Fluid Phase Equilibria 411 (2016) 33–42

Contents lists available at ScienceDirect

Fluid Phase Equilibria

journal homepage: www.elsevier.com/locate/fluid

Parametrization of two-center Lennard-Jones plus point-quadrupole force field models by multicriteria optimization

K. Stöbener [a], P. Klein [a], M. Horsch [b,*], K. Küfer [a], H. Hasse [b]

[a] Fraunhofer Institute for Industrial Mathematics (ITWM), Kaiserslautern, Germany
[b] Laboratory of Engineering Thermodynamics (LTD), University of Kaiserslautern, Germany

ARTICLE INFO

Article history:
Received 27 October 2015
Received in revised form
22 November 2015
Accepted 24 November 2015
Available online 30 November 2015

Keywords:
Optimization
Force field models
United-atom
Lennard-Jones
Multicriteria
Visualization

ABSTRACT

To ensure the quantitative precision and reliability of molecular simulations, force field models of molecular fluids need to be adjusted to e.g. experimental data. An optimal agreement for different properties is often not achieved by a single model parametrization. Applying multicriteria optimization, based on the evaluation and analysis of the Pareto set, solves this problem. The Pareto set contains all optimal compromises between multiple conflicting objectives. Its computation and suitable visualization enables the end user to freely choose a model parametrization, tailored to his particular application scenario.

We apply multicriteria optimization to the two-center Lennard-Jones plus point-quadrupole model class (2CLJQ), which has four adjustable parameters. The Pareto set is determined and analyzed for ten real fluids: Ethane, ethylene, acetylene, fluorine, chlorine, bromine, perfluoroethylene, perchloroethylene, nitrogen, and oxygen. Thereby, two multicriteria optimization scenarios are considered, based on two criteria (saturated liquid density and vapor pressure) and three criteria (saturated liquid density, vapor pressure, and surface tension), respectively. It is shown that literature models for these fluids can be further improved in these criteria. We visualize our results by self-organizing patch plots, which facilitate the representation of the entire Pareto set and its corresponding model parametrizations.

1. Introduction

Molecular simulations of fluids contribute to solving problems in physics, biology, chemistry, and engineering. They rely on the choice of suitable molecular force fields. Adjusting the force field parameters, aiming at an accurate representation of different thermophysical properties, is a multicriteria optimization problem. Different properties normally cannot simultaneously be optimized without a trade-off, i.e. the objectives are conflicting and a single optimal solution cannot be determined. The goal of multicriteria optimization is to identify the Pareto set, which represents best possible compromises between conflicting objectives. A solution is defined to be Pareto optimal if a further improvement in one objective can only be achieved at the expense of at least one other objective. The Pareto set is a subset of all feasible solutions in the objective space. To each Pareto optimal point in the objective space corresponds one parametrization in the parameter space, representing a Pareto optimal model. Hence, identifying the Pareto set does not yield one model, but a set of optimal models from which a user can choose the one best fitting a particular application scenario.

It is attractive to use multicriteria optimization for the parametrization of molecular models. In a preceding work, the Pareto optimal Lennard-Jones models for argon and methane were identified by brute force evaluation of 200×200 parameter combinations [1]. The same procedure was used by Werth et al. [2] to identify Pareto optimal parameters for representing carbon dioxide with a two-center Lennard-Jones plus point-quadrupole force field model. The grid in the parameter space employed by Werth et al. [2] was $60 \times 60 \times 60 \times 60$. This brute force enumeration is only feasible for scenarios in which the evaluation of the objective functions is not expensive and the number of parameters is not high. This is rarely the case for optimizations of molecular models. A more efficient strategy to approximate the Pareto set is necessary. Apart from Refs. [1] and [2], we are only aware of one other work in which multicriteria optimization was used in the context of developing force field models. Mosthagim et al. [3] developed Pareto sets by Particle Swarm Optimization with a focus on intramolecular potentials fitted to ab initio data.

* Corresponding author.
E-mail address: martin.horsch@mv.uni-kl.de (M. Horsch).

http://dx.doi.org/10.1016/j.fluid.2015.11.028
0378-3812/

In the present work, we determine the Pareto set by a combination of the sandwiching and hyperboxing algorithm [4]. The algorithm aims at efficiently approximating the Pareto set for a predefined approximation quality.

Multicriteria optimization is applied to the parametrization of the two-center Lennard-Jones plus point-quadrupole (2CLJQ) potential model in two optimization scenarios. The 2CLJQ potential model has four parameters, which are adjusted to optimize first two and then three different objective functions. However, for both optimization tasks, the Pareto optimal parameter sets are evaluated regarding four functions. They describe the quantitative agreement between the simulation data and the experimental data for the saturated liquid density, the vapor pressure, the surface tension, and the critical temperature. For simplicity, we refer to the saturated liquid density only as liquid density in the following.

In the first optimization scenario the conflicting objective functions for the liquid density and the vapor pressure are used (*two-criteria scenario*). Previous work by Fischer and collaborators [5,6] shows that model parameters adjusted to the liquid density and the vapor pressure are suitable to predict e.g. caloric or further thermodynamic properties. In the second scenario, the surface tension is included as an example for a third objective function (*three-criteria scenario*). For both scenarios the Pareto set is determined for ten different fluids: ethane (C_2H_6), ethylene (C_2H_4), acetylene (C_2H_2), fluorine (F_2), chlorine (Cl_2), bromine (Br_2), nitrogen (N_2), oxygen (O_2), perfluoroethylene (C_2F_4), and perchloroethylene (C_2Cl_4). The obtained Pareto optimal solutions are also compared to the performance of molecular models from the literature. The discussed literature models were developed on the basis of results of direct numerical molecular simulations using MD or MC methods. Molecular models obtained by theoretical studies as e.g. perturbation theory [7,8] were not taken into account.

In the main text, first the results for the two-criteria scenario are discussed for all fluids. The results for the three-criteria scenario are then discussed for acetylene and used for a comparison with those of the two-criteria scenario as a representative for all studied fluids. Finally, a brief summary of a comparison between the two- and three-criteria scenario for all fluids is given. The Pareto optimal solutions for both optimization tasks for all studied fluids can be found in the Supplementary Material.

We also introduce a novel technique for simultaneously visualizing the Pareto set in the objective space, the corresponding Pareto optimal model parameter space as well as additional functions, which were not included in the multicriteria optimization as an objective function. It is based on self-organizing maps (SOM), which are also referred to as Kohonen maps in the literature. With a SOM an interpolation of a high-dimensional data set can be displayed in low-dimensional maps [9,10]. Hunger and Huttner [11] used SOM to gain insight into the dependence of their single-criteria optimization on parameters specifying a force field description for tripod metal templates. SOMs have been used before in a few studies in the literature to represent results from multicriteria optimizations [12,13]. The drawback of SOMs is that they do not only represent the given input data set, but an interpolation, resulting in an approximation of the data set. Thus we enhanced the SOM and used it as a projection method for a given data set, in our case the Pareto set. Furthermore, by combining them with Voronoi diagrams [14], we achieved that only the exact Pareto optimal solutions are displayed. We refer to the representation as *self-organizing patch plot* (SOPP). It is a visualization specifically adapted to display any Pareto set.

By displaying the Pareto sets with the novel SOPP technique, a comprehensive overview over the optimal solutions is gained. It enables assessing how good simultaneous representations of different properties of pure fluids by the 2CLJQ model can be without having to refer to only one single solution for each studied fluid. As many pure fluids are studied and similar results are obtained, the present statements can probably be generalized: They allow an assessment of how good the 2CLJQ model can be.

The methods presented have a much wider scope. They can be used for parametrizing molecular force fields in general and well beyond for parametrizing any thermodynamic model.

2. Multicriteria optimization

A multicriteria optimization problem is characterized by multiple objective functions $f_i(x)$ which have to be minimized simultaneously:

$$\min f(x) = (f_1(x), \ldots, f_r(x)) \in \mathbb{R}^r. \tag{1}$$

They span the objective space $\mathbb{R}^r$ and depend on the decision vector $x \in \mathbb{R}^q$ where $\mathbb{R}^q$ is the design space. The solution to such a problem is a set of best compromises: For any improvement in a single objective $f_i(x)$, a decline in at least one other objective $f_k(x)$, $i \neq k$ has to be accepted. The set of all best compromises is called the Pareto set. In the context of force field parametrization the design space is spanned by the parameters describing the model. Thus we refer to it as parameter space in the subsequent text. Mapped to each point in the parameter space, hence, to each model, is one point in the objective space.

For more information on multicriteria optimization see e.g. Refs. [15–17].

2.1. Sandwiching and hyperboxing algorithm

The Pareto set is a subset of all feasible points in the objective space and needs to be approximated by a suitable numerical strategy. In the present work the sandwiching and hyperboxing algorithm is used for determining suitable approximations of the Pareto sets. This algorithm is basically taken from Ref. [4].

By scalarizations of the objective functions, single criterion optimization problems are obtained. The solutions for these scalarizations belong to the Pareto set. Thus by solving several single criterion optimizations a point–wise approximation of the Pareto set is obtained. The sandwiching and hyperboxing algorithm is used to subsequently identify suitable scalarizations to efficiently approximate the Pareto set. The algorithm focuses on exploring the Pareto set in regions, where the curvature of the Pareto set is highest and starts by identifying the extreme compromises, i.e. at first the minimum of each objective function is located. The sandwiching algorithm then alternately finds inner and outer approximations for the Pareto set, thereby assuming the Pareto set is convex. The sandwiching algorithm is applied, until a specified approximation quality is reached. Then, if regions of non-convex behavior are identified, the non-convex regions are approximated with the hyperboxing algorithm. The sandwiching algorithm uses the weighted sum scalarization (see e.g. Ref. [17]), whereas the hyperboxing algorithm uses a scalarization proposed by Pascoletti and Serafini [18]. A brief sketch of the concept of the sandwiching and hyperboxing algorithm in a chemical engineering context can be found in Bortz et al. [19]. To solve the single criterion optimization tasks with which the Pareto set is approximated, we employ the Quasi-Newton solver NLPQLP of Schittkowski [20].

During the multicriteria optimization numerical problems may occur. Depending on the choice and settings of the single criterion solver, the sandwiching and hyperboxing algorithm may identify solutions as Pareto optimal, which are not. Thus we additionally check the Pareto sets for Pareto optimality by comparing the solutions to each other and sort out solutions, which are not Pareto