

Data-driven Approximation of Thermodynamic Phase Equilibria

Ashfaq Iftakher,[†] Chinmay M. Aras,[†] Mohammed Sadaf Monjur,[†] and M. M. Faruque Hasan*

Artie McFerrin Department of Chemical Engineering, Texas A&M University
College Station, TX 77843-3122, USA.

Abstract

We present a new data-driven approach for both accurate and computationally efficient approximation of vapor liquid equilibria (VLE) models. Our method is able to provide guaranteed enclosure to limit the approximation errors over the entire domain of interest, all just by sampling only at select points. The approximation relies on a mixed-integer linear programming (MILP) formulation that exploits vertex polyhedral properties of theoretically guaranteed lower and upper bounds to enclose nonlinear and nonconvex equations of state (EOS) and empirical models. Another advantage is that, unlike traditional full simulation-based data-driven approaches, we do not solve nonlinear system of equations ($\mathbf{f}(\mathbf{x}) = 0$) for sampling. Instead of looking for only feasible samples, we evaluate $\mathbf{f}(\mathbf{x})$ over $\mathbf{x}$ -domain. This functional evaluation eliminates the need for computationally-demanding full-scale simulations and the associated convergence issues. We demonstrate excellent performance of the proposed MILP formulation in predicting the solubility of hydrofluorocarbon (HFC) refrigerants in ionic liquids (IL).

Keywords: Data-driven Modeling and Optimization, Surrogate Modeling, Global Optimization, Gamma-Phi Approach, Equation of State.

*Correspondence should be addressed to M.M. Faruque Hasan at hasan@tamu.edu, Phone: (979) 862-1449.

[†]These authors have contributed equally.

Introduction

Various models are used to describe thermodynamic phase equilibria that relate a system’s temperature and pressure with the volume, phase and composition of different chemical species that constitute the system at equilibrium conditions. A cubic equation of state (EOS) is an example of a thermodynamic model that describes the distribution of a chemical species under vapor liquid equilibria (VLE). Many processing operations, such as distillation, absorption, adsorption, phase separation and phase transition, that occur in chemical, petrochemical, pharmaceutical, energy-generating, gas processing, food and cosmetics industry are governed by phase equilibria.¹ Accurate estimation of physical properties and phase equilibrium data are important to ensure the safety, operability and economic margin of these chemical process systems.

Thermodynamic models are frequently used in computer-aided systematic process design, simulation, optimization and intensification activities. Therefore, it is imperative to develop accurate and, at the same time, computationally efficient models to represent the thermodynamic relations. With the advent of computing power and numerical algorithms, many commercial process simulators are now available that are routinely used in the chemical process industry. A key to the recent success of process models and computer codes as “virtual twins” can be attributed to the use of detailed property estimation packages, experimentally validated or fitted equation of states (e.g., Peng Robinson, Redlich–Kwong, etc.), and activity coefficient models (UNIFAC or NRTL in Gamma-Phi approach) to reliably estimate the process conditions and performance. While useful, these models come in various forms and often involve large systems of highly nonlinear equations, thereby requiring the use of sophisticated numerical simulation and/or global optimization techniques^{2,3}. Since the primary goal of these models is to mimic the physical reality, modelers tend to adhere to first-principles based rigorous mathematical descriptions that increase the model complexity. Incorporating them in decision-making often pose significant algorithmic and computational challenges. For example, a process simulator may not converge when simulating a large process flowsheet with one or more recycle loops. Similarly, a steady-state process design and global optimization problem may not converge to a globally optimal solution within a reasonable time limit. Dealing with dynamic problems is even more challenging as the full discretization⁴ in both temporal and spatial domains typically leads to large system of algebraic equations and degeneracy.

At least two major data-driven approaches have been found to be useful to tackle some of these challenges. The first is the model-free approach where a complex thermodynamic and property estimation model is treated as a blackbox to generate input-output sampling data

over the domains of interest. The selection of thermodynamic models, property estimation packages, numerical convergence criteria, etc., are all kept in the blackbox simulation environment, thereby maintaining the high-fidelity nature of the models. Based on the data provided by these models, an optimizer then performs design of computer experiments⁵, and derivative-free optimization based on direct search^{6,7}, blackbox optimization^{8,9}, trust region modeling^{10,11}, or logic-based hybrid optimization¹². Considerable advances have been made in finding feasible solutions (due to the adherence to the original models) for constrained problems^{13,14}. However, the convergence to optimal solutions is only guaranteed for dense sampling, thereby making the global optimization even a harder proposition. In many cases, solutions are obtained only after a long time if the simulations are computationally demanding and the overall strategy does not support parallel computing.

The second is the model-based approach that involves model reduction¹⁵ or surrogate modeling techniques^{16–23} that replace the original models. Although different surrogate models have different functional forms, they are fitted with sample/simulated data to mimic the original trends. This way, surrogate models are used to strike a balance between prediction accuracy and computational efficiency. The surrogate modeling approaches that exist in the literature can be divided into two main categories: Non-interpolating and interpolating. Examples of non-interpolating models include single/piecewise linear approximation, quadratic response surfaces, polynomial or generalized regression functions, and artificial neural networks (ANN). These methods may not reliably predict the response surface^{24,25} or sufficiently capture multi-dimensional and highly nonlinear correlations²⁴ due to entrapment to local regions or over fitting in the case of small data set. Interpolating methods such as Kriging²⁶ and radial basis functions (RBF)^{23,27,28} incorporate different basis functions to exactly predict the training points²⁹. Key advantages of interpolating methods include better handling of highly nonlinear functions²⁶ with global convergence³⁰ given that the surrogate model is probabilistically fully linear. Since surrogate models are often composed of simpler basis functions and expressions, generalized algorithms, methods and codes (e.g., ALAMO³¹, ARGONAUT²², NOMAD⁷, UNIPOPT³²) have been developed and used for different purposes.

Considerable work has been done on developing surrogate models for predicting phase and chemical equilibria. For example, Nentwich and Engell³³ applied adaptive sampling to develop surrogate models for simplifying phase equilibrium calculations. Zhu and Müller³⁴ combined ANN and Gaussian processes to approximate equation of states and predict thermodynamic properties. Carranza-Abaid et al.³⁵ deployed a hybrid approach that combines machine learning (ML) with thermodynamically consistent additional constraints to improve the prediction accuracy. Eason et al.³⁶ developed and applied

surrogate EOS for process optimization. These approaches have their own advantages and disadvantages. While surrogate models often improve the computational efficacy, there are cases where sampling-based surrogate models are shown to diverge from the original model predictions, as more data are included in the building phase of these surrogate models. Furthermore, surrogates may be good approximators but cannot always guarantee the type of approximation (true underestimation or overestimation). It is important to identify the type of approximation, for instance, for conservative (under)estimation of safety ratings for critical process operations. Any overestimation may lead to safety concerns. Similarly, overestimation of purity of pharmaceutical molecules may lead to health concerns. On the other hand, underestimation of costs and utilities may disrupt the development of a process technology. Therefore, it is important to provide both lower and upper bounds on the prediction errors when approximating critical properties.

