
7. ρ_l and p_σ at six different temperatures,
8. ρ_l and p_σ at six different temperatures with artificial noise.

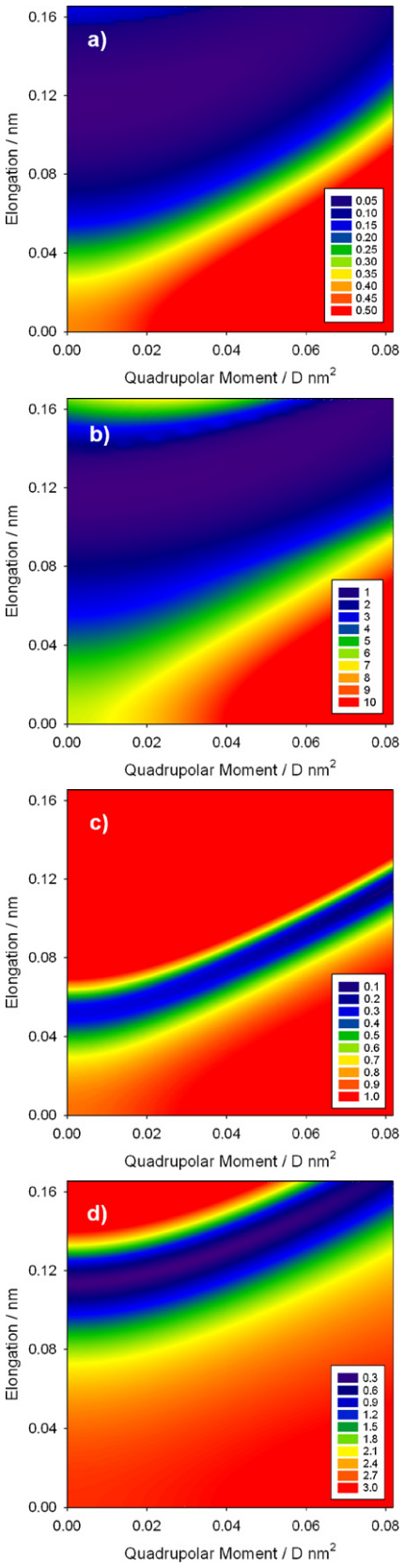


Fig. 2. Contour plots of the projected loss function, where $\epsilon = 0.3101$ kJ/mol and $\sigma = 0.331$ nm are fixed to their initial values, whereas the quadrupolar moment Q^2 and the elongation L vary. The plots a)–d) correspond to the optimization tasks 1, 3, 5, and 7, respectively. When six temperatures are considered, the loss function takes much higher values than in the case of one temperature only. A three-dimensional steep rain drain can be observed in all four plots, where the minimum may be located.

Table 1
Reduction formulas of different physical quantities.

Physical quantity	Reduction formula
Quadrupolar moment	$Q^{*2} = Q^2/(\epsilon\sigma^5)$
Elongation	$L^* = L/\sigma$
Temperature	$T^* = Tk_B/\epsilon$
Density	$\rho^* = \rho\sigma^3$
Vapor pressure	$p_\sigma^* = p_\sigma\sigma^3/\epsilon$
Enthalpy of vaporization	$\Delta h_v^* = \Delta h_v/\epsilon$

Fig. 2 shows the contour plots of the projected loss functions, where σ and ϵ are fixed to their initial values, cf. Section 4.1, and Q^2 and L vary. The plots a)–d) correspond to the optimization tasks 1, 3, 5, and 7, respectively. By these contours, a rough impression of the four-dimensional shape is given. However, these plots indicate the broad range of loss function values that may appear. Furthermore, it can be observed that there are domains where the function has the shape of a steep rain drain where the minimum may be situated.

3. Fit functions for VLE properties

Critical values of temperature and density, as well as the saturated liquid density, saturated vapor density, and vapor pressure are available as fit functions of the four molecular model parameters [52]. In the cited work, the VLE data was determined by molecular simulations for 30 individual 2CLJQ fluids. The calculated properties thereof were fitted by nonlinear regression functions, whose coefficients were determined individually for each property over a range of temperatures. In order to obtain VLE data for the whole range of $Q^2/(\epsilon\sigma^5)$, L/σ and Tk_B/ϵ, where k_B is the Boltzmann constant, the simulation data was globally fitted. In this regard, the critical data $T_c(\epsilon, \sigma, Q^2, L)$ and $\rho_c(\epsilon, \sigma, Q^2, L)$, the saturated liquid density $\rho_l(\epsilon, \sigma, Q^2, L, T)$, the saturated vapor density $\rho_v(\epsilon, \sigma, Q^2, L, T)$, as well as the vapor pressure $p_\sigma(\epsilon, \sigma, Q^2, L, T)$ were considered to be the key VLE data for an adjustment to real fluids.

In [52], the properties as well as Q^2, L, T, and T_c were presented in the reduced form, cf. Table 1. The reduced functions $T_c^*(Q^{*2}, L^*)$ and $\rho_c^*(Q^{*2}, L^*)$ were assumed to be linear combinations of elementary functions one of which is a constant c, the others depend either on Q^{*2}, i.e. $\psi_i(Q^{*2})$, or on L^*, i.e. $\xi_i(L^*)$, or on both, i.e. $\chi_i(Q^{*2}, L^*)$. The number of elementary functions was restricted to up to two for both the Q^{*2}- and the L^*-dependence. Writing the linear combination of any of the aforementioned functions represented by y, one gets

$$y(Q^{*2}, L^*) = c + \sum_{i=1}^{\leqslant 2} \alpha_i \cdot \psi_i(Q^{*2}) + \sum_{j=1}^{\leqslant 2} \beta_j \cdot \xi_j(L^*) + \sum_{k=1}^{\leqslant 4} \gamma_k \cdot \chi_k(Q^{*2}, L^*). \quad (5)$$

The crucial step was the selection of elementary functions ψ_i, ξ_i, and χ_i.

It is known that the density–temperature dependence near the critical point is well described by $\rho \sim (T_c - T)^{1/3}$, as given by Guggenheim [62]. In terms of reduced properties, ρ^* and T^*, the fit functions for the saturated liquid density ρ_l^* and vapor density ρ_v^* are given by

$$\rho_l^* = \rho_c^* + C_1 \cdot (T_c^* - T^*)^{1/3} + C_2' \cdot (T_c^* - T^*) + C_3' \cdot (T_c^* - T^*)^{3/2}, \quad (6)$$

$$\rho_v^* = \rho_c^* - C_1 \cdot (T_c^* - T^*)^{1/3} + C_2'' \cdot (T_c^* - T^*) + C_3'' \cdot (T_c^* - T^*)^{3/2}. \quad (7)$$

The simultaneous fitting of saturated liquid and saturated vapor densities yields not only the coefficients C_1, C_2', C_3', C_2'', and C_3'' but also the critical data ρ_c^*, T_c^*. The coefficient functions $C_1(Q^{*2}, L^*)$, $C_2'(Q^{*2}, L^*)$, $C_3'(Q^{*2}, L^*)$, $C_2''(Q^{*2}, L^*)$, and $C_3''(Q^{*2}, L^*)$ were also linear combinations of elementary functions in the sense of Eq. (5).

The fit function for the logarithm of the vapor pressure, $\ln p_\sigma^*$, reads

$$\ln p_\sigma^*(Q^{*2}, L^*, T^*) = c_1(Q^{*2}, L^*) + \frac{c_2(Q^{*2}, L^*)}{T^*} + \frac{c_3(Q^{*2}, L^*)}{T^{*4}}. \quad (8)$$

Again, the coefficients $c_1(Q^{*2}, L^*)$, $c_2(Q^{*2}, L^*)$, and $c_3(Q^{*2}, L^*)$ were assumed to be linear combinations of the elementary functions in the sense of Eq. (5).

The enthalpy of vaporization Δh_v^* was calculated in this work via the Clausius–Clapeyron equation

$$\frac{\partial \ln p_\sigma^*}{\partial T^*} = \frac{\Delta h_v^*}{p_\sigma^* T^* (1/\rho_v^* - 1/\rho_l^*)}, \quad (9)$$

where $(\partial \ln p_\sigma^*)/(\partial T^*)$ was determined analytically following Eq. (8). The values for ρ_l^* and ρ_v^* were taken from Eqs. (6) and (7), respectively. Please note that the vapor pressure fit (7) is only valid for $T/T_c \geqslant 0.7$. Hence, for lower temperatures, the ideal gas equation $\rho_v^* = p_\sigma^*/T^*$ was used for the reduced saturated vapor density.

Fig. 3 shows how the fit functions are used within the optimization procedure of this work. Their usage makes it possible to avoid expensive molecular simulation runs and therefore, they enable a detailed assessment of the optimization algorithms.

4. Optimizing the molecular model for nitrogen

For the detailed description of the general optimization workflow, cf. [46]. The considered numerical algorithms start with an initial guess $\boldsymbol{x}^0$, which is improved within the optimization procedure. The output of the fit functions, i.e. the reduced values for saturated liquid density plus enthalpy of vaporization or vapor pressure, are converted into values with physical units, and inserted into the loss function (1). Hence, the calculated physical properties are compared with the experimental ones [63]. If they fulfill a specified stopping criterion, the parameters are final and the workflow ends. Otherwise, the current parameter vector is passed on to the optimization algorithm, which finds a new parameter vector with a lower loss function value via one of the different investigated gradient-based methods.

4.1. Initial guess and boundary values

The values determined in [61] were taken as initial model parameters: $Q^2 = 0.020727$ (Dnm)^2 and $L = 0.10464$ nm. Furthermore, $\epsilon = 0.3101$ kJ/mol and $\sigma = 0.331$ nm were chosen, which differ from the optimal values found in [61] but are still physically reasonable values. The reason for this modification is the fact that not all four initial model parameters can be chosen as optimal parameters from literature, as the focus of this work lies on the assessment of the behavior of numerical optimization algorithms. This would be trivial, if very good or optimal values were taken as initial guess. The algorithms must also be successful, if bad model parameters are initially chosen. Of course, the underlying workflow described in [46] can be used for an unlimited number of model parameters. In this work, i.e. for the assessment of the

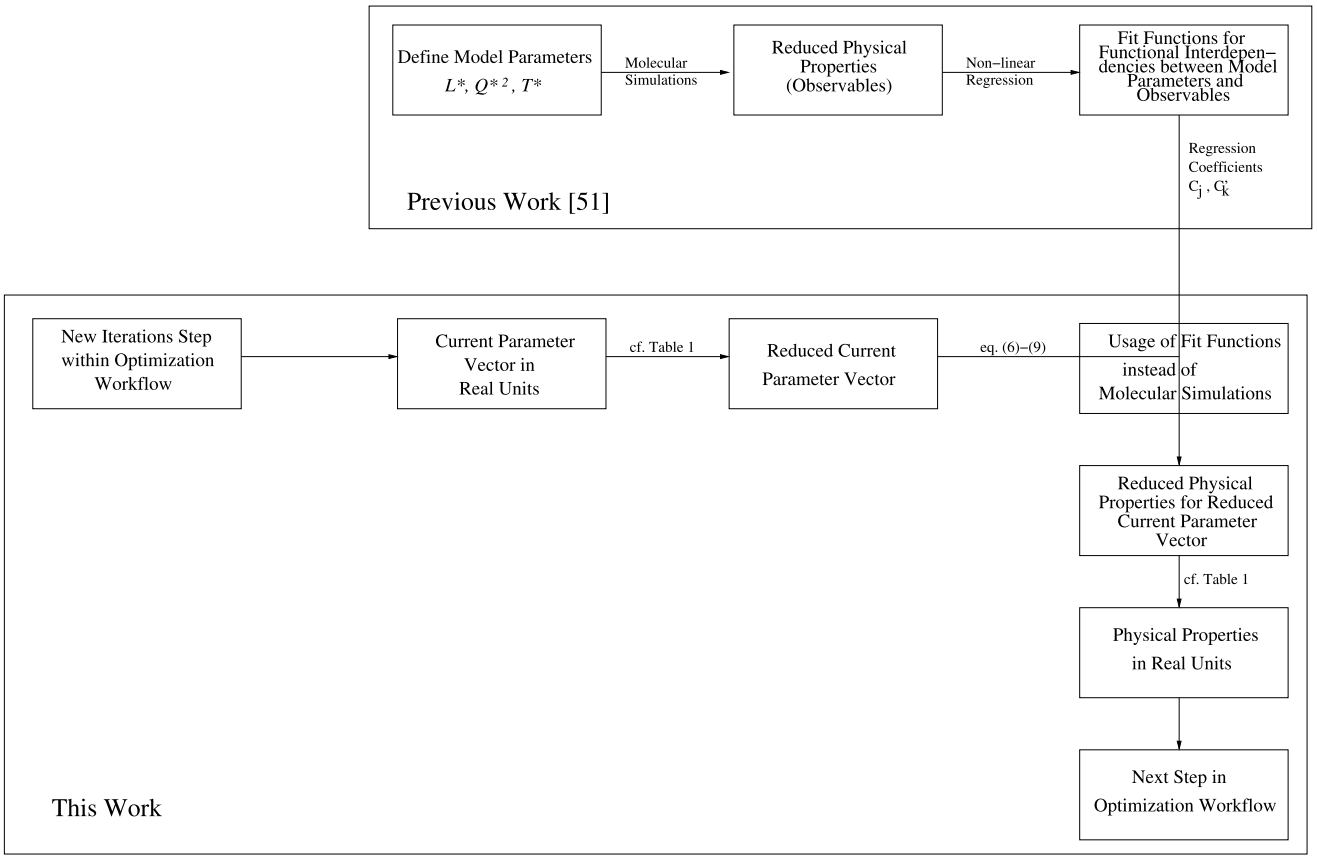


Fig. 3. Diagram describing how the fit functions for VLE properties developed by Stoll et al. [52] are used in this work in order to replace molecular simulations and hence to enable an assessment of the optimization algorithms.

algorithms, four parameters are sufficient to obtain reliable conclusions.

The optimization procedure has to be performed within an admissible domain: The range of validity of the fits [52] should be repeated here: ϵ and σ can be arbitrary, but they must be positive. However, in order to keep physically reasonable values, σ was not changed by more than 10% and ϵ not by more than 40%. The elongation is limited to a fraction of the LJ size parameter, i.e. $L/\sigma \in [0, 0.8]$, and at the same time, for the quadrupolar moment, $Q^2/(\epsilon\sigma^5) \in [0, 4]$ is required. The smaller σ and ϵ are, the lower are the maximal values for Q^2 and L. Therefore, if the initial parameters are taken as indicated above, the ranges are $0.2979 \leqslant \sigma/\mathrm{nm} \leqslant 0.3641$, $0.18606 \leqslant \epsilon/(\mathrm{kJ/mol}) \leqslant 0.43414$, $0 \leqslant (Q/\mathrm{Dnm})^2 \leqslant 0.029$ and $0 \leqslant L/\mathrm{nm} \leqslant 0.14895$.

Naturally, the choice of a different initial parameter vector may lead to a different local optimum. A study of the behavior of the algorithms with respect to different initial parameter vectors was not performed here for reasons of brevity, as its result is clear from the outset. Using molecular simulations, the initial model parameters as well as the admissible domain have always to be chosen physically reasonable. The initial model parameters may, for example, be taken from literature [41,43,44].

4.2. Different temperatures and artificial noise

The employed fit functions (6) and (8) are avlid in the temperature range $T/T_c \in [0.55, 0.95]$, where T_c is the critical temperature, which of course depends on the four molecular parameters. Note that [52] also provides a fit for T_c. The six temperatures considered here are $T/T_c \in \{0.52, 0.59, 0.67, 0.75, 0.83, 0.91\}$, corresponding to $T/\mathrm{K} \in \{65, 75, 85, 95, 105, 115\}$. If the target data were fitted at one temperature only, $T/T_c = 0.59$ was taken, which corresponds to $T/\mathrm{K} = 75$. Please note that $T/T_c = 0.52$ is slightly outside the temperature range where the fit functions are valid. However, the triple point temperature of nitrogen is $T = 63.151$ K and for reasons of continuity, the fit functions can handle reduced temperatures which are somewhat lower than 0.55.

In order to mimic the calculation of molecular simulation data, statistical noise was superimposed artificially on the physical properties from the fit functions. The artificial noise consists of uniformly distributed random numbers and is 0.5% for ρ_l, 1.0% for Δh_v and 3.0% for p_σ. Please note that thereby different loss function values may appear, if the loss function is evaluated for the same parameter vector. The values of physical properties obtained in independent molecular simulations are usually normally distributed but in this work, i.e. for the assessment of the behavior of the numerical algorithms with respect to noise, the assumption of uniformly distributed errors is more reliable because this is a much serious type of noise.

Moreover, the stopping criterion depends on what one can expect from the molecular model and the optimization workflow: Optimizing the model parameters to physical properties at different temperatures with a four-dimensional parameter vector is much more elaborate than fitting them at one temperature only. The problem becomes even more difficult in the case of artificial noise. For details, cf. [46].

For the assessment of the different numerical optimization algorithms, here $\forall_i\ w_i = 1$ was specified in Eq. (1), because all considered properties were treated equally.

4.3. Numerical optimization algorithms

The focus of the present work lies on applying fast algorithms with low complexity, i.e. efficient algorithms, because they must be

applicable to time-consuming molecular simulations as well. There is a large number of candidate methods in the field of gradient-based optimization. However, in general, the number of function evaluations, i.e. the number of simulations to be performed at each iteration, increases with the speed of convergency. Hence, the following numerical algorithms, which have very good convergency properties but do not require too much computation time, were chosen here:

1. **Descent methods**: $\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + t_k \boldsymbol{d}^k$
   - Steepest Descent: $\boldsymbol{d}^k = -\nabla F(\boldsymbol{x}^k)$
   - Newton Raphson: $\boldsymbol{d}^k = -(D^2 F(\boldsymbol{x}^k))^{-1} \nabla F(\boldsymbol{x}^k)$
   - Quasi Newton: $\boldsymbol{d}^k = -H_k^{-1} \nabla F(\boldsymbol{x}^k)$. The approximation H_k of the Hessian is updated at every iteration, which can be achieved by different Quasi Newton methods.
     - Powell Symmetric Broyden (PSB)
     - Davidon Fletcher Powell (DFP)
     - Broyden Fletcher Goldfarb Shanno (BFGS)

   The step length t_k is obtained by a function adaptive step length control method, the so-called *Armijo* step length control.
2. **Conjugate Gradient (CG) methods**: $\boldsymbol{d}^{k+1} = -\nabla F(\boldsymbol{x}^{k+1}) + \beta_k \boldsymbol{d}^k$, $\boldsymbol{d}^0 = -\nabla F(\boldsymbol{x}^0)$
   - Fletcher Reeves: $\beta_k^{\mathrm{FR}} = \|\nabla F(\boldsymbol{x}^{k+1})\|^2 / \|\nabla F(\boldsymbol{x}^k)\|^2$
   - Polak Ribière: $\beta_k^{\mathrm{PR}} = \langle \nabla F(\boldsymbol{x}^{k+1}) - \nabla F(\boldsymbol{x}^k), \nabla F(\boldsymbol{x}^{k+1}) \rangle / \|\nabla F(\boldsymbol{x}^k)\|^2$
3. **Trust Region (TR) methods**: $\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \boldsymbol{d}^k$ with the so-called *Trust Region subproblem*: $d^k = d^k(\Delta)$ for some pre-defined step length Δ.

   There are two ways to solve the subproblem:
   - Double Dog Leg Algorithm (DD): geometric approach
   - Exact solution: eigenvalue decomposition of Hessian

The algorithms and the Armijo step length control are described and discussed in detail in [46,66].

4.4. Stopping criteria

The stopping criterion $\|\nabla F(\boldsymbol{x})\| < \tau$ for some $\tau > 0$ cannot be expected to be always fulfilled because of the slow convergency of the Armijo step length control algorithm in a neighborhood of the minimum. For example, this can be due to the rain drain phenomenon mentioned in Section 2.2. Hence, the amount of computation time to fulfill this stopping criterion would be too high in comparison to the benefit obtained by the resulting step lengths, which are very small due to the Armijo constraint [66].

Therefore, the stopping criterion

$$F(\boldsymbol{x}) < \tau$$

was chosen for $\tau = 10^{-4}$ or even $\tau = 10^{-5}$. For the exact choice of τ, the Steepest Descent method was taken as a reference method: Whenever the Armijo step length control did not converge within a reasonable number of steps (usually 100), the actual parameter vector was taken as the optimal one and the optimization workflow ended. Then, an accurate upper bound of the actual loss function value was chosen for τ. The value of τ depended on the optimization task.

An algorithm was considered as successful, if and only if the stopping criterion was fulfilled with a reasonable amount of computation time. This means that the optimization workflow was interrupted, whenever the Armijo step length control did not converge within 100 steps at some iteration.

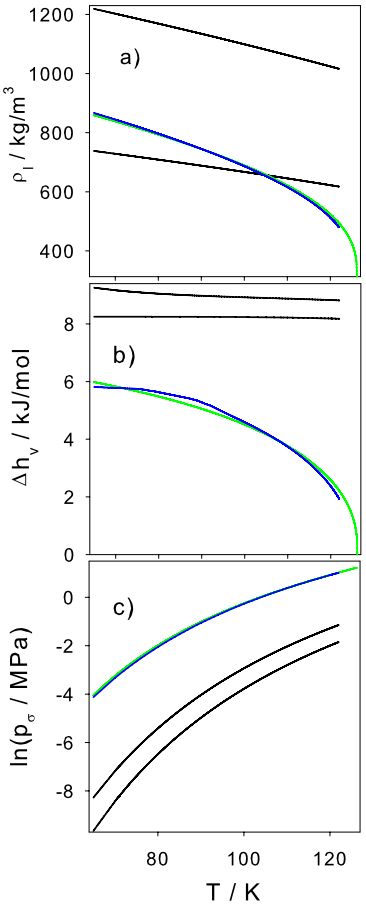


Fig. 4. Curves showing the considered physical properties over the temperature range from 65 K to 122 K using boundary values (black curves) and the optimal parameters from the corresponding optimization tasks (blue curves). The green curves indicate the experimental values for comparison, ranging from 65 K to 126 K. Plot a) shows saturated liquid density vs. temperature, plot b) enthalpy of vaporization vs. temperature, and plot c) the logarithm of vapor pressure vs. temperature. The plots indicate that boundary values can lead to extremely wrong results. This can also be true using model parameters that significantly differ from the minimum of the loss function. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5. Results and discussion

5.1. Optimal model parameters

As mentioned in Section 4, the loss function is defined within an admissible domain where it has to be minimized. From a physical point of view, it would be interesting to investigate the behavior of the physical properties over a range of temperatures using boundary values of the admissible domain as model parameters instead of the optimal ones. Fig. 4 presents ρ_l vs. T (plot a)), Δh_v vs. T (plot b)), and $\ln p_\sigma$ vs. T (plot c)). A temperature range of [65 K, 122 K] was chosen.

Each plot consists of four curves: A green curve showing the experimental target data, a blue one obtained by using optimal parameters and two black curves obtained by using boundary values of the admissible domain, as discussed below.

The experimental values over the complete temperature range (green curves) were taken from [67]. The values for the optimal parameters (blue curves) were taken from the results of the corresponding optimization tasks defined in Section 2.2:

- ρ_l vs. T: Optimization task 3.
- Δh_v vs. T: Optimization task 3.
- $\ln p_\sigma$ vs. T: Optimization task 7.

The first black curve (boundary values 1) is drawn by inserting initial values for ϵ and σ and minimal values for Q^2 and L and the second one (boundary values 2) by inserting maximal values for ϵ and σ and initial values for Q^2 and L into the corresponding fit functions.

The curves obtained by the optimal parameters are always very close to the experimental curves. However, using boundary values leads to extremely diverging results. Therefore, an automated optimization is indispensable.

5.2. Assessment of numerical optimization algorithms

The detailed results of the investigated numerical optimization algorithms are shown in the tables of Appendix A, where they are described and explained as well. Please note that all successful methods led to model parameters situated inside the admissible domain and not at its boundary.

In order to discuss and evaluate the algorithms, both the number of iterations, indicating the speed of convergency, and the number of function evaluations, indicating the number of loss function or partial derivatives evaluations, have to be considered. The number of function evaluations equals the number of molecular simulations which would have to be performed. The Newton Raphson and the Trust Region methods use a Hessian matrix at each step, which requires $\mathcal{O}(N^2)$ calculations of properties, where N is the number of the simultaneously optimized molecular model parameters, in addition to the calculation of the gradient that requires $\mathcal{O}(N)$ calculations of properties. Furthermore, at each Armijo iteration, one loss function evaluation has to be performed. Please note that for the current objective, the number of iterations is crucial, since the first goal is to assess the algorithms mathematically and the speed of convergency is given by the number of iterations. Whenever the methods are applied to molecular simulations, one must always care about the fact that the computation time required for the function evaluations should be as small as possible. This can be achieved, e.g. by parallelizing the calculation of the independent gradient components or Hessian entries. Furthermore, mathematical tricks can be applied in order to increase the efficiency of the optimization workflow, if the amount of noise is sufficiently low.

The evaluation criteria of the eight loss functions to be optimized, corresponding to the eight optimization tasks, always have to consider what can be expected from the optimization procedure. If target data at only one temperature is fitted, the computed properties can be expected to deviate only very little from the experimental ones. However, as the problem is underdetermined, the number of global optima of the loss function can be infinite. The proper definition of the initial guess and the admissible search domain ensure that a physically meaningful minimum is found.

If noise is added artificially, the properties can only be situated within an error bar around the minimum. Hence, a deviation smaller than the statistical uncertainty of the respective property indicated in Section 4.2 is an optimal result. However, the results could have been produced by chance. If an optimization is repeated, it may create a better or worse output. Therefore, some replications were performed, in order to get statistically representative assessments. The respective result Tables 4, 6, 8, and 10 contain the average values and Figs. 5–8, cf. Appendix B, show the corresponding box plots for each column of the result tables produced by ten replicates of each optimization workflow.

If more than one temperature is considered, the problem becomes more elaborate, as one parameter vector cannot predict exactly all properties at all desired temperatures at the same time: From the mathematical point of view, the problem is overdetermined. Together with noise, this is even more difficult.

The simplex algorithm is the only iterative method applied to the optimization of LJ parameters so far [31]. It is very robust with respect to noise in finding a local optimum but it is a quite heuristic approach with a slow convergency. Furthermore, at some point, it starts to hop around the minimum. In a practical situation, using Molecular Dynamics simulations executed with the simulation tool *YASP* [64], it took about 70–100 iterations for the optimization of the saturated liquid density and enthalpy of vaporization at one temperature with four model parameters and comparable accuracy [65]. Therefore, a reduction to 30–50 iterations would already be a benefit. If the optimization ends after ten iterations or even less, this can even be considered as an achievement and a very good performance of the respective algorithm. But one must be careful, as the simplex algorithm only requires $\mathcal{O}(P)$ loss function evaluations, where P is the number of iterations. The methods used here require more function evaluations per iteration. As the evaluation was made via the fit functions introduced in Section 3 and not via molecular simulations, the computational effort was irrelevant in the present work. If, however, molecular simulations are used in order to calculate the physical properties, the complexity by applying additional auxiliary methods should be reduced from $\mathcal{O}(PN^2)$ to $\mathcal{O}(PN)$ or even to $\mathcal{O}(P)$.

5.3. Unsuitable methods

We claim that the Quasi Newton methods are unsuitable for the present type of optimization tasks. This is especially true, if the function values are noisy. In most cases, the Armijo step length control does not converge within 100 steps after a few iterations only. This is due to the fact that an approximation H_k of the Hessian matrix is used and it is not guaranteed that the Quasi Newton direction is a descent direction, as H_k is not guaranteed to be symmetric positive definite (spd). For details, cf. [46]. Therefore, the methods tend to lead to the boundary of the admissible domain, where no improved loss function value can be found. Firstly, the gradient is only an inaccurate approximation in the case of noise, and secondly, H_k is another approximation, namely of the actual Hessian. Of course, this can lead to the fact that the Quasi Newton direction can lead anywhere but not to the minimum. In general, the Quasi Newton methods deliver poor results. However, in the case of noise, the probability to find a lower loss function value, in the sense of the Armijo step length, becomes higher. Thus, the Quasi Newton methods sometimes deliver quite good results but they are not robust with respect to noise.

Also, the solution of the Trust Region subproblem by the DD method is unsuitable: The Hessian matrix is not spd in most cases so that the so-called *Cauchy point* [66] is taken as a solution. But then, a good quality of the interpolation model [46,66] can only be achieved by chance. Hence, only in the case of noise, the DD was successful. But it cannot be guaranteed that this is always the case.

5.4. Detailed discussion of the results

Table 2 contains the results of the initial model parameters of each optimization task. When the saturated liquid density and the vapor pressure were optimized at one temperature (optimization tasks 5 and 6), the initial parameter vector was obtained by three steepest descent iterations with a heuristic step length [46], which led to the boundary of the admissible domain. Therefore, the new admissible domain for ϵ was reduced but not for σ, as this led to convergency problems of the Armijo step length control. Hence, the initial parameter vector is not the same as the original one in this case.

The results for the initial parameter vectors are poor, and the optimization algorithms could improve them drastically: In the following, the tables in Appendix A are discussed in detail:

Table 2
Initial model parameters and the corresponding results for all optimization tasks. For the first four, the relative errors on enthalpy of vaporization and for the last four, the ones on vapor pressure are indicated. If six temperatures are considered, the absolute average values over all temperatures are indicated. In the case of artificial noise, the averages over ten replicates are shown. Please note that different initial parameters were taken for the optimization tasks 5 and 6, as no minimum was found in the original admissible domain. Hence, it had to be modified.

Results for initial model parameters					
	$\boldsymbol{x}^0$	ρ_l	Δh_v, p_σ	$F(\boldsymbol{x}^0)$	$\|\nabla F(\boldsymbol{x}^0)\|$
Optimization task 1 (Table 3): ρ_l, Δh_v, one temperature, without noise	0.310100 0.331000 0.020727 0.104640	3.49%	10.88%	0.0131	1.84
Optimization task 2 (Table 4): ρ_l, Δh_v, one temperature, with noise	0.310100 0.331000 0.020727 0.104640	3.42%	10.97%	0.0132	389.13
Optimization task 3 (Table 5): ρ_l, Δh_v, six temperatures, without noise	0.310100 0.331000 0.020727 0.104640	5.49%	21.86%	0.3848	57.84
Optimization task 4 (Table 6): ρ_l, Δh_v, six temperatures, with noise	0.310100 0.331000 0.020727 0.104640	5.51%	21.86%	0.3850	2149.57
Optimization task 5 (Table 7): ρ_l, Δp_σ, one temperature, without noise	0.323221 0.338112 0.028995 0.088137	8.21%	68.87%	0.4811	131.96
Optimization task 6 (Table 8): ρ_l, Δp_σ, one temperature, with noise	0.323221 0.338112 0.028995 0.088137	8.30%	68.92%	0.4819	3794.30
Optimization task 7 (Table 9): ρ_l, Δp_σ, six temperatures, without noise	0.310100 0.331000 0.020727 0.104640	5.49%	34.60%	0.7524	121.44
Optimization task 8 (Table 10): ρ_l, Δp_σ, six temperatures, with noise	0.310100 0.331000 0.020727 0.104640	5.46%	34.56%	0.7510	1939.62

Table 3 contains the results obtained by optimizing the saturated liquid density and enthalpy of vaporization at $T = 75$ K without noise. The Steepest Descent and the Newton Raphson iterations are identical in this case, as the Newton direction was never accepted. The relative errors on ρ_l and Δh_v are very small (-0.03% on ρ_l and 0.19% on Δh_v). The Fletcher Reeves and the Trust Region methods led to even better results. The Polak Ribière and the Trust Region methods had the smallest number of iterations. As $\|\nabla F\| \approx 0$ for all these methods, their results can be considered as optimal.

In Table 4, the results of the same problem together with artificial noise are indicated: None of the methods fell below both error bars (0.5% on ρ_l and 1% on Δh_v). Only the Newton Raphson and the Fletcher Reeves methods reached the error bar of ρ_l. The PSB method gives the best result for Δh_v on average. The corresponding box plots (Fig. 5) show that the Newton Raphson, the Conjugate Gradient methods and, in this case, both Trust Region methods are the most robust ones with respect to noise. The Newton Raphson and the Fletcher Reeves methods deliver the best results, which can be seen from the position of the box plots, but only in the case of the exact Trust Region method, all ten replicates were successful.

Tables 5 and 6 show the results for ρ_l and Δh_v at six different temperatures, without and with artificial noise, respectively: As mentioned earlier, the loss function cannot be expected to reach zero, as the problem is overdetermined. All suitable methods, i.e. all methods except the Quasi Newton methods and the Trust Region method together with the DD algorithm, yield satisfactory results concerning the saturated liquid density (≈0.7% without and ≈1.7–2% with noise). Also for the enthalpy of vaporization, the average errors (≈3–3.5%) are satisfactory. Without noise, the Steepest Descent, the Newton Raphson, and the Fletcher Reeves methods are quite poor, as they require more than 50 iterations, whereas the Trust Region method is clearly superior with only 30 iterations. However, when noise is introduced, the Newton Raphson method yields the best results (0.63% on ρ_l). All the other methods—also the DFP and the DD methods—lead to similar satisfactory results. An interesting aspect is the fact that the noise only influences the error on ρ_l but not on Δh_v. From Fig. 6, it can be seen that the DFP method yields low relative errors on average but it is not robust with respect to noise. The most robust algorithms are the Steepest Descent method, the Conjugate Gradient methods and the exact Trust Region method. The low error on ρ_l in the case of the Newton Raphson method was maintained over ten replicates, as can be seen from the corresponding box plot.

The output of the optimization of ρ_l together with the vapor pressure p_σ at one temperature without and with noise is indicated in Tables 7 and 8, respectively: All suitable methods yield a notably small error for the vapor pressure (<0.5%), if no artificial noise is present. Otherwise, the error is always less than 1% but as can be seen from Fig. 7, the variations on this average error can be quite large. The best results for the vapor pressure in the case of no noise are created by the Newton Raphson method (0.03%), which requires more than 100 iterations, as well as the

Steepest Descent method. However, after the introduction of noise, no Newton Raphson replicate was successful. One workflow, which had been executed previously, gave an outlier result: After 23 iterations, the stopping criterion was fulfilled with an error on ρ_l of 6.86% and on p_σ of 0.04%. For the vapor pressure, this is a very good result but the probability that such a result is achieved is very low. In the case of the other methods, the error on the saturated liquid density is quite high (>0.9% without and even >6% with noise). Two exceptions are the Trust Region method in the case of no noise (0.68%) and the PSB method in the case of noise (4.34%). However, the PSB result was obtained by chance, because only two replicates were successful. The box plots in Fig. 7 show that the most robust methods are the Steepest Descent method, the Conjugate Gradient methods, and the Trust Region method. They also give similar results. It is worth to mention that in this case, also the DFP method works quite well. However, only three DFP replicates were successful.

When ρ_l and p_σ are optimized at six different temperatures simultaneously without noise, all suitable methods yield similar satisfactory results as shown in Table 9: The error on ρ_l is about 1.2–1.3% and the error on p_σ is about 3.6–3.8%. In the case of noise, which is summarized by Table 10, the error on ρ_l is slightly lower, and the error on Δh_v remains approximately the same. The Newton Raphson method yielded a quite high error on ρ_l ($\approx 4\%$) and only one replicate was successful. Interestingly, the DFP and BFGS methods produced very good results in this case. The stopping criterion was fulfilled by 7 and 5 replicates, respectively, within about 8–9 iterations. The lowest loss function value was given again by the Trust Region method with exact solution. However, exceptionally, the Trust Region method also delivered very good results, if the subproblem was solved by the DD algorithm. The box plots shown in Fig. 8 approve that all methods except the Newton Raphson and PSB methods achieve good results. Only the Trust Region method together with the DD algorithm—exceptionally being superior in this case—delivers narrow box plots. Only the relative error on ρ_l is produced robustly. All other methods give comparable results but it is worth to mention that the Conjugate Gradient methods are quite robust with respect to the noise on the loss function value itself and that the Trust Region method delivers the lowest loss function values.

5.5. Particular aspects with respect to noise

In general, the number of iterations is much smaller in the case of noise. However, the Armijo step length often requires more iterations, as the artificial noise may lead to greater loss function values. Hence, the probability to find a lower loss function value in the sense of the Armijo step length decreases. But at some point, the superimposed random number will be negative, which leads to much smaller loss function values. Hence, the algorithm converges faster to the minimum and may reach final parameters with lower or higher loss function values, depending on the sign of the last random number.

The two whiskers in each box plot of Figs. 5–8 show the best and the worst cases, i.e. the replicates with the highest and lowest loss function values, respectively. The worst case occurs, of course, if a replicate is not successful. However, also below the upper bound of the stopping criterion, better and worse replicates have to be differentiated. If the whiskers are situated far away from each other, there is a large span between the best and the worst cases. This is mostly the case for the Newton Raphson method and the Quasi Newton methods. This means that by chance, an extraordinarily good result can be achieved but also the opposite may occur as well. In the case of the Quasi Newton methods, the probability that a replicate is successful is very low. So, the occurrence of the best case is rare. If the distance between the lower whisker and the lower quartile is large, a very good result may be achieved but this case is most unlikely. The outliers marked by circles cannot be evaluated as best or worst cases. Only the whiskers can be chosen for this.

5.6. Further improvements

Another aspect is the fact that the tables only allow evaluations with respect to the stopping criterion, i.e. how 'optimal' the last iteration is, meaning how small the last loss function value is. All algorithms terminate, as soon as $F(\boldsymbol{x}) \leqslant \tau$ for some $\tau > 0$, and τ was defined with the Steepest Descent method as a reference. Therefore, all suitable methods will end up with approximately the same function values. But possibly there are methods which still get closer to the minimum. Hence, for the Conjugate Gradient methods and the Trust Region method with exact solution of the subproblem, a further investigation was performed, determining how close to the minimum the methods really get, until the Armijo step length does not converge within 100 steps or the Trust Region radius becomes too small.

For example, in the case of the enthalpy of vaporization at one temperature without noise, the Fletcher Reeves method gave a better result compared to Table 3: A relative error on ρ_l of 0.04% and on Δh_v of −0.06% was obtained after three additional iterations.

As can be seen from Table 4, both error bars were reached by none of the methods within the corresponding optimization task, which seems to be quite hard to solve for the considered algorithms. However, there was one exception obtained by further studies: After eight iterations, the Trust Region method with an exact solution of the Trust Region subproblem obtained 0.28% on ρ_l and −0.23% on Δh_v.

In the case of six temperatures (compared to Table 5), only slightly better results were obtained with much more effort: the Polak Ribière method yields a result of 0.61%/3.09% after 62 additional iterations, and the Trust Region method yielded a result of 0.62%/3.12% after 70 additional iterations. In the case of noise (compared to Table 6), only the Fletcher Reeves method showed improved results: After 16 iterations, the result was 0.79%/2.83%.

Improvements were achieved in the case of the vapor pressure as well: At $T = 75$ K without noise (compared to Table 7), the Polak Ribière method gave an error rate of 0.51% on ρ_l and of 0.09% on p_σ after only four additional iterations, and the result of the Trust Region method was even 0.24% and 0.03%, respectively, after 11 additional iterations. When noise was introduced (compared to Table 8), the Fletcher Reeves method achieved the error rates 2.81%/−0.14% after 18 iterations and the Polak Ribière method even 0.45%/0.25% after 25 iterations. This result cannot be improved, as it is situated within the error bars due to noise. When six temperatures are considered (compared to Table 9), the average error is 0.88% on ρ_l and 3.53% on p_σ in the case of Fletcher Reeves after 91 more iterations, the Polak Ribière method gave a result of 0.96% and 3.50%, respectively, after 40 additional iterations. The best result was achieved by the Trust Region method with average errors of 0.73%/3.12% with 33 additional iterations. With noise (compared to Table 10), only the Polak Ribière method could improve the parameter vector yielding 0.72%/2.25% with 11 iterations.

Please note that in the case of noise, the result of only one replicate was indicated above in order to show exemplarily the capacity of the algorithms. Of course, the hypothesis that no improvement was achieved was tested with replicates.

6. Conclusion

An optimization procedure for the parameterization of molecular models was presented, wherein numerical optimization algo-

Table 3
Optimization results of loss function containing the density and the enthalpy of vaporization at $T = 75$ K, without introduction of artificial noise. The stopping criterion was $F(\boldsymbol{x}) \leqslant 10^{-5}$. The asterisk indicates that the stopping criterion was not fulfilled and that the Armijo step length search did not converge within 100 steps. In the case of PSB, the convergency was too slow. Therefore, the descent direction was not normalized which led to better values in much less time. The Newton Raphson results are exactly concordant with the Steepest Descent results, as the Newton direction was never accepted. The Trust Region constraint had to be weakened. Therefore, $\eta_1 = 0.2$.

Density and enthalpy of vaporization, $T = 75$ K, without noise							
Algorithm	# Iter.	# Eval.	$\boldsymbol{x}^{\text{opt}}$	ρ_l	Δh_v	$F(\boldsymbol{x}^{\text{opt}})$	$\|\nabla F(\boldsymbol{x}^{\text{opt}})\|$
Steepest Descent	9	62	0.300120 0.325710 0.013278 0.115343	−0.03%	0.19%	4×10^{-6}	0.02
Newton Raphson	9	152	0.300120 0.325710 0.013278 0.115343	−0.03%	0.19%	4×10^{-6}	0.02
PSB (Quasi Newton)	2*	$\boldsymbol{x}^2 =$	0.301444 0.327477 0.014151 0.114339	−0.58%	1.46%	2.5×10^{-5}	0.17
DFP (Quasi Newton)	4*	$\boldsymbol{x}^4 =$	0.301445 0.327476 0.014145 0.114337	−0.58%	1.46%	2.5×10^{-5}	0.17
BFGS (Quasi Newton)	8*	$\boldsymbol{x}^8 =$	0.301446 0.327473 0.014145 0.114338	−0.83%	0.35%	8.1×10^{-5}	7.21
Fletcher Reeves (CG)	36	235	0.300137 0.325656 0.013175 0.115355	−0.01%	0.16%	3×10^{-3}	0.02
Polak Ribière (CG)	6	41	0.300215 0.325784 0.013409 0.115276	−0.03%	0.29%	8×10^{-6}	0.03
Trust Region (DD)	No solution of Trust Region partial problem found: Hessian not spd						
Trust Region (Exact)	6	105	0.299802 0.325402 0.013067 0.115660	−0.03%	−0.14%	2×10^{-6}	0.02

Table 4
Optimization results of loss function containing the density and the enthalpy of vaporization at $T = 75$ K, with introduction of artificial noise (0.5% for the density and 1% for the enthalpy of vaporization). The stopping criterion was $F(\boldsymbol{x}) \leqslant 3.5 \times 10^{-3}$. In all columns, the average values of ten replicates are shown. In order to weaken the Armijo and Trust Region constraints, the settings $\zeta_A = 0.001$ and $\eta_1 = 0.2$ were made, respectively.

Density and enthalpy of vaporization, $T = 75$ K, with noise								
Algorithm	# Repl.	# Iter.	# Eval.	$\boldsymbol{x}^{\text{opt}}$	ρ_l	Δh_v	$F(\boldsymbol{x}^{\text{opt}})$	$\|\nabla F(\boldsymbol{x}^{\text{opt}})\|$
Steepest Descent	8	7–8	63–64	0.305494 0.331350 0.016490 0.110247	−0.80%	5.22%	2.9×10^{-3}	0.72
Newton Raphson	8	8–9	161–162	0.306133 0.331285 0.016791 0.109278	0.10%	5.68%	3.3×10^{-3}	0.82
PSB (Quasi Newton)	5	5–6	62	0.313424 0.326680 0.024047 0.126885	−2.81%	2.94%	2.4×10^{-3}	0.71
DFP (Quasi Newton)	4	9	95–96	0.306095 0.331706 0.017046 0.109435	−0.23%	5.67%	3.3×10^{-3}	0.80
BFGS (Quasi Newton)	No replicate successful							
Fletcher Reeves (CG)	3	10–11	191–192	0.306708 0.331669 0.017046 0.108642	−0.12%	5.86%	3.4×10^{-3}	0.86
Polak Ribière (CG)	8	6–7	52	0.305550 0.331365 0.016605 0.110351	−0.61%	5.19%	2.8×10^{-3}	0.69
Trust Region (DD)	7	2–3	51–52	0.306105 0.331997 0.016823 0.109954	−0.71%	5.22%	2.8×10^{-3}	0.70
Trust Region (Exact)	10	3–4	70–71	0.305743 0.331171 0.015991 0.111001	−1.01%	4.98%	2.6×10^{-3}	0.60

Table 5
Optimization results of loss function containing the density and the enthalpy of vaporization at six temperatures, without introduction of artificial noise. The stopping criterion was $F(\boldsymbol{x}) \leqslant 0.01$. The errors on density and enthalpy of vaporization are averages over the absolute percental errors at the different temperatures. The asterisk indicates that the stopping criterion was not fulfilled and that the Armijo step length search did not converge within 100 steps. In the case of BFGS, the approximation of the Hessian matrix was singular. Two asterisks indicate that the convergency was too slow and that the workflow was artificially interrupted. The Armijo constraint had to be weakened in the cases of PSB, DFP and Fletcher Reeves. Therefore the setting $\zeta_A = 0.1$ was made.

Density and enthalpy of vaporization, 6 temperatures, without noise							
Algorithm	# Iter.	# Eval.	$\boldsymbol{x}^{\text{opt}}$	ρ_l	Δh_v	$F(\boldsymbol{x}^{\text{opt}})$	$\|\nabla F(\boldsymbol{x}^{\text{opt}})\|$
Steepest Descent	50	346	0.304200 0.326315 0.011931 0.114235	0.73%	3.29%	9.9×10^{-3}	0.53
Newton Raphson	42	715	0.304112 0.326501 0.011745 0.114031	0.69%	3.31%	9.9×10^{-3}	0.31
PSB (Quasi Newton)	100**	$\boldsymbol{x}^{100} =$	0.312485 0.327299 0.017722 0.111616	3.83%	10.00%	0.1336	31.26
DFP (Quasi Newton)	24*	$\boldsymbol{x}^{24} =$	0.301446 0.327473 0.014145 0.114338	1.72%	3.52%	0.0134	1094.36
BFGS (Quasi Newton)	0*	$\boldsymbol{x}^{0} =$	0.310100 0.331000 0.020737 0.104640	5.49%	21.86%	0.3848	65.72
Fletcher Reeves (CG)	93	569	0.304203 0.326076 0.012550 0.114217	0.72%	3.31%	9.9×10^{-3}	0.19
Polak Ribière (CG)	33	224	0.304313 0.325951 0.012119 0.114163	0.72%	3.30%	9.9×10^{-3}	0.24
Trust Region (DD)	No solution of Trust Region partial problem found: Hessian not spd						
Trust Region (Exact)	30	553	0.304172 0.326158 0.012438 0.114130	0.70%	3.32%	9.9×10^{-3}	0.29

Table 6
Optimization results of loss function containing the density and the enthalpy of vaporization at six temperatures, with introduction of artificial noise (0.5% for the density and 1% for the enthalpy of vaporization). The stopping criterion was $F(\boldsymbol{x}) \leqslant 0.013$. The errors on density and enthalpy of vaporization are averages over the absolute percental errors at the different temperatures. In all columns, the average values of ten replicates are shown. The Armijo constraint had to be weakened in the case of PSB and Fletcher Reeves. Therefore, $\zeta_A = 0.1$ was set. The Trust Region constraint had to be weakened as well. The setting $\eta_1 = 0.5$ was made.

Density and enthalpy of vaporization, 6 temperatures, with noise								
Algorithm	# Repl.	# Iter.	# Eval.	$\boldsymbol{x}^{\text{opt}}$	ρ_l	Δh_v	$F(\boldsymbol{x}^{\text{opt}})$	$\|\nabla F(\boldsymbol{x}^{\text{opt}})\|$
Steepest Descent	5	9–10	84	0.303881 0.329132 0.016636 0.114858	1.93%	3.21%	0.0125	3.15
Newton Raphson	7	7–8	134–135	0.306369 0.327030 0.017582 0.116017	0.63%	3.47%	0.0118	4.63
PSB (Quasi Newton)	No replicate successful							
DFP (Quasi Newton)	3	9–10	91–92	0.304235 0.328549 0.016598 0.115224	1.80%	3.19%	0.0118	2.46
BFGS (Quasi Newton)	No replicate successful							
Fletcher Reeves (CG)	8	8–9	61–62	0.303858 0.327724 0.016460 0.115816	1.78%	3.19%	0.0121	2.88
Polak Ribière (CG)	7	7–8	81–82	0.304027 0.328910 0.016597 0.114834	1.74%	3.33%	0.0124	3.51
Trust Region (DD)	3	5–6	111–112	0.304029 0.329058 0.016557 0.115179	2.09%	3.16%	0.0120	2.69
Trust Region (Exact)	3	5	97–98	0.303701 0.328640 0.016466 0.115193	1.84%	3.22%	0.0118	2.40

Table 7
Optimization results of loss function containing the density and the vapor pressure at $T = 75$ K, without introduction of artificial noise. The stopping criterion was $F(\boldsymbol{x}) \leqslant 10^{-4}$. The initial parameter vector was obtained by 3 steepest descent iterations with a heuristic step length, which led on the boundary of the admissible domain. Therefore, the new admissible domains for σ and ϵ were reduced. The asterisk indicates that the stopping criterion was not fulfilled and that the Armijo step length search did not converge within 100 steps. The Armijo constraint had to be weakened in the case of PSB. Therefore, the setting $\zeta_A = 0.1$ was made.

Density and vapor pressure, $T = 75$ K, without noise							
Algorithm	# Iter.	# Eval.	$\boldsymbol{x}^{\text{opt}}$	ρ_l	p_σ	$F(\boldsymbol{x}^{\text{opt}})$	$\|\nabla F(\boldsymbol{x}^{\text{opt}})\|$
Steepest Descent	106	670	0.329080 0.349806 0.032471 0.084349	0.91%	0.34%	9.5×10^{-5}	0.37
Newton Raphson	145	2343	0.324721 0.351530 0.027617 0.079118	0.94%	0.03%	8.9×10^{-5}	0.17
PSB (Quasi Newton)	14	214	0.309356 0.354817 0.041032 0.074182	0.85%	0.48%	9.6×10^{-5}	0.51
DFP (Quasi Newton)	88*	$\boldsymbol{x}^{88} =$	0.342390 0.335250 0.007313 0.081177	13.58%	1.9%	0.018812	3.31
BFGS (Quasi Newton)	16*	$\boldsymbol{x}^{16} =$	0.323386 0.338228 0.028883 0.076116	14.16%	1.99%	0.020433	3.30
Fletcher Reeves (CG)	157*	$\boldsymbol{x}^{157} =$	0.328757 0.346543 0.032610 0.083449	4.02%	0.34%	0.001630	0.77
Polak Ribière (CG)	93	580	0.329154 0.349809 0.032640 0.084373	0.92%	−0.13%	8.1×10^{-5}	0.25
Trust Region (DD)	No solution of Trust Region partial problem found: Hessian not spd						
Trust Region (Exact)	29	572	0.329153 0.350085 0.032568 0.084390	0.68%	−0.18%	5×10^{-5}	0.26

Table 8
Optimization results of loss function containing the density and the vapor pressure at $T = 75$ K, with introduction of artificial noise (0.5% for the density and 3% for the vapor pressure). The stopping criterion was $F(\boldsymbol{x}) \leqslant 5 \times 10^{-3}$. In all columns, the average values of ten replicates are shown. The initial parameter vector was obtained by 3 steepest descent iterations with a heuristic step length, which led on the boundary of the admissible domain. Therefore, the new admissible domain for ϵ were reduced but not for σ, as this led to convergency problems of the Armijo step length. The Armijo constraint also had to be weakened drastically. Therefore, $\zeta_A = 0.001$ was set. The Trust Region constraint had to be weakened as well. The setting $\eta_1 = 0.2$ was made. In the case of DFP, the last iterations were steepest descent iterations, as the approximation of the Hessian was singular.

Density and vapor pressure, $T = 75$ K, with noise								
Algorithm	# Repl.	# Iter.	# Eval.	$\boldsymbol{x}^{\text{opt}}$	ρ_l	p_σ	$F(\boldsymbol{x}^{\text{opt}})$	$\|\nabla F(\boldsymbol{x}^{\text{opt}})\|$
Steepest Descent	4	14–15	155–156	0.327347 0.343816 0.032180 0.081839	6.69%	0.68%	4.8×10^{-3}	2.41
Newton Raphson	No replicate successful							
PSB (Quasi Newton)	2	11–12	121	0.338901 0.344088 0.017238 0.084665	4.34%	0.16%	2.6×10^{-3}	1.77
DFP (Quasi Newton)	3	20	298–299	0.328169 0.343964 0.029897 0.081085	6.65%	−0.47%	4.7×10^{-3}	2.33
BFGS (Quasi Newton)	No replicate successful							
Fletcher Reeves (CG)	7	12–13	94–95	0.327520 0.344102 0.032243 0.081782	6.60%	0.24%	4.5×10^{-3}	1.64
Polak Ribière (CG)	7	12	79–80	0.327505 0.344212 0.032151 0.082050	6.47%	0.53%	4.5×10^{-3}	2.32
Trust Region (DD)	No solution of Trust Region partial problem found: Hessian not spd							
Trust Region (Exact)	7	8–9	158–159	0.327377 0.344348 0.031970 0.081577	6.40%	0.81%	4.4×10^{-3}	2.18

Table 9
Optimization results of loss function containing the density and the vapor pressure at six temperatures, without introduction of artificial noise. The stopping criterion was $F(\boldsymbol{x}) \leqslant 0.012$. The errors on density and vapor pressures are averages over the absolute percental errors at the different temperatures. The Newton Raphson results are exactly concordant with the Steepest Descent results, as the Newton direction was never accepted. The asterisk indicates that the stopping criterion was not fulfilled and that the Armijo step length search did not converge within 100 steps. In the case of BFGS, the approximation of the Hessian matrix was singular. Two asterisks indicate that the algorithm tends to move in the direction of the boundary of the admissible domain: Therefore, the workflow was interrupted at some point. The Armijo constraint had to be weakened in the case of DFP ($\zeta_A = 0.1$) and drastically in the case of PSB ($\zeta_A = 0.001$).

Density and vapor pressure, 6 temperatures, without noise							
Algorithm	# Iter.	# Eval.	$\boldsymbol{x}^{\text{opt}}$	ρ_l	p_σ	$F(\boldsymbol{x}^{\text{opt}})$	$\|\nabla F(\boldsymbol{x}^{\text{opt}})\|$
Steepest Descent	20	143	0.304677 0.328113 0.017685 0.111711	1.26%	3.65%	0.011915	1.15
Newton Raphson	20	143	0.304677 0.328113 0.017685 0.111711	1.26%	3.65%	0.011915	1.15
PSB (Quasi Newton)	552	9438	0.339904 0.317919 0.021602 0.134778	2.85%	2.73%	0.011883	14.07
DFP (Quasi Newton)	300**	$\boldsymbol{x}^{300} =$	0.351747 0.319875 0.021347 0.133568	4.54%	28.39%	0.515222	106.82
BFGS (Quasi Newton)	58*	$\boldsymbol{x}^{58} =$	0.311995 0.340508 0.00955 0.116957	9.53%	15.97%	0.214957	60.16
Fletcher Reeves (CG)	20	144	0.304690 0.328137 0.017624 0.111662	1.26%	3.72%	0.011999	1.67
Polak Ribière (CG)	20	143	0.304710 0.328130 0.017709 0.111691	1.27%	3.75%	0.011961	1.97
Trust Region (DD)	No solution of Trust Region partial problem found: Hessian not spd						
Trust Region (Exact)	7	140	0.304695 0.328128 0.017678 0.111701	1.26%	3.70%	0.011938	1.52

rithms were applied and assessed, in order to fit the saturated liquid density, the enthalpy of vaporization and the vapor pressure to experimental data by minimizing a loss function. This was performed at one and at six different temperatures simultaneously. A direct relation of the parameters to be optimized and the physical properties was given by fit functions for VLE data based on the 2CLJQ potential. As an example, the 2CLJQ model for nitrogen was chosen.

The gradient-based iterative procedures considered have very good convergency qualities. The gradient and the Hessian were approximated by finite differences of the physical properties. The Steepest Descent method, the Newton Raphson method, three Quasi Newton methods, two Conjugate Gradient methods and the Trust Region method with two different solutions of the Trust Region subproblem were investigated.

The algorithms were assessed for the parameterization problem, also with respect to some artificial noise in the calculated properties. Thereby, it was very important which performance could be expected at each single problem, leading to different stopping criteria for each optimization task. A very strict stopping criterion for the loss function was fulfilled in the case of the optimization of the saturated liquid density and the enthalpy of vaporization at one temperature, a somewhat weaker one in the case of vapor pressure. It became increasingly difficult, when the properties were noisy and when several temperatures were considered. Therefore, even weaker stopping criteria were used.

The Quasi Newton methods and the solution of the Trust Region subproblem by a DD algorithm turned out not to be suitable for this optimization task, as the matrices involved in those methods were not spd and at some iteration steps even singular. The Steepest Descent, the Newton Raphson, the Fletcher Reeves, the Polak Ribière, and the Trust region methods with an exact solution of the subproblem mostly fulfilled the respective stopping criterion within a reasonable number of iterations, regardless of the presence of noise, where the physical properties were mostly predicted within an error bar around the minimum. However, the Newton Raphson method often used the Steepest Descent direction, as the Hessian was not spd in most cases. Furthermore, please note that in the case of noise, the Newton direction may be accepted by mistake, although the Hessian is not spd. This, in turn, leads to the fact that the Newton Raphson method does not fulfill the stopping criterion in some cases. But as it worked quite well for the enthalpy of vaporization, it is considered as a suitable method anyway.

The Conjugate Gradient methods and the Trust Region method were found to be the best numerical optimization algorithms for the present scenarios, as they led to the best results in most cases and were very robust with respect to noise. The Trust Region method converged within less iterations than the Conjugate Gradient methods but it needs more function evaluations due to the calculation of the Hessian.

So, we can conclude that gradient-based numerical optimization algorithms are suitable, even if there is noise in the loss function evaluations. If they are applied to molecular simulations, one must always take care of the complexity, as for each evaluation of the loss function, simulations have to be performed. At each iteration, the gradient has to be computed and an Armijo step length control must be used. Some algorithms also require the calculation of the Hessian. Hence, some considerations have to be undertaken

Table 10
Optimization results of loss function containing the density and the vapor pressure at six temperatures, with introduction of artificial noise (0.5% for the density and 3% for the vapor pressure). The stopping criterion was $F(\boldsymbol{x}) \leqslant 0.015$. The errors on density and vapor pressure are averages over the absolute percental errors at the different temperatures. In all columns, the average values of ten replicates are shown. The Armijo constraint had to be weakened drastically in the case of Newton Raphson ($\zeta_A = 0.001$). In the case of PSB and DFP, the last iterations were steepest descent iterations, as the approximation of the Hessian was singular.

Density and vapor pressure, 6 temperatures, with noise								
Algorithm	# Repl.	# Iter.	# Eval.	$\boldsymbol{x}^{\text{opt}}$	ρ_l	p_σ	$F(\boldsymbol{x}^{\text{opt}})$	$\|\nabla F(\boldsymbol{x}^{\text{opt}})\|$
Steepest Descent	9	6	46–47	0.305610 0.328488 0.017561 0.111967	1.06%	3.79%	0.0118	8.68
Newton Raphson	1	11	258	0.309644 0.325261 0.015911 0.112473	3.93%	2.58%	0.0142	12.74
PSB (Quasi Newton)	No replicate successful							
DFP (Quasi Newton)	7	8	101–102	0.306359 0.327853 0.017664 0.112784	1.21%	3.35%	0.0105	3.45
BFGS (Quasi Newton)	5	8–9	73–74	0.308596 0.327424 0.020029 0.114783	1.12%	3.90%	0.0119	11.31
Fletcher Reeves (CG)	10	6–7	82	0.305510 0.328466 0.017439 0.112224	0.86%	3.69%	0.0125	10.11
Polak Ribière (CG)	10	5–6	55–56	0.305590 0.328411 0.017556 0.111956	1.13%	3.90%	0.0132	15.63
Trust Region (DD)	8	4–5	89–90	0.305605 0.328515 0.017454 0.112047	0.92%	3.82%	0.0112	8.30
Trust Region (Exact)	5	5	113–14	0.305578 0.328435 0.017474 0.111974	1.00%	3.39%	9.8×10^{-3}	9.18

in order to reduce the complexity when using molecular simulations, as the number of function evaluations was always quite high. However, the present work shows that it is possible to apply high-performance gradient-based iteration procedures in this field, and their application to molecular simulations will be presented in a coming publication [68].

Acknowledgements

We are grateful to Axel Arnold, Thorsten Köddermann, and Florian Müller-Plathe for valuable discussions and appreciate their intellectual support to our work. Marco Hülsmann acknowledges the financial support for his scholarship of the University of Cologne (Germany).

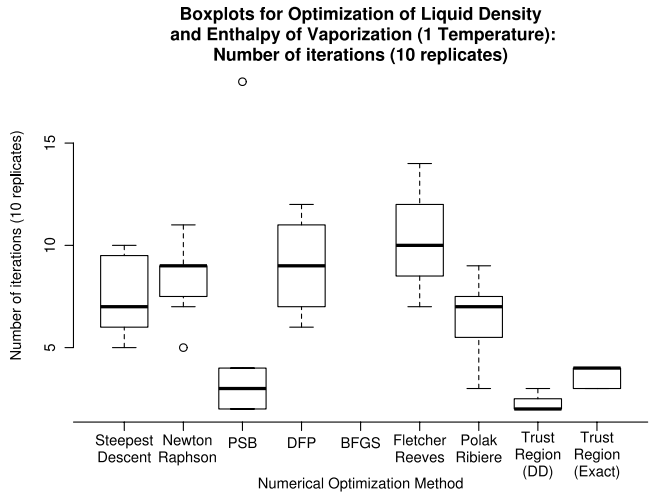
Appendix A. Result tables

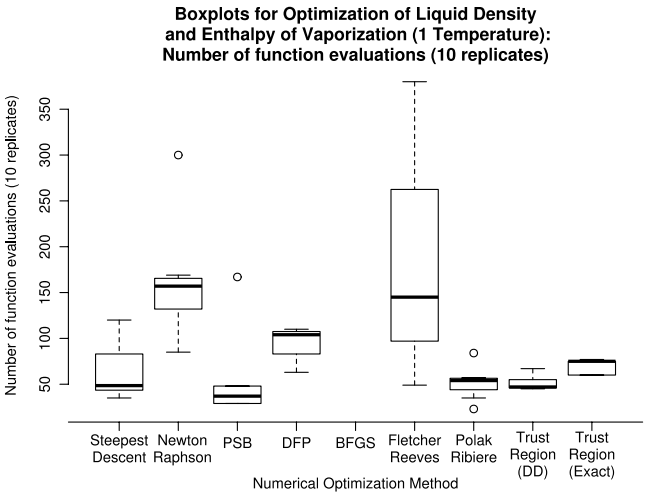
Tables 3–10 show the results for all eight optimization tasks, without and with artificial noise.

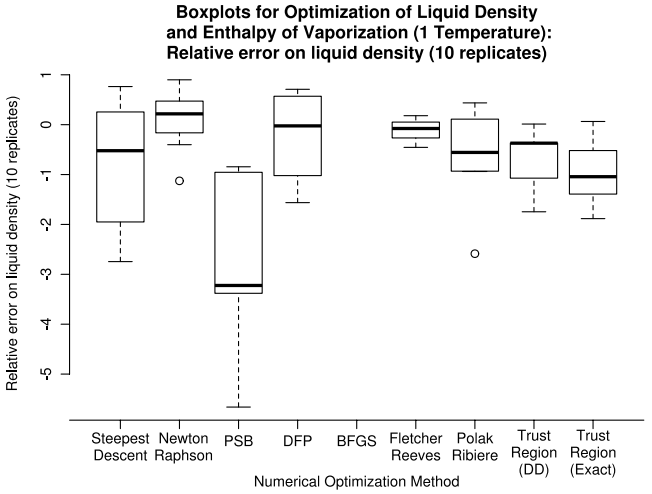
The tables contain all gradient-based numerical optimization methods considered, i.e. Steepest Descent, Newton Raphson, three Quasi Newton methods (PSB, DFP and BFGS), two Conjugate Gradient methods (Fletcher Reeves and Polak Ribière) as well as the Trust Region method with two different solutions of the Trust Region subproblem (DD algorithm and exact solution).

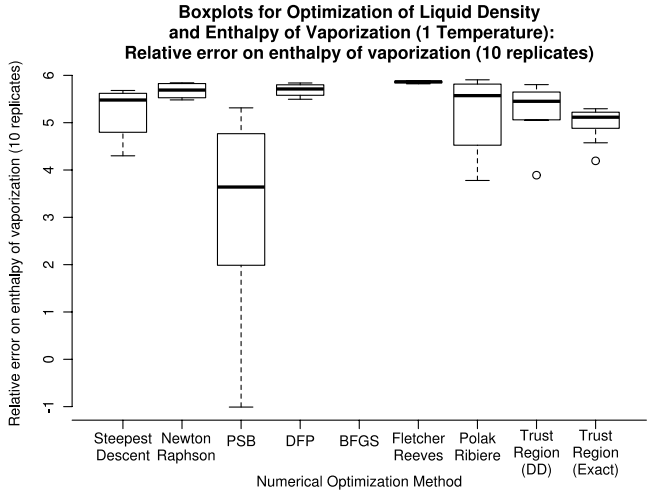
The tables contain the following details:

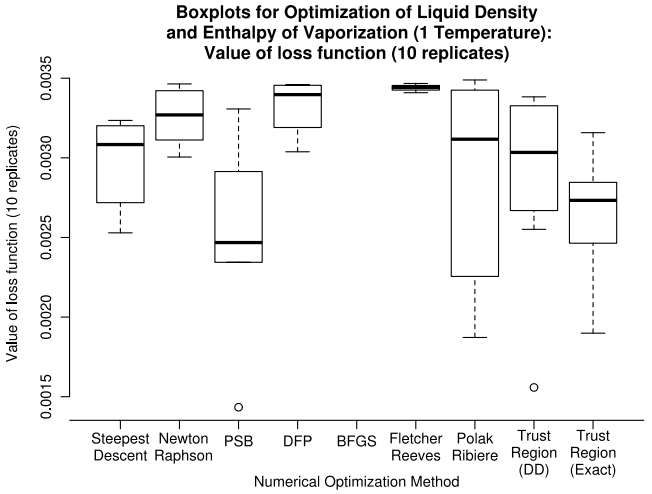
- Optimization method (Algorithm): One of the above mentioned numerical optimization methods.
- Number of iterations (# Iter.): Those are the iteration steps required, in order to fulfill the stopping criterion. Hence, it is the $P \in \mathbb{N}$ so that $\boldsymbol{x}^P = \boldsymbol{x}^{\text{opt}}$, starting from $\boldsymbol{x}^0$. The number of iterations indicates the speed of convergency.
- Number of function evaluations (# Eval.): This number indicates how often the loss function or its partial derivatives had to be evaluated, i.e. how often the physical properties had to be calculated, e.g. by molecular simulations, until the stopping criterion was fulfilled. This number increases with the computation of the gradient and the Hessian as well as the number of iterations required for the Armijo step length control. If the algorithm was interrupted for some reason before the stopping criterion was fulfilled, the column contains '$\boldsymbol{x}^{\tilde{P}} =$', where $\tilde{P}$ is the number of iterations so far, followed by the parameter $\boldsymbol{x}^{\tilde{P}}$ itself in the next column.
- Final parameter ($\boldsymbol{x}^{\text{opt}}$): The final parameter vector is given, either the optimal one, for which the stopping criterion is fulfilled, or the last one before the optimization was interrupted.
- Error on saturated liquid density (ρ_l): This is the percental error on the saturated liquid density. In the case of one temperature, a negative sign indicates that ρ_l was underestimated, a positive sign that it was overestimated. In the case of six different temperatures, the Mean Absolute Percental Error (MAPE) over all temperatures is indicated.
- Error on enthalpy of vaporization or vapor pressure (Δh_v or p_σ): The percental error on enthalpy of vaporization or vapor pressure. It is computed exactly in the same way as the error on the saturated liquid density.
- Value of loss function ($F(\boldsymbol{x}^{\text{opt}})$): The value of the loss function for the final or last parameter vector. Whenever the stopping











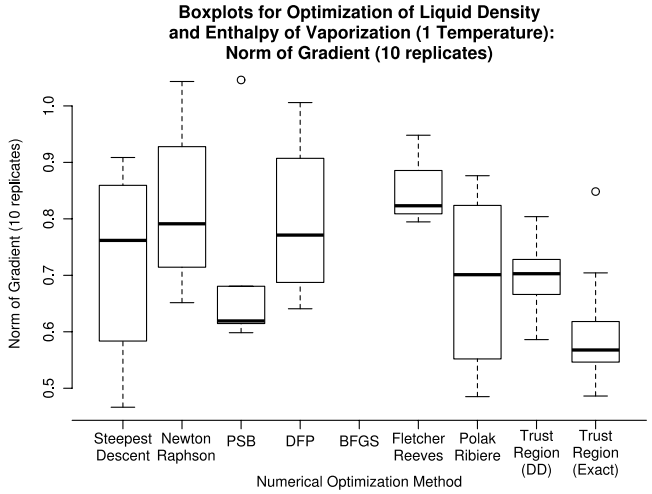
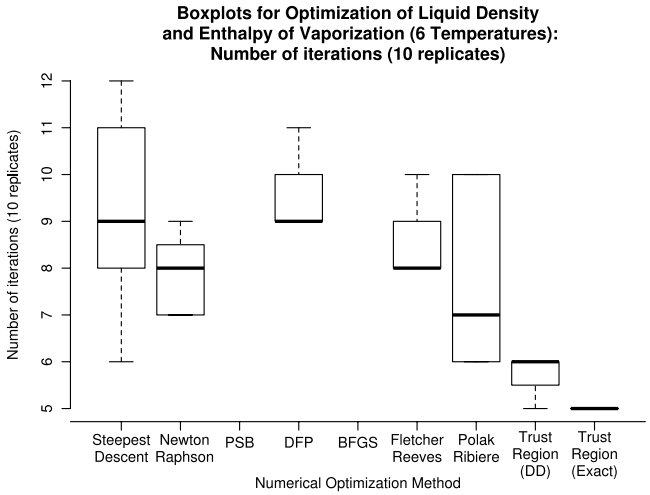
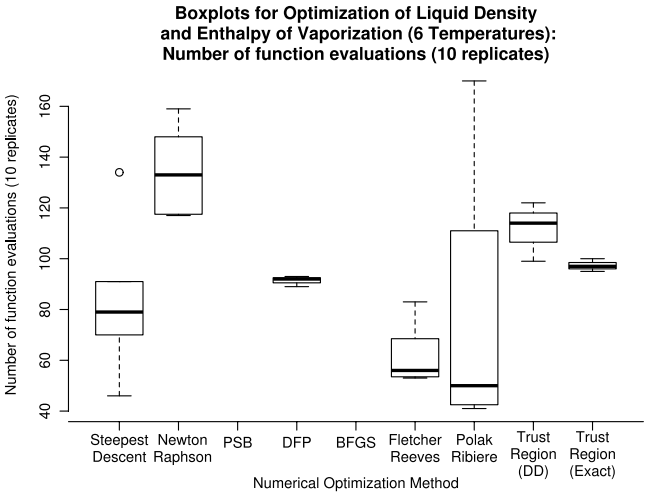
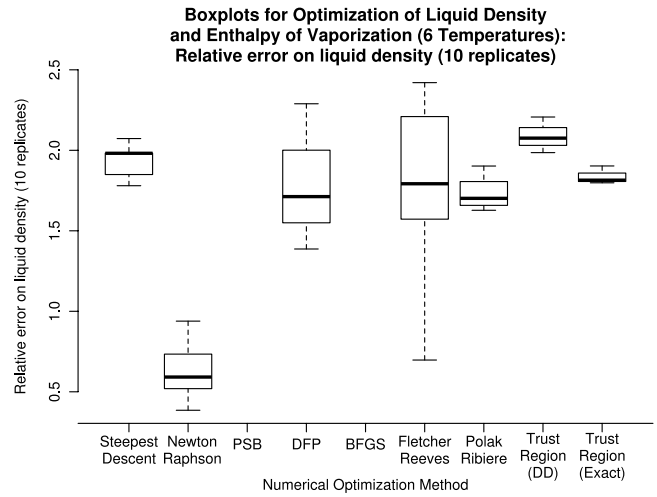
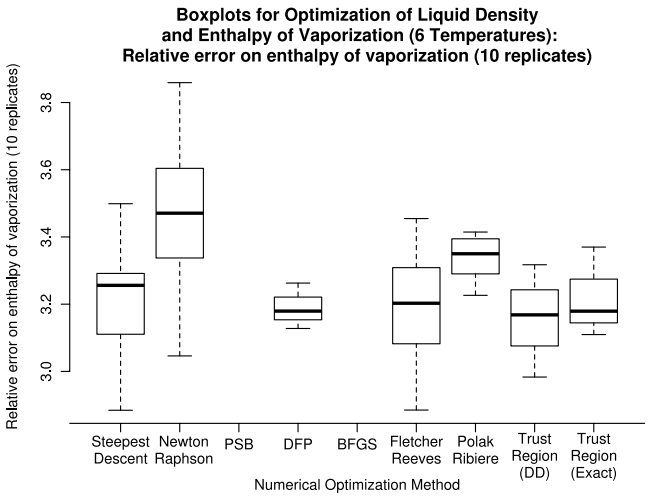


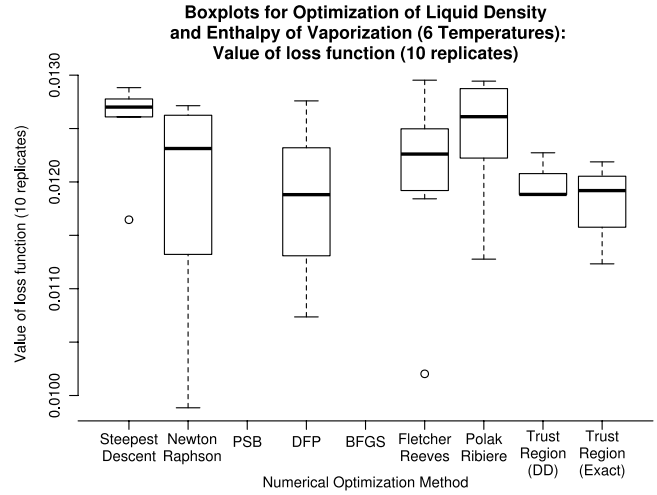
Fig. 5. Box plots corresponding to the columns of Table 4: Saturated liquid density plus enthalpy of vaporization at $T = 75$ K.











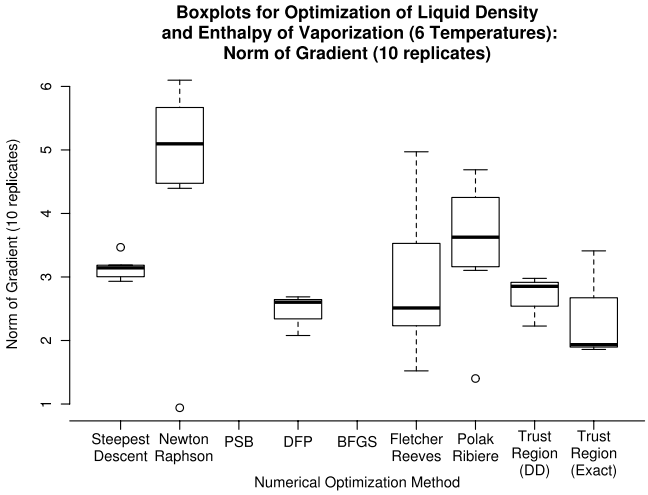


Fig. 6. Box plots corresponding to the columns of Table 6: Saturated liquid density plus enthalpy of vaporization at six different temperatures.

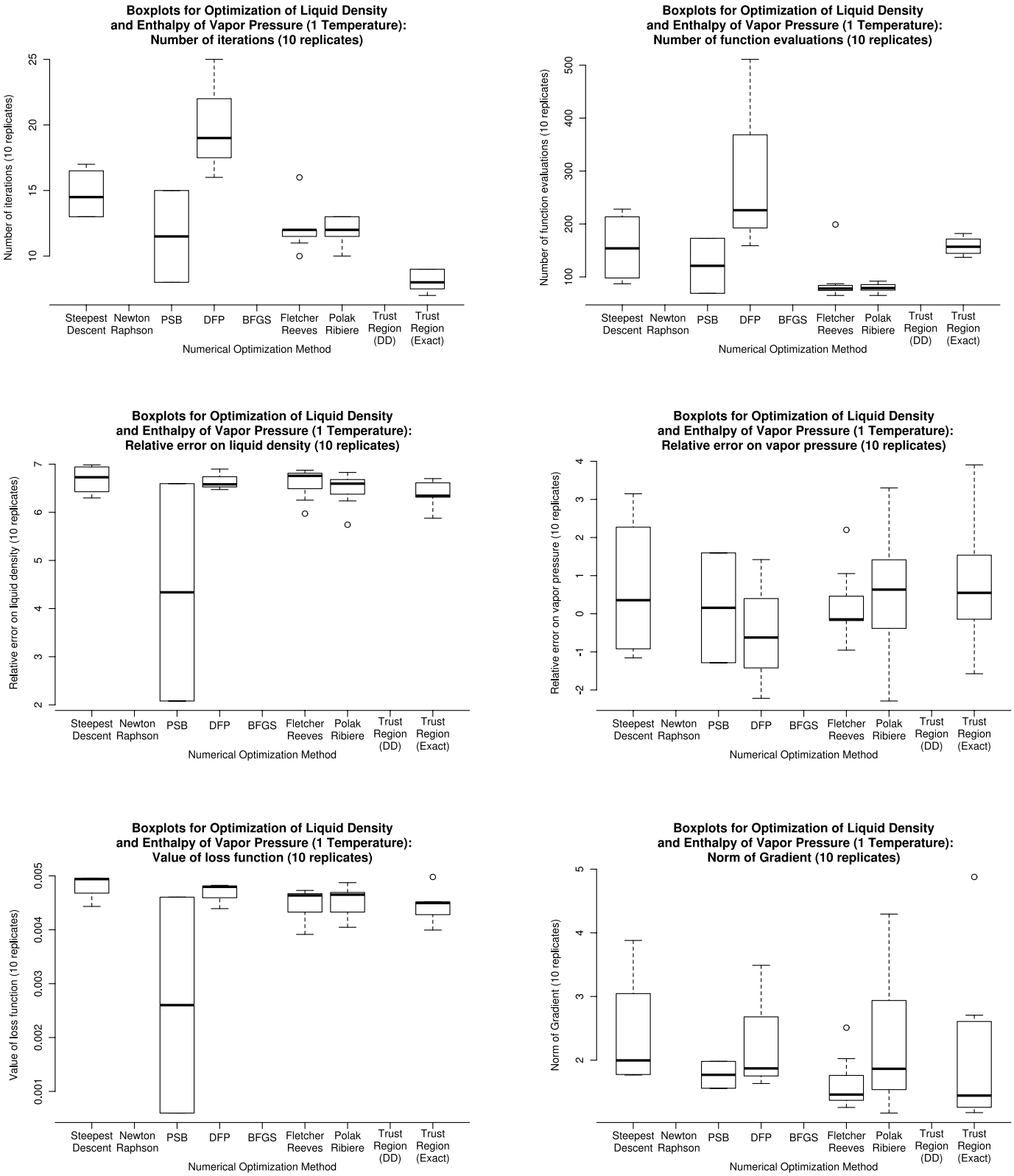
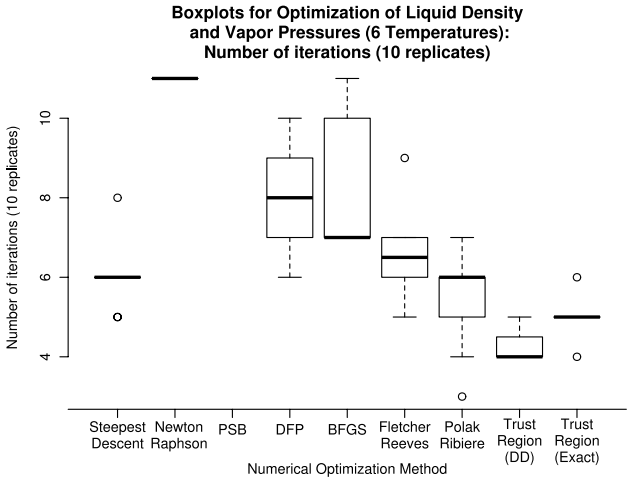
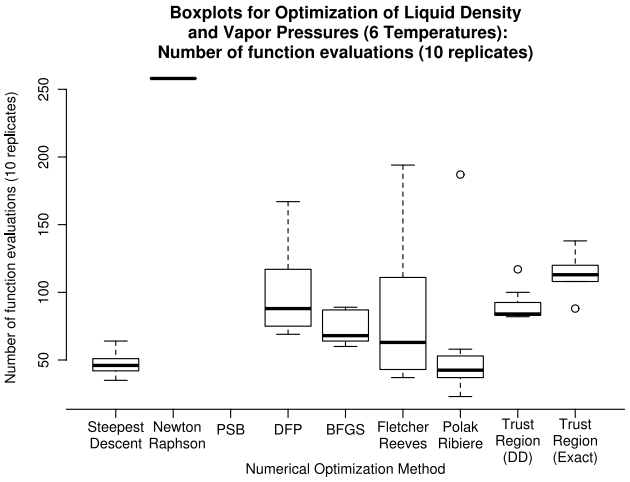
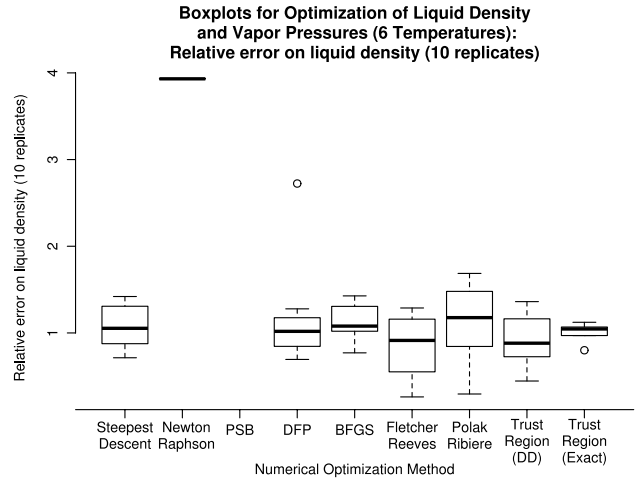
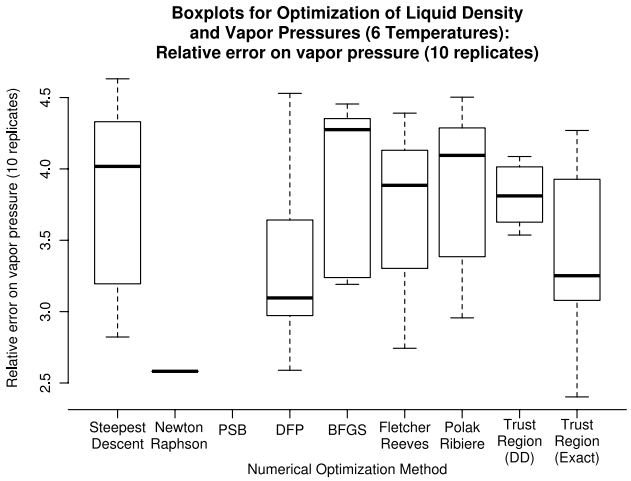


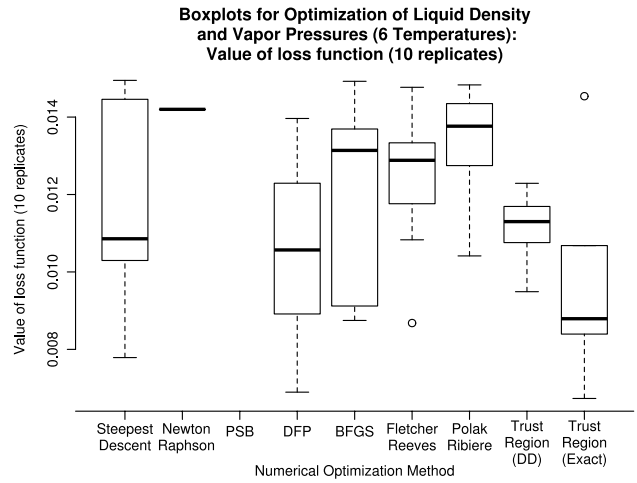
Fig. 7. Box plots corresponding to the columns of Table 8: Saturated liquid density plus vapor pressure at $T = 75$ K.











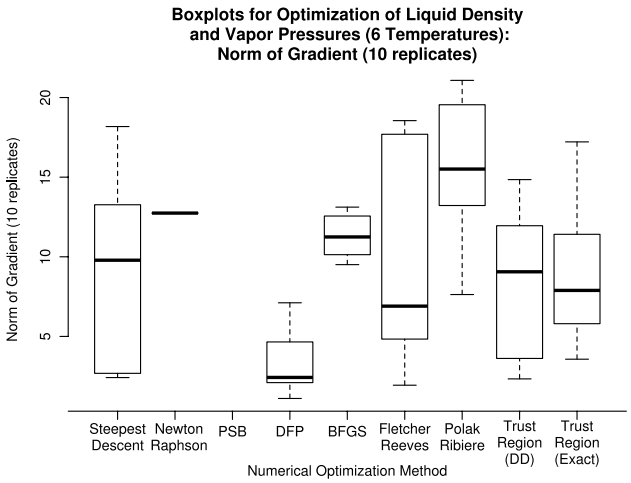


Fig. 8. Box plots corresponding to the columns of Table 10: Saturated liquid density plus vapor pressure at six different temperatures.

criterion is fulfilled, this value is lower than the upper bound of the criterion. The loss function should be as close to zero as possible.

- Norm of the gradient ($\|\nabla F(\boldsymbol{x}^{\mathrm{opt}})\|$): The norm of the gradient for the final iteration. This value should be close to zero as well, but it cannot be expected that it gets as small as $F(\boldsymbol{x}^{\mathrm{opt}})$ for the reasons mentioned in Section 4.

In the case of noise, the average values over ten replicates are indicated. The second column (# Repl.) contains the number of successful replicates (maximum: 10), i.e. the replicates which fulfilled the stopping criterion without being interrupted before. Please note again that a workflow was interrupted whenever the Armijo step length control algorithm did not converge within 100 steps or, in the case of the Quasi Newton methods, if H_k was singular.

Appendix B. Box plots for artificial noise

Figs. 5–8 show the box plots corresponding to the columns of the result Tables 4, 6, 8, and 10, respectively. The width of a box plot indicates the robustness of the method with respect to noise: The narrower the box plot, the more robust is the corresponding method. Furthermore, the position of a box plot is crucial: The bold line stands for the median of the sample consisting of ten replicates. Therefore, if the median is located at a value close to zero and the box plot is narrow with only a small number of outliers, the method features a good performance and is robust with respect to noise as well.

The lower and upper boundaries of a box plot are the first and third quartiles of the sample, the largest non-outlier observations are marked by error bars (so-called *whiskers*). Outliers are marked by circles.

References

[1] S.J. Singer, G.L. Nicolson, Science 175 (1972) 720.
[2] M.P. Allen, D.J. Tildesley, Computer Simulations of Liquids, Oxford Science, Oxford, 1987.
[3] Y. Zhou, G. Stell, J. Chem. Phys. 96 (1992) 1504.
[4] J.I. Siepmann, S. Karaborni, B. Smit, Nature 365 (1993) 330.
[5] S. O'Connell, P.A. Thompson, Phys. Rev. E 52 (1995) 5792.
[6] J. Kolafa, I. Nezbeda, M. Lisal, Mol. Phys. 99 (2001) 1751.
[7] R. Valiullin, S. Naumov, P. Galvosas, J. Kärger, H.-J. Woo, F. Porcheron, P.A. Monson, Nature 443 (2006) 965.
[8] I.P. Batra, B.I. Bennett, F. Herman, Phys. Rev. B 11 (1975) 4972.
[9] T.P. Fehlner, J. Solid State Chem. 154 (2000) 110.
[10] C.N. Della, S. Dongwei, Diffus. Defect Data Pt. B Solid State Phenom. 136 (2008) 45.
[11] S.-T. Lin, M. Blanco, W.A. Goddard, III, J. Chem. Phys. 119 (2003) 11792.
[12] D.E. Bien, V.A. Chiriac, in: Proceedings of the 9th Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems, IEEE, New Jersey, 2004, p. 748.
[13] J. Vrabec, J. Gross, J. Phys. Chem. B 112 (2008) 51.
[14] A. Hodgkin, A. Huxley, J. Physiol. 117 (1952) 500.
[15] B.J. Barkla, O. Pantoja, Annu. Rev. Plant Physiol. Plant Mol. Biol. 47 (1996) 159.
[16] M. Levitt, A. Warshe I, Nature 253 (1975) 694.
[17] J. Gsponer, A. Caflisch, Proceedings of the National Academy of Sciences (PNAS) 99 (10) (2002) 6719.
[18] C.D. Snow, E.J. Sorin, Y.M. Rhee, V.S. Pandel, Annu. Rev. Biophys. Biomol. Struct. 34 (2005) 43.
[19] F. Müller-Plathe, D. Reith, Comput. Theor. Polymer Sci. 9 (1999) 203.
[20] P. Bordat, D. Reith, F. Müller-Plathe, J. Chem. Phys. 115 (2001) 8978.
[21] G. Guevara-Carrion, C. Nieto-Draghi, J. Vrabec, H. Hasse, J. Phys. Chem. B 112 (2008) 16664.
[22] G.S. Grest, K. Kremer, Phys. Rev. A 33 (1986) 3628.
[23] F. Müller-Plathe, Acta Polymer 45 (1994) 259.
[24] K. Binder, Monte Carlo and Molecular Dynamics Simulations in Polymer Science, Oxford University Press, Oxford, 1995.
[25] K. Kremer, F. Müller-Plathe, Mol. Sim. 28 (2002) 729.
[26] M. Praprotnik, C. Junghans, L. Delle Site, K. Kremer, Comput. Phys. Comm. 179 (2008) 51.
[27] B.E. Poling, J.M. Prausnitz, J.P. O'Connell, The Properties of Gases and Liquids, 5th edition, McGraw–Hill Professional, New York, 2000.
[28] A.R. Leach, Molecular Modelling: Principles and Applications, 2nd edition, Prentice-Hall, New Jersey, 2001.
[29] W.L. Jorgensen, J.D. Madura, C.J. Swensen, J. Am. Chem. Soc. 106 (1984) 6638.
[30] M.G. Martin, J.I. Siepmann, J. Phys. Chem. B 102 (1998) 256.
[31] R. Faller, H. Schmitz, O. Biermann, F. Müller-Plathe, J. Comp. Chem. 20 (1999) 1009.
[32] P. Ungerer, C. Beauvais, J. Delhommelle, A. Boutin, B. Rousseau, A.H. Fuchs, J. Phys. Chem. 112 (2000) 5499.
[33] E. Bourasseau, M. Haboudou, A. Boutin, A.H. Fuchs, P. Ungerer, J. Chem. Phys. 118 (2003) 3020.
[34] J. Stoll, J. Vrabec, H. Hasse, J. Chem. Phys. 119 (2003) 11396.
[35] D. Reith, M. Pütz, F. Müller-Plathe, J. Comp. Chem. 24 (2003) 1624.
[36] C. Oostenbrink, A. Villa, A.E. Mark, W.F. van Gunsteren, J. Comp. Chem. 25 (2004) 1656.
[37] H. Sun, Fluid Phase Eq. 217 (2004) 59.
[38] K.N. Kirschner, A.B. Yongye, S.M. Tschampel, J. Gonzalez-Outeirino, C.R. Daniels, B.L. Foley, R. Woods, J. Comp. Chem. 29 (2008) 622.
[39] B. Eckl, J. Vrabec, H. Hasse, Fluid Phase Eq. 274 (2008) 16.
[40] B. Eckl, J. Vrabec, H. Hasse, J. Phys. Chem. B 112 (2008) 12710.
[41] W.L. Jorgensen, D.S. Maxwell, J. Tirado-Rives, J. Am. Chem. Soc. 118 (1996) 11225.
[42] J. Wang, P.A. Kollman, J. Comp. Chem. 22 (2001) 1219.
[43] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, J. Comp. Chem. 25 (2004) 1157.
[44] D. Yin, A.D. MacKerell Jr., J. Comp. Chem. 19 (1998) 334.
[45] N. Foloppe, A.D. MacKerell Jr., J. Comp. Chem. 21 (2000) 86.
[46] M. Hülsmann, J. Vrabec, T. Koeddermann, D. Reith, Comput. Phys. Comm. 181 (2010) 499.
[47] A. Rahman, Phys. Rev. 136 (1964) 405.
[48] H. Bertagnolli, M.D. Zeidler, Mol. Phys. 35 (1978) 177.
[49] F. Kohler, N. van Nhu, Mol. Phys. 80 (1993) 795.
[50] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, J. Chem. Phys. 79 (1983) 926.
[51] H.C. van Nesse, Pure Appl. Chem. 67 (1995) 859.
[52] J. Stoll, J. Vrabec, H. Hasse, J. Fischer, Fluid Phase Eq. 179 (2001) 339.
[53] J. Stoll, J. Vrabec, H. Hasse, Fluid Phase Eq. 209 (2003) 29.
[54] H. Zho, M. Gong, Y. Zhang, J. Wu, J. Chem. Eng. Data 51 (2006) 1201.
[55] P. Wagner, S. Horstmann, K. Fischer, Monatsh. Chem. Chem. Mon. 138 (2007) 637.
[56] W.R. Ji, E. Stiebing, G. Hradetzky, D.A. Lempe, Fluid Phase Eq. 260 (2007) 113.
[57] R.P.S. Peguin, G. Kamath, J.J. Potoff, S.R.P. da Rocha, J. Phys. Chem. B 113 (2009) 178.
[58] A.D. Buckingham, Q. Rev. Biophys. 13 (1959) 183.
[59] C.S. Murthy, K. Singer, M.L. Klein, I.R. McDonald, Mol. Phys. 41 (1980) 1387.
[60] C. Kriebel, A. Muller, M. Mecke, J. Winkelmann, J. Fischer, Int. J. Thermodyn. 17 (1996) 1349.
[61] J. Vrabec, J. Stoll, H. Hasse, J. Phys. Chem. B 105 (2001) 12126.
[62] E.A. Guggenheim, J. Chem. Phys. 13 (1945) 253.
[63] N.I.S.T. Chemistry Webbook, National Institute of Standards and Technology, USA, http://webbook.nist.gov/chemistry/.
[64] YASP Molecular Dynamics Simulation package (by F. Müller-Plathe), Documentation: http://www.theo.chemie.tu-darmstadt.de/group/services/yaspdoc/yaspdoc.html.
[65] T.J. Müller, S. Roy, W. Zhao, A. Maaß, D. Reith, Fluid Phase Eq 274 (2008) 27.
[66] J. Nocedal, S.J. Wright, Numerical Optimization, Springer-Verlag, New York, 1999.
[67] R. Span, E.W. Lemmon, R.T. Jacobsen, W. Wagner, A. Yokozeki, J. Phys. Chem. Ref. Data 29 (2000) 1361.
[68] M. Hülsmann, T.J. Müller, T. Köddermann, D. Reith, Mol. Sim. (2010), submitted for publcation.