The choice of surrogate models is often made based on the type of problems we seek to solve. Replacing rigorous thermodynamic models with a purely data-driven surrogate may not guarantee good prediction over the entire domain of interest. To avoid computationally expensive exploration of the entire domain, many approaches use expected values for the less explored regions. However, such statistical approach cannot provide guaranteed/deterministic bounds on the prediction error. Significant uncertainty (and hence, lack of confidence) in using surrogate or metamodels may originate from the model form and fitted parameters.³⁷ For the case of thermodynamics, both noninterpolating and interpolating surrogate models may lead to infeasible solutions due to inaccurate approximations. For example, nonideal liquid phase behavior deviates significantly from the linear solubility prediction based on Henry’s law. To circumvent this issue, one possible way is to train ANN based surrogate thermodynamic models on a large dataset. However, availability of the large dataset itself relies upon computationally expensive simulations. Additionally, ANN based models are not able to identify whether the prediction overestimates or underestimates the true values, which may potentially hinder its application to conservative cases (e.g., safety ratings). To that end, guaranteed underestimators and overestimators of thermodynamic models are highly desired so that the improved relaxation achieved from tighter lower and upper bounds leads to faster convergence in global optimization of process synthesis problems.

Deterministic global optimization techniques are appropriate for providing theoretically guaranteed bounds on prediction errors. Among the few notable works related to thermodynamic models, McDonald and Floudas² developed GLOPEQ which is a computational tool to locate global minima of the Gibbs free energy for finding solutions of thermodynamic phase and chemical equilibria. Schweidtmann et al.³⁸ developed an

approach that learns the thermodynamic equilibria of mixtures using ANNs with a guarantee on the worst-case performance of the surrogate, provided that the algebraic formulation of the underlying model is available. Keßler et al.³⁹ developed an approach for the global optimization of problems with multiple input-output multiplicities using explicit and implicit surrogate models. Jones et al.⁴⁰ proposed an efficient global optimization (EGO) algorithm for blackbox optimization. Schweidtmann et al.⁴¹ and Schweidtmann and Mitsos⁴² considered the deterministic global optimization of problems that have embedded ANN models in them. Boukouvala et al.⁴³ developed a global optimization method for general constrained grey-box problems that contain a combination of black-box constraints and constraints with known functional forms. These global optimization approaches require gradient information over the entire domain, which may not always be available. Because of this reason, majority of the global optimization works focus on already developed surrogate models without paying much attention to approximation errors.

Purely data-driven under- and overestimation of blackbox models is challenging. Rebennack and Kallrath⁴⁴ proposed a mixed-integer linear programming (MILP) formulation to obtain data-driven piecewise linear under- and overestimators for univariate and bivariate functions. However, these estimators are guaranteed only at the simulated points, but not over the entire domain. Recent works, such as by Bajaj and Hasan⁴⁵ and Song et al⁴⁶, indicate that theoretically guaranteed affine underestimators and lower bounds of certain classes of models can be tractably obtained just by black-box sampling. Specifically, Bajaj and Hasan⁴⁵ developed a new method to deterministically find global solutions to problems where only available information includes the global upper bound on the diagonal Hessian elements. An edge-concave underestimator⁴⁷ can be constructed with vertex polyhedral solution, thereby leading to affine underestimators of the original black-box problem solely based on simulation at the bound vertices. When the full model is known, one can estimate theoretically guaranteed upper bounds on the diagonal Hessian elements using analytical methods or using automatic differentiation in conjunction with interval analysis (in this case, one can also learn an ANN model with worst-case error bounds³⁸). However, for black-box models where only the input-output data are available, one can estimate near-exact (albeit somewhat loose) upper bound from physical intuition and interpretation based on domain knowledge (e.g., bounds on flowrates, mass fractions, reaction rates, etc.). Since it is possible to derive the bounds on the diagonal Hessian elements for known thermodynamic models, the edge concave relaxation provides an attractive way towards developing surrogate/approximate thermodynamic models with theoretically guaranteed bounds on the prediction errors. As elaborated in the subsequent sections, this is, in fact, one of the key ideas of the current work.

In this work, we present a new data-driven approach for both accurate and computationally favorable (cheap to evaluate) approximation of VLE models. Based on functional evaluation or data-sampling of EOS models, we construct edge-concave underestimators and edge-convex overestimators that get tighter as the number of evaluation points increases. These estimators provide an enclosure to the original model from both below and top over the entire domain of interest. Both the under- and overestimators require minimal information. The requirement includes the existence of continuity and twice differentiability of the thermodynamic model with bounded Hessian. Our method is able to provide guaranteed bounds and improve the maximum approximation errors, all just by sampling only at select points. The approximation relies on a MILP formulation that exploits vertex polyhedral properties of theoretically guaranteed lower and upper bounds on the thermodynamic models. Another advantage is that, unlike traditional full simulation-based data-driven approaches, we do not solve nonlinear system of equations ($\mathbf{f}(\mathbf{x}) = 0$) for sampling. Instead of sampling the feasible solutions for ($\mathbf{f}(\mathbf{x}) = 0$), we sample the function $\mathbf{f}(\mathbf{x})$ itself and estimate $\mathbf{x}$ that would satisfy $\mathbf{f}(\mathbf{x}) = 0$. This eliminates the need for computationally-demanding full-scale simulations and avoids associated convergence issues.

This paper is organized as follows: First, we introduce the overall method with theoretical details and a description of the MILP model formulation. Then, we present a case study on predicting the solubility of Hydrofluorocarbon (HFC) in ionic liquid (IL) systems. We illustrate excellent solubility prediction capability of our MILP based approximation model as compared to the original Gamma-Phi approach based thermodynamic model. Finally, we provide some concluding remarks.

Data-driven Bounded Approximation

Consider a true simulation model:

$$f(\mathbf{x}) = 0 \tag{1}$$

where, $\mathbf{x} \in S = \{\mathbf{x}^l \leq \mathbf{x} \leq \mathbf{x}^u\} \subseteq \mathbb{R}^n$, and $f(\mathbf{x}) : S \rightarrow \mathbb{R}$ is a continuous and twice differentiable deterministic function representing physical system behavior (examples include thermodynamic VLE model, property estimation model, conservation laws, etc). We do not need to know the exact form of $f(\mathbf{x})$ but information needs to be available about the upper bounds on the diagonal elements of the Hessian of $f(\mathbf{x})$. Our goal is to find an approximate solution, $\mathbf{x}^*$, such that $f(\mathbf{x}^*) \rightarrow 0$. We do this only using functional evaluations of $f(\mathbf{x})$ at different points over the domain. We postulate that $\mathbf{x}^*$ can be obtained by solving the following optimization problem:

$$\min_{\mathbf{x} \in S} \left(U(\mathbf{x}) - L(\mathbf{x}) \right) \quad (2)$$

s.t.

$$L(\mathbf{x}) \leq 0 \quad (3)$$

$$U(\mathbf{x}) \geq 0 \quad (4)$$

where, $L(\mathbf{x})$ and $U(\mathbf{x})$ are guaranteed the underestimator and overestimator of $f(\mathbf{x})$ respectively. For good approximation, the key is to obtain tight $L(\mathbf{x})$ and $U(\mathbf{x})$. As we increase the tightness of both estimators, $L(\mathbf{x}) \rightarrow f(\mathbf{x})$ and $U(\mathbf{x}) \rightarrow f(\mathbf{x})$, thereby increasing the quality of approximation as $\mathbf{x}^* \rightarrow \mathbf{x}$. In the subsequent subsections, we present the forms of these estimators and their linear facets leading to tractable formulation and solution of the above optimization problem.

Model Enclosure via Underestimation and Overestimation

There are several candidate estimators that have been developed in the literature. However, most of them require gradient information over the entire domain. To that end, we adopt the edge-concave underestimator^{45,47} because of its special vertex polyhedral property⁴⁸ that allows us to construct linear facets of the convex envelope solely based on evaluation of $f(\mathbf{x})$ at the domain bounds and interior points.

The mathematical form of the edge-concave underestimator, $L(\mathbf{x})$, for a twice-differentiable non-convex function $f(\mathbf{x})$ is given by

$$L(\mathbf{x}) = f(\mathbf{x}) - \sum_{i=1}^n \theta_i^L (x_i - x_i^{Int})^2, \quad (5)$$

where x_i^{Int} is a fixed point and the parameter θ_i^L is defined as

$$\theta_i^L = \max \left\{ 0, \frac{1}{2} \left[\frac{\partial^2 f}{\partial x_i^2} \right]^U \right\} \quad (6)$$

Similar to the edge-concave underestimator, we propose an edge-convex overestimator, $U(\mathbf{x})$, as follows:

$$U(\mathbf{x}) = f(\mathbf{x}) + \sum_{i=1}^n \theta_i^U (x_i - x_i^{Int})^2 \quad (7)$$

$$\theta_i^U = \max \left\{ 0, \frac{1}{2} \left[\frac{-\partial^2 f}{\partial x_i^2} \right]^U \right\} \quad (8)$$

Since $\theta_i^L \geq 0$ and $\theta_i^U \geq 0$, $L(\mathbf{x})$ is a guaranteed underestimator and $U(\mathbf{x})$ is a guaranteed overestimator of $f(\mathbf{x})$ as the summation term will always be non-negative.

$L(\mathbf{x})$ is edge-concave where all nonedge-concavities are overpowered by the summation term in Eq. 5. Similarly, $U(\mathbf{x})$ is edge-convex where all nonedge-convexities are overpowered by the summation term in Eq. 7.

$L(\mathbf{x}^{Int}) = f(\mathbf{x}^{Int})$ at the sample points x^{Int} , irrespective of the convexity/edge-concavity of $f(\mathbf{x})$. Similarly, $U(\mathbf{x}^{Int}) = f(\mathbf{x}^{Int})$ at the sample points x^{Int} .

$L(\mathbf{x})$ and $U(\mathbf{x})$ themselves are nonlinear and care must be taken to use the estimators in their current form. Since we know the value of $L(\mathbf{x})$ and $U(\mathbf{x})$ at the domain bounds and the simulated points only, it is possible to generate linear facets (affine relaxations of the estimators themselves). Consider the following one-dimensional function

$$g(x) = x^3 - 4x^2 - 5x + 16, \quad x \in [-4, 6] \quad (9)$$

The upper bound on the Hessian for this function over the given domain is 28. Therefore, $\theta^L = 14$. The original function and its edge-concave underestimator generated at $x^{Int} = 1$ along with its linear facets that underestimate $L(\mathbf{x})$ are shown in Figure 1a. Similarly, one can generate an edge-convex overestimator at $x^{Int} = 1$. It is possible to generate a unique underestimator for each and every sample point. Multiple underestimators generated at sample points $x^{Int} = \{-2, 1, 4\}$ and their linear facets are shown in Figure 1b.

Let J be the set of interior sampling points. For an n -dimensional problem, for each point $j \in J$, one can obtain an underestimator, the linear facets of which result in $|J|$ number of $n+1$ dimensional simplices (polytopes) each having $2^n + 1$ vertices (2^n vertices at the domain bounds and one interior vertex at the sample point).

[Figure 1 about here]

Piecewise Linear Bounding

Let A_j be the set of vertices at domain bounds of polytope $j \in J$ where each set A_j contains 2^n vertices. For example, given a 2-D problem with variables x_1 and x_2 , for every polytope j , we will have $2^2 = 4$ vertices at the domain bounds, one each at

$L(x_1^L, x_2^L), L(x_1^L, x_2^U), L(x_1^U, x_2^L), L(x_1^U, x_2^U)$. Since the number of polytopes generated is equal to the number of interior points, we define another singleton set B_j for every j which contains the interior (sampled) point, that is, the interior vertex of polytope j given by $L(x_1^{Int}, x_2^{Int}) = f(x_1^{Int}, x_2^{Int})$. Therefore, the set of vertices of polytope j is the union of both sets ($V_j = A_j \cup B_j$) whose final element corresponds to the interior point.

One can obtain a data-driven underestimation by finding the polytope with the closest underestimating linear facet to the original function. This can be achieved by solving the following MILP problem (P-L):

$$P - L : \max \sum_{j \in J} l_j \quad (10)$$

s.t.

$$\sum_{j \in J} z_j = 1 \quad (11)$$

$$x_i = \sum_{v \in V_j} \lambda_{v,j} \hat{x}_{j,i,v} \quad i \in n, j \in J \quad (12)$$

$$\sum_{v \in V_j} \lambda_{v,j} = 1 \quad j \in J \quad (13)$$

$$0 \leq \lambda_{v,j} \leq 1 \quad v \in V_j, j \in J \quad (14)$$

$$lf_j = \sum_{v \in V_j} \lambda_{v,j} L(\hat{x}_{v,j}) \quad j \in J \quad (15)$$

$$l_j \geq lf_j^L z_j \quad j \in J \quad (16)$$

$$l_j \geq lf_j^U z_j + lf_j - lf_j^U \quad j \in J \quad (17)$$

$$l_j \leq lf_j^L z_j + lf_j - lf_j^L \quad j \in J \quad (18)$$

$$l_j \leq lf_j^U z_j \quad j \in J \quad (19)$$

$$\sum_{v \in V_j} r_{v,j} = |n| + 1 \quad j \in J \quad (20)$$

$$\lambda_{v,j} \leq r_{v,j} \quad v \in V_j, j \in J \quad (21)$$

$$r_{v,j} = 1 \quad v \in B_j, j \in J \quad (22)$$

$$\sum_{v \in LDV_j} r_{v,j} \leq |n| \quad j \in J \quad (23)$$

where,

$$z_j = \begin{cases} 1 & \text{if polytope } j \text{ is selected that contains the closest linear facet with the} \\ & \text{tightest underestimation of } f(x), \\ 0 & \text{otherwise} \end{cases}$$

and

$$r_{v,j} = \begin{cases} 1 & \text{if vertex } v \text{ is selected for convex combination in polytope } j, \\ 0 & \text{otherwise.} \end{cases}$$

Here, $l_j = lf_j z_j$. This bilinear expression is exactly linearized as shown in Eqs. 16, 17, 18, 19. $lf_j^L = \min_{v \in V_j} [L(\hat{x}_{v,j})]$ and $lf_j^U = \max_{v \in V_j} [L(\hat{x}_{v,j})]$ and $L(\hat{x}_{v,j})$ is the value of the edge-concave underestimator at vertex v of polytope j . Eq. 11 means that the closest linear facet should lie on only one of the polytopes. In this formulation, all x_i values are known, that is they are the input parameters where we wish to evaluate the tightest underestimated value of $f(x)$. Also, $\hat{x}_{j,i,v}$ is the value of variable i at vertex v of polytope j . The key variable is $\lambda_{v,j}$ which is known as the convex combination variables that refers to the "weight" given to vertex v of polytope j . To ensure that the convex combination results in a point that lies in the interior of the vertex, we bound $\lambda_{v,j}$ to be between 0 and 1. Eqs. 12, 13, 14, refer to the convex combination formulation where any point in the domain can be expressed as a linear combination of the vertices of polytope j . lf_j is the value of the linear facet of polytope j at the point of interest given by Eq. 15.

For an n -dimensional problem, the simplex generated is $(n + 1)$ dimensional. Hence, $n + 1$ linearly independent points are sufficient to determine a unique value in $(n + 1)$ dimensional space given by Eqs. 20, 21. It is important to include the interior vertex for convex combination as there may exist facets not including the interior point which have a higher value than those facets including the interior point itself resulting in unwanted solutions. Eq. 22 is included in the model for this purpose. To that effect, we add cuts to the model where LDV_j is the set of linearly dependent vertices in polytope j and such vertices must not be chosen at the same time for convex combination given by Eq. 23. The closest linear facets of the function given by Eq. 9 generated using five interior (simulation)

points at $x = \{-4, -2, 1, 4, 6\}$ are shown in Figure 2. The maximum separation distance of the closest linear underestimating facets from the original function reduces with an increasing number of sample points as shown in Table 1.

[Figure 2 about here]

[Figure 3 about here]

[Table 1 about here]

Similarly, one can obtain a data-driven overestimation by finding the polytope with the closest overestimating linear facet to the original function. This can be achieved by solving the following MILP problem (P-U):

$$P - U : \min \sum_{j \in J} u_j \quad (24)$$

s.t. Eqs. 11 - 14

$$uf_j = \sum_{v \in V_j} \lambda_{v,j} U(\hat{x}_{v,j}) \quad j \in J \quad (25)$$

$$u_j \geq uf_j^L z_j \quad j \in J \quad (26)$$

$$u_j \geq uf_j^U z_j + uf_j - uf_j^U \quad j \in J \quad (27)$$

$$u_j \leq uf_j^L z_j + uf_j - uf_j^L \quad j \in J \quad (28)$$

$$u_j \leq uf_j^U z_j \quad j \in J \quad (29)$$

Eqs. 20 - 23

where,

$$z_j = \begin{cases} 1 & \text{if polytope } j \text{ is selected that contains the closest linear facet with the} \\ & \text{tightest overestimation of } f(x), \\ 0 & \text{otherwise} \end{cases}$$

and

$$r_{v,j} = \begin{cases} 1 & \text{if vertex } v \text{ is selected for convex combination in polytope } j, \\ 0 & \text{otherwise.} \end{cases}$$

Here, $u_j = uf_j z_j$. This bilinear expression is exactly linearized as shown in Eqs. 26, 27, 28, 29. $uf_j^L = \min_{v \in V_j} [U(\hat{x}_{v,j})]$ and $uf_j^U = \max_{v \in V_j} [U(\hat{x}_{v,j})]$ and $U(\hat{x}_{v,j})$ is the value of the edge-convex overestimator at vertex v of polytope j .

The closest underestimating and overestimating facets of Eq. 9 are shown in Figure 3.

Remark 1: In the context of global optimization, in general, if one is interested in relaxation problems to find a lower bound and upper bound on $f(x)$ over $x^L \leq x \leq x^U$, it is possible to do so using the MILP formulations $P-L$ and $P-U$ respectively. The quality of the lower bound and upper bounds increases with increasing number of simulation (interior) points as more interior points result in tighter facets.

MILP-based Approximate Solution to $f(\mathbf{x}) = 0$

Now that we can generate linear facets for guaranteed underestimators and overestimators, given some sampled points of a function $f(\mathbf{x})$, it is possible to find an approximate solution to the problem $f(\mathbf{x})=0$. For this purpose, we do not require gradient-based information over the entire domain, but rather the simulated data points and θ_i values are sufficient.

In formulations $P-L$ and $P-U$, our objective was to find the closest linear facet of $L(\mathbf{x})$ and $U(\mathbf{x})$ given any point in the domain, i.e., all x_i in Eq. 12 were treated as parameters, since we decided their values. In this section, since we plan to approximate the values of those free x_i at which $f(x) = 0$, x_i are now free variables. With this, we present an MILP formulation (P-S) to find approximate solution of $f(\mathbf{x}) = 0$ as follows:

$$P-S : \min \sum_{j=1}^J (u_j - l_j) \quad (30)$$

s.t. Eqs. 11 - 23

Eqs. 25 - 29

$$lf_j \leq 0 \quad j \in J \quad (31)$$

$$uf_j \geq 0 \quad j \in J \quad (32)$$

According to Property , since $(u_j - l_j) = 0$ at the simulated points, the formulation results in the value of x_i where the value of $(u_j - l_j)$ attains a minimum subject to the constraints Eqs. 31 and 32. That is, we get a value of x_i at one of the simulated points which is closest to the real value of x_i where $f(x) = 0$. The feasible regions pertaining to these equations are shown in Figure 4. It can be seen that the minimum indeed occurs an interior point closest to the real root.

[Figure 4 about here]

Remark 2.3.1: If one is interested only in the approximation of f , then it is a simpler problem than what we presented in this section as we are fixing all x_i . For the fixed x_i , this results in a range of values, where the real value of $f(x)$ could be. This range decreases with increasing number of interior points as the linear facets of both the underestimator and overestimator converge to the real function itself.

This completes the theoretical background and the development of MILP-based approximation with guaranteed enclosure.

Prediction of HFC Solubility in Ionic Liquids

Hydrofluorocarbons (HFCs) are commonly used as refrigerants in domestic and commercial cooling systems. These compounds account for 2–3% of the global greenhouse gas (GHG) emissions⁴⁹ which prompted the Kigali Amendment to the Montreal Protocol to recommend cutting the global HFC emissions by 80–85% by 2047.⁵⁰ Therefore, it is desirable for these HFCs to be recycled so that the amount of newly manufactured HFC molecules can be minimized. However, recycling through separation of HFCs pose a significant challenge as these are azeotropic or close-boiling mixtures that often behave as single fluids. Consequently, conventional separation techniques, such as cryogenic distillation, when used for HFC separation are not a viable option as they are very energy and cost-intensive and pose significant operational challenges.⁵⁰ To that end, ILs have garnered significant attention as potential solvents for extractive distillation based separation of these azeotropic mixtures. ILs can significantly improve the efficiency of absorption refrigeration processes. In addition, negligible vapor pressure of ILs prevent contamination of the refrigerant gas with solvent.⁵¹

To predict the amount of refrigerant absorbed in IL at a given temperature and pressure, an accurate solubility model is required. Significant amount of work has been done to model the solubility of HFC in IL, ranging from simple Henry’s law to more rigorous/empirical Gamma-Phi and Equation of state (EOS) models.^{52–54} Henry’s law is computationally inexpensive and provides a linear surrogate to solubility prediction at dilute conditions. However, it is not applicable to systems that operate at high pressures and/or concentrated solutions. To facilitate this, Gamma-Phi or EOS models can be used for more accurate solubility prediction. However, the exponential terms and different form of mixing rules make the model nonconvex and highly nonlinear. Employing these models to an optimization framework for process synthesis purpose make it computationally expensive. Therefore, simpler surrogate models that are able to reliably predict the solubility, and at the same

time can be employed in superstructure based optimization of process synthesis problems, are of paramount importance in order to provide guaranteed optimal process flowsheet for the separation of HFC using ILs within tractable solution time.

[Figure 5 about here]

Recently, we have proposed a process synthesis, optimization, and intensification framework named SPICE⁵⁵ based on building block-based process representation.^{56,57} Detailed description on buildings block-based representation method can be found elsewhere^{58–60}. Here, we briefly describe on how the VLE separation is represented and modeled in building block. Figure 5a, depicts a conventional two phase/flash separator, where the inlet stream is in liquid phase. After entering the separator, vapor and liquid phases separate depending on the phase equilibrium. The vapor phase leaves the separator from the top and the liquid phase leaves from the bottom. For representing two phases in building block, two blocks are required, one for each phase. The phase boundary or the contact between two phases is represented by a common semi-restricted boundary. The material flow between the phases takes place through this boundary and the rate of mass transfer is determined by the VLE. Figure 5b represents such kind of VLE separation in building block. Here, the block is right hand side is in liquid phase and the block in left hand side is in vapor phase. The phase contact is represented by the dotted blue vertical line. The inlet feed stream enters the VLE separator in the liquid block (right hand side block). After separation, the vapor products leaves the separator from the left hand side block and the liquid product leaves from the right hand side block. In similar fashion, any two phase separation can be represented. For instance, by using two blocks we can represent a single tray of a distillation column. To represent the whole distillation column, we need two rows of blocks arranged in series (refer to Demirel et al.⁵⁸ for details).

Calculating θ

Here, we first derive the diagonal elements of the Hessian of the HFC/IL thermodynamic model (see Appendix A). We then use a nonlinear program (NLP) to find the upper bounds of these Hessian elements.

The Hessian, $\mathbf{H}_f$ of the function $f(P, T, \tilde{x}_1)$ described in Eq. 43 is as follows:

$$\mathbf{H}_f = \begin{bmatrix} \frac{\partial^2 f}{\partial P^2} & \frac{\partial^2 f}{\partial P \partial T} & \frac{\partial^2 f}{\partial P \partial \tilde{x}_1} \\ \frac{\partial^2 f}{\partial T \partial P} & \frac{\partial^2 f}{\partial T^2} & \frac{\partial^2 f}{\partial T \partial \tilde{x}_1} \\ \frac{\partial^2 f}{\partial \tilde{x}_1 \partial P} & \frac{\partial^2 f}{\partial \tilde{x}_1 \partial T} & \frac{\partial^2 f}{\partial \tilde{x}_1^2} \end{bmatrix} \quad (33)$$

The principal diagonal elements of $\mathbf{H}_f$ comprise the vector, $\mathbf{h}_f$ as follows:

$$\mathbf{h}_f = \left[\frac{\partial^2 f}{\partial P^2} \quad \frac{\partial^2 f}{\partial T^2} \quad \frac{\partial^2 f}{\partial \tilde{x}_1^2} \right]^T \quad (34)$$

where,

$$\frac{\partial^2 f}{\partial P^2} = -\frac{1}{P^2} \quad (35)$$

$$\frac{\partial^2 f}{\partial \tilde{x}_1^2} = \frac{1}{\tilde{x}_1^2} - \left[\frac{1}{\gamma_1} \frac{\partial^2 \gamma_1}{\partial \tilde{x}_1^2} + \frac{\partial \gamma_1}{\partial \tilde{x}_1} \left(-\frac{1}{\gamma_1^2} \frac{\partial \gamma_1}{\partial \tilde{x}_1} \right) \right] \quad (36)$$

$$\begin{aligned} \frac{\partial^2 f}{\partial T^2} = & \frac{1}{R} \left[\left(\frac{\partial B_1}{\partial T} - \frac{\partial \tilde{V}_1}{\partial T} \right) \left\{ \frac{T \left(-\frac{\partial P_1^s}{\partial T} \right) - (P - P_1^s)}{T^2} \right\} \right] \\ & + \frac{1}{R} \left\{ \left(\frac{P - P_1^s}{T} \right) \left(\frac{\partial^2 B_1}{\partial T^2} - \frac{\partial^2 \tilde{V}_1}{\partial T^2} \right) + \frac{(B_1 - \tilde{V}_1)}{T} \left(-\frac{\partial^2 P_1^s}{\partial T^2} \right) \right\} \\ & + \frac{1}{R} \left[\left\{ T \left(-\frac{\partial P_1^s}{\partial T} \right) - (P - P_1^s) \right\} \left\{ \frac{T \left(\frac{\partial B_1}{\partial T} - \frac{\partial \tilde{V}_1}{\partial T} \right) - 2(B_1 - \tilde{V}_1)}{T^3} \right\} \right] \\ & - \frac{1}{P_1^s} \frac{\partial^2 P_1^s}{\partial T^2} - \frac{\partial P_1^s}{\partial T} \left(-\frac{1}{P_1^{s^2}} \frac{\partial P_1^s}{\partial T} \right) - \frac{1}{\gamma_1} \frac{\partial^2 \gamma_1}{\partial T^2} - \frac{\partial \gamma_1}{\partial T} \left(-\frac{1}{\gamma_1^2} \frac{\partial \gamma_1}{\partial T} \right) \end{aligned} \quad (37)$$

Now that we have the diagonal elements of the Hessian, we use Eqs. 6 and 8 to find the θ^L and θ^U values over the entire domain $P^L \leq P \leq P^U$, $T^L \leq T \leq T^U$, and $\tilde{x}_1^L \leq \tilde{x}_1 \leq \tilde{x}_1^U$.

For the case of HFC/IL system we set $P^L = 0.01$ MPa, $P^U = 1$ MPa, $T^L = 280$ K, $T^U = 375$ K, $\tilde{x}_1^L = 0.01$, $\tilde{x}_1^U = 1$. The obtained θ values are listed in Table 2. Interestingly, θ_P values are zero for both NRTL and Margules-based activity coefficient models considered in this work. Although all θ_T values are not zero, most of them take near-zero values. These indicate that the function f (Eq. 43) is already edge-concave and near-edge-concave in terms of pressure (P) and temperature (T), respectively. That is, the functions are pointwise concave and a linear surrogate model is not a good choice but rather, we obtain the tightest linear facets of their edge-concave underestimators by exploiting the vertex polyhedral properties of edge-concavity. This justifies the use of the edge-concave underestimator (and edge-convex overestimator) for thermodynamic models, since these estimators nearly match with the original function over the entire domain of P and T for fixed $\tilde{x}_1$, thereby minimizing the approximation error.

[Table 2 about here]

Results

We demonstrate the benefit of the developed surrogate modeling technique in a process synthesis platform by implementing it in SPICE. To predict the solubility of HFC in IL ([bmim][PF₆]), a single tray of an extractive distillation column is simulated. First, the solubility is determined using the Gamma-Phi approach. Afterwards, the MILP based surrogate model is used to predict the solubility. For each of the simulation, it is considered that the inlet flow rate of HFC is 1 mol/s and the inlet flow rate of IL is 0.1 mol/s. While using the Gamma-Phi method based thermodynamic model, a dummy objective function is used for solubility prediction, because all the variables are pre-defined beforehand. On the other hand, when the MILP surrogate model is used, Eq. 30 is considered as the objective function. We solve each of the reported simulation with ANTIGONE v1.1⁶¹ in GAMS v35.1.0 environment to global optimality.

Solubility of R-134a in [bmim][PF₆]

The solubility of R-134a is estimated at four different (P, T) conditions. The solubility prediction by the Margules model is presented in Table 3. For the prediction of solubility using the MILP based surrogate model, four sets of interior points, i.e., 11, 21, 51, and 101 are considered. With the increasing number of interior points, the number of variables also increase. For instance, when there are 11 interior points, the model has 166 continuous variables, 99 binary variables and 311 equations. On the other hand, when 101 interior points are considered, the model has 1516 continuous variables, 909 binary variables, and 2741 equations. However, with the increasing number of interior points, the solubility prediction becomes more accurate. For instance, at 298.1 K and 0.35 MPa, the R-134a liquid phase mol% prediction with 11 interior points has a prediction error of 9.7%. When the interior points are increased to 101, the prediction error is reduced to only -1.1%. Similar trend is observed at other (P, T) conditions as well. For most cases, the prediction accuracy of the surrogate model irrespective of the number of interior points is within acceptable range ($< 5\%$).

[Table 3 about here]

The solubility prediction with the NRTL model is presented in Table 4. As the NRTL based activity coefficient model is more rigorous (hence, more complex) compared to the Margules model, the solubility prediction with the Gamma-Phi approach is very close to the experimental solubility data.⁵² However, even with the detailed nature of the NRTL based model, the solubility prediction with our MILP based surrogate model is accurate. For some

cases with 11 interior points, the prediction error is large, which reduces when more interior points are used. For instance, at 323.2 K and 0.349 MPa, the prediction error with 11 interior points is 35.5%, which reduces to only 0.7% with 101 interior points.

[Table 4 about here]

When generating isotherms at four different temperatures (283.1 K, 298.1 K, 323.2 K, 348.2 K), 101 interior points are used since the solubility prediction with 101 interior points provides higher prediction accuracy. Figure 6a and Figure 6b show the solubility isotherms of R-134a determined with the Margules and NRTL activity coefficient models, respectively. For comparison purposes, in the same plot, the prediction with the MILP based surrogate model (circle symbols) are also shown. It is observed that the prediction using MILP based surrogate is very close to the solubility determined with the Gamma-Phi approach (solid lines) under both activity coefficient models.

[Figure 6 about here]

Solubility of R-32 in [bmim][PF₆]

Table 5 reports the solubility prediction of R-32 in [bmim][PF₆] with Margules activity coefficient model. To the best of our knowledge, the Margules binary interaction parameters for R-32/[bmim][PF₆] are not reported in the literature. Therefore, the parameters are estimated from the experimental solubility data.⁵² With these estimated parameters, the solubility prediction with the Gamma-Phi approach is found to be very close to the experimental solubility data as shown in Table 5. Similar to R-134a/[bmim][PF₆], the solubility prediction using the MILP-based surrogate model also improves for R-32/[bmim][PF₆] system with the increasing number of interior points. For example, at 323.1 k and 0.6996 MPa, the prediction error reduces from -8.5% to -1.4% as the number of interior points is increased from 11 to 101. It is interesting to note that, at lower temperatures, using only 11 interior points results in less than 1% prediction error.

[Table 5 about here]

The predicted solubility of R-32 in [bmim][PF₆] with NRTL activity coefficient model is presented in Table 6. Similar to the analysis presented above for the Margules activity coefficient model, the prediction becomes more accurate with the increasing number of interior points. Figure 7a and Figure 7b presents the solubility isotherms of R-32 in [bmim][PF₆] at four different temperatures with Margules and NRTL activity coefficient

model respectively. The solubility prediction with the MILP based surrogate model with 101 interior points (circle symbols) shows excellent agreement with the solubility estimated with the Gamma-Phi based model (solid lines). The accuracy of the P-S model depends on the number of data points used. However, increasing the number of data points would increase the size and complexity of the P-S model, thereby requiring more time to solve it. To study this, we further solve the P-S model for a varying number of sample points and record the CPU times. As shown in Figure 8., the computational time grows exponentially with number of samples. The $P - S$ formulation was implemented in GAMS 28.2.0 in an Intel[®] Core i7-4790 CPU at 3.60 GHz running Linux. This points to the fact that one needs to strike a balance between the accuracy and the computational burden when selecting the number of samples.

[Table 6 about here]

[Figure 7 about here]

[Figure 8 about here]

Conclusions

The data-driven approach presented in this work is not only applicable for the approximation of vapor liquid equilibria (VLE) models, but it also shows a promising pathway for solving general data-driven global optimization problems. We observed that the quality of approximation depends on several factors including the number and selection of interior points. Future work may include the optimal choice of these interior points in conjunction with other under- and overestimation types, such as α -BB⁶² and general McCormick relaxation⁶³, among others. The MILP formulation depends on the evaluation at the bound vertices, which may potentially limit the applicability of the current formulation to problems with large number of variables. Further work is needed to be able to reduce the number of needed evaluation points. However, as noted in this work, thermodynamic problems typically involve fewer variables which make the proposed method an attractive option for efficient approximation of thermodynamic properties. This is highlighted by the quality performance of the current MILP formulation in predicting the solubility of HFC refrigerants in ionic liquids. To that end, one might also consider replacing our MILP-based $P - S$ formulation using other MILP- or NLP-based ANN models³⁸. However, in the context of process synthesis, both would result in MINLP formulations due to the presence of systems-level discrete decisions related to the selection of unit operations and processing pathways.

Even when there are no discrete decisions involved, it is difficult to have a generalized conclusion on which of the approaches (the current MILP-based $P - S$ formulation or the ANN-based approximation) would always perform better. We anticipate that our approach will enable a new paradigm of derivative-free optimization-based process simulation and process optimization based on global optimization concepts. Lastly, it could be an efficient way for accelerating computationally demanding process simulations employing complex thermodynamic models, and further provides a means for confirming the global convergence of simulation data-driven process optimization problems.

Acknowledgement

The authors gratefully acknowledge support from the NSF EFRI DChEM (2029354), NSF CAREER (CBET-1943479), and the DOE AICHE RAPID Institute (DE-EE0007888-09-03) grants. Part of the research was conducted with the computing resources provided by Texas A&M High Performance Research Computing.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

- [1] Gmehling J, Kleiber M, Kolbe B, Rarey J. *Chemical thermodynamics for process simulation*. John Wiley & Sons. 2019.
- [2] McDonald CM, Floudas CA. GLOPEQ: A new computational tool for the phase and chemical equilibrium problem. *Computers & Chemical Engineering*. 1997;21(1):1–23.
URL <https://www.sciencedirect.com/science/article/pii/0098135495002502>
- [3] Mitsos A, Bolas GM, Barton PI. Bilevel optimization formulation for parameter estimation in liquid–liquid phase equilibrium problems. *Chemical Engineering Science*. 2009;64(3):548–559.
URL <https://www.sciencedirect.com/science/article/pii/S0009250908005113>
- [4] Wächter A, Biegler LT. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical Programming*. 2006;106(1):25–57.
URL <https://doi.org/10.1007/s10107-004-0559-y>
- [5] Garud SS, Karimi IA, Kraft M. Design of computer experiments: A review. *Computers & Chemical Engineering*. 2017; 106:71–95.
- [6] Rios LM, Sahinidis NV. Derivative-free optimization: a review of algorithms and comparison of software implementations. *Journal of Global Optimization*. 2013;56(3):1247–1293.
- [7] Le Digabel S. Algorithm 909: NOMAD: Nonlinear optimization with the MADS algorithm. *ACM Transactions on Mathematical Software (TOMS)*. 2011;37(4):1–15.

- [8] Bajaj I, Arora A, Hasan MF. Black-Box Optimization: Methods and Applications. In: *Black Box Optimization, Machine Learning, and No-Free Lunch Theorems*, pp. 35–65. Springer. 2021;.
- [9] Javaloyes-Antón J, Ruiz-Femenia R, Caballero JA. Rigorous design of complex distillation columns using process simulators and the particle swarm optimization algorithm. *Industrial & Engineering Chemistry Research*. 2013;52(44):15621–15634.
- [10] Conn AR, Scheinberg K, Vicente LN. Global convergence of general derivative-free trust-region algorithms to first-and second-order critical points. *SIAM Journal on Optimization*. 2009;20(1):387–415.
- [11] Eason JP, Biegler LT. A trust region filter method for glass box/black box optimization. *AIChE Journal*. 2016; 62(9):3124–3136.
- [12] Caballero JA. Logic hybrid simulation-optimization algorithm for distillation design. *Computers & Chemical Engineering*. 2015;72:284–299.
- [13] Bunin GA, François G, Bonvin D. Feasible-side global convergence in experimental optimization. *arXiv preprint arXiv:1406.4063*. 2014;.
- [14] Bajaj I, Iyer SS, Hasan MF. A trust region-based two phase algorithm for constrained black-box and grey-box optimization with infeasible initial point. *Computers & Chemical Engineering*. 2018;116:306–321.
- [15] Agarwal A, Biegler LT. A trust-region framework for constrained optimization using reduced order modeling. *Optimization and Engineering*. 2013;14(1):3–35.
- [16] Caballero JA, Grossmann IE. An algorithm for the use of surrogate models in modular flowsheet optimization. *AIChE Journal*. 2008;54(10):2633–2650.
URL <https://aiche.onlinelibrary.wiley.com/doi/abs/10.1002/aic.11579>
- [17] Henao CA, Maravelias CT. Surrogate-based superstructure optimization framework. *AIChE Journal*. 2011; 57(5):1216–1232.
- [18] Fahmi I, Cremaschi S. Process synthesis of biodiesel production plant using artificial neural networks as the surrogate models. *Computers & Chemical Engineering*. 2012;46:105–123.
- [19] Cozad A, Sahinidis NV, Miller DC. Learning surrogate models for simulation-based optimization. *AIChE Journal*. 2014; 60(6):2211–2227.
- [20] Eason J, Cremaschi S. Adaptive sequential sampling for surrogate model generation with artificial neural networks. *Computers & Chemical Engineering*. 2014;68:220–232.
- [21] Boukouvala F, Hasan MF, Floudas CA. Global optimization of general constrained grey-box models: new method and its application to constrained PDEs for pressure swing adsorption. *Journal of Global Optimization*. 2017;67(1-2):3–42.
- [22] Boukouvala F, Floudas CA. ARGONAUT: AlgoRithms for Global Optimization of coNstrAined grey-box compUTational problems. *Optimization Letters*. 2017;11(5):895–913.
URL <https://doi.org/10.1007/s11590-016-1028-2>
- [23] Bhosekar A, Ierapetritou M. Advances in surrogate based modeling, feasibility analysis, and optimization: A review. *Computers & Chemical Engineering*. 2018;108:250–267.
URL <https://www.sciencedirect.com/science/article/pii/S0098135417303228>
- [24] Jones DR. A Taxonomy of Global Optimization Methods Based on Response Surfaces. *Journal of Global Optimization*. 2001;21(4):345–383.
URL <https://doi.org/10.1023/A:1012771025575>
- [25] Wan X, Pekny JF, Reklaitis GV. Simulation-based optimization with surrogate models—Application to supply chain management. *Computers & Chemical Engineering*. 2005;29(6):1317–1328. Selected Papers Presented at the 14th European Symposium on Computer Aided Process Engineering.
URL <https://www.sciencedirect.com/science/article/pii/S0098135405000190>
- [26] Jones DR, Schonlau M, Welch WJ. Efficient Global Optimization of Expensive Black-Box Functions. *Journal of Global Optimization*. 1998;13(4):455–492.
URL <https://doi.org/10.1023/A:1008306431147>
- [27] Forrester AI, Keane AJ. Recent advances in surrogate-based optimization. *Progress in Aerospace Sciences*. 2009; 45(1):50–79.
URL <https://www.sciencedirect.com/science/article/pii/S0376042108000766>

- [28] Jakobsson S, Patriksson M, Rudholm J, Wojciechowski A. A method for simulation based optimization using radial basis functions. *Optimization and Engineering*. 2010;11(4):501–532.
URL <https://doi.org/10.1007/s11081-009-9087-1>
- [29] Wang C, Duan Q, Gong W, Ye A, Di Z, Miao C. An evaluation of adaptive surrogate modeling based optimization with two benchmark problems. *Environmental Modelling Software*. 2014;60:167–179.
URL <https://www.sciencedirect.com/science/article/pii/S1364815214001698>
- [30] Shukla HA, Ferreira TdA, Faulwasser T, Bonvin D, Jones CN. Convergence certificate for stochastic derivative-free trust-region methods based on Gaussian processes. *arXiv preprint arXiv:201001120*. 2020;.
- [31] Cozad A, Sahinidis NV, Miller DC. Learning surrogate models for simulation-based optimization. *AIChE Journal*. 2014; 60(6):2211–2227.
URL <https://aiche.onlinelibrary.wiley.com/doi/abs/10.1002/aic.14418>
- [32] Bajaj I, Hasan MF. UNILOPT: Univariate projection-based optimization without derivatives. *Computers & Chemical Engineering*. 2019;127:71–87.
- [33] Nentwich C, Engell S. Surrogate modeling of phase equilibrium calculations using adaptive sampling. *Computers Chemical Engineering*. 2019;126:204–217.
URL <https://www.sciencedirect.com/science/article/pii/S0098135418312626>
- [34] Zhu K, Müller EA. Generating a Machine-Learned Equation of State for Fluid Properties. *The Journal of Physical Chemistry B*. 2020;124(39):8628–8639.
URL <https://doi.org/10.1021/acs.jpcb.0c05806>
- [35] Carranza-Abaid A, Svendsen HF, Jakobsen JP. Surrogate modelling of VLE: Integrating machine learning with thermodynamic constraints. *Chemical Engineering Science: X*. 2020;8:100080.
URL <https://www.sciencedirect.com/science/article/pii/S2590140020300265>
- [36] Eason JP, Kang J, Chen X, Biegler LT. Surrogate Equations of State for Equation-Oriented Optimization of Polymerization Processes. In: *13th International Symposium on Process Systems Engineering (PSE 2018)*, edited by Eden MR, Ierapetritou MG, Towler GP, vol. 44 of *Computer Aided Chemical Engineering*, pp. 781–786. Elsevier. 2018;.
URL <https://www.sciencedirect.com/science/article/pii/B9780444642417501257>
- [37] Hüllen G, Zhai J, Kim SH, Sinha A, Realff MJ, Boukouvala F. Managing uncertainty in data-driven simulation-based optimization. *Computers & Chemical Engineering*. 2020;136:106519.
- [38] Schweidtmann AM, Bongartz D, Huster WR, Mitsos A. Deterministic Global Process Optimization: Flash Calculations via Artificial Neural Networks. In: *29th European Symposium on Computer Aided Process Engineering*, edited by Kiss AA, Zondervan E, Lakerveld R, Özkan L, vol. 46 of *Computer Aided Chemical Engineering*, pp. 937–942. Elsevier. 2019;.
URL <https://www.sciencedirect.com/science/article/pii/B9780128186343501570>
- [39] Keßler T, Kunde C, McBride K, Mertens N, Michaels D, Sundmacher K, Kienle A. Global Optimization of Distillation Columns using Explicit and Implicit Surrogate Models. *Chemical Engineering Science*. 2018;197.
- [40] Jones DR, Schonlau M, Welch WJ. Efficient global optimization of expensive black-box functions. *Journal of Global optimization*. 1998;13(4):455–492.
- [41] Schweidtmann AM, Huster WR, Lüthje JT, Mitsos A. Deterministic global process optimization: Accurate (single-species) properties via artificial neural networks. *Computers Chemical Engineering*. 2019;121:67–74.
URL <https://www.sciencedirect.com/science/article/pii/S009813541830886X>
- [42] Schweidtmann AM, Mitsos A. Deterministic global optimization with artificial neural networks embedded. *Journal of Optimization Theory and Applications*. 2019;180(3):925–948.
- [43] Boukouvala F, Hasan MMF, Floudas CA. Global optimization of general constrained grey-box models: new method and its application to constrained PDEs for pressure swing adsorption. *Journal of Global Optimization*. 2017;67(1):3–42.
URL <https://doi.org/10.1007/s10898-015-0376-2>
- [44] Rebennack S, Kallrath J. Continuous piecewise linear delta-approximations for bivariate and multivariate functions. *Journal of Optimization Theory and Applications*. 2015;167(1):102–117.
- [45] Bajaj I, Hasan MF. Deterministic global derivative-free optimization of black-box problems with bounded hessian. *Optimization Letters*. 2019;pp. 1–16.
- [46] Song Y, Cao H, Khan KA. Bounding Convex Relaxations of Process Models from Below by Tractable Black-Box Sampling. *Computers & Chemical Engineering*. 2021;p. 107413.

- [47] Hasan MMF. An edge-concave underestimator for the global optimization of twice-differentiable nonconvex problems. *Journal of Global Optimization*. 2018;71(4):735–752.
URL <https://doi.org/10.1007/s10898-018-0646-x>
- [48] Tardella F. On the existence of polyhedral convex envelopes. In: *Frontiers in Global Optimization*, edited by Floudas CA, Pardalos P. Boston, MA: Springer US. 2004; pp. 563–573.
- [49] Purohit P, Höglund-Isaksson L. Global emissions of fluorinated greenhouse gases 2005–2050 with abatement potentials and costs. *Atmospheric Chemistry and Physics*. 2017;17(4):2795–2816.
URL <https://acp.copernicus.org/articles/17/2795/2017/>
- [50] Pardo F, Gutiérrez-Hernández SV, Zarca G, Urtiaga A. Toward the Recycling of Low-GWP Hydrofluorocarbon/Hydrofluoroolefin Refrigerant Mixtures Using Composite Ionic Liquid–Polymer Membranes. *ACS Sustainable Chemistry & Engineering*. 2021;9(20):7012–7021.
URL <https://doi.org/10.1021/acssuschemeng.1c00668>
- [51] Faúndez CA, Barrientos LA, Valderrama JO. Modeling and thermodynamic consistency of solubility data of refrigerants in ionic liquids. *International Journal of Refrigeration*. 2013;36(8):2242–2250.
URL <https://www.sciencedirect.com/science/article/pii/S0140700713001539>
- [52] Shiflett MB, Yokozeki A. Solubility and diffusivity of hydrofluorocarbons in room-temperature ionic liquids. *AIChE Journal*. 2006;52(3):1205–1219.
URL <https://aiche.onlinelibrary.wiley.com/doi/abs/10.1002/aic.10685>
- [53] Shiflett MB, Harmer MA, Junk CP, Yokozeki A. Solubility and diffusivity of 1,1,1,2-tetrafluoroethane in room-temperature ionic liquids. *Fluid Phase Equilibria*. 2006;242(2):220–232.
URL <https://www.sciencedirect.com/science/article/pii/S0378381206000689>
- [54] Ren W, Scurto AM. Phase equilibria of imidazolium ionic liquids and the refrigerant gas, 1,1,1,2-tetrafluoroethane (R-134a). *Fluid Phase Equilibria*. 2009;286(1):1–7.
URL <https://www.sciencedirect.com/science/article/pii/S0378381209002696>
- [55] Monjur MS, Demirel SE, Li J, Hasan MF. A Computer-Aided Platform for Simultaneous Process Synthesis and Intensification. In: *Computer Aided Chemical Engineering*, vol. 50, pp. 287–293. Elsevier. 2021;.
- [56] Demirel SE, Li J, Hasan MF. Systematic process intensification using building blocks. *Computers & Chemical Engineering*. 2017;105:2–38.
- [57] Demirel SE, Li J, Hasan MF. A general framework for process synthesis, integration, and intensification. *Industrial & Engineering Chemistry Research*. 2019;58(15):5950–5967.
- [58] Demirel SE, Li J, El-Halwagi M, Hasan MF. Sustainable Process Intensification Using Building Blocks. *ACS Sustainable Chemistry & Engineering*. 2020;8(48):17664–17679.
- [59] Monjur MS, Demirel SE, Li J, Hasan MF. SPICE_MARS: A Process Synthesis Framework for Membrane-Assisted Reactive Separations. *Industrial & Engineering Chemistry Research*. 2021;60(20):7635–7655.
- [60] Demirel SE, Li J, Hasan MF. Membrane Separation Process Design and Intensification. *Industrial & Engineering Chemistry Research*. 2021;60(19):7197–7217.
- [61] Misener R, Floudas CA. ANTIGONE: algorithms for continuous/integer global optimization of nonlinear equations. *Journal of Global Optimization*. 2014;59(2-3):503–526.
- [62] Androulakis IP, Maranas CD, Floudas CA. α BB: A global optimization method for general constrained nonconvex problems. *Journal of Global Optimization*. 1995;7(4):337–363.
- [63] Bongartz D, Mitsos A. Deterministic global optimization of process flowsheets in a reduced space using McCormick relaxations. *Journal of Global Optimization*. 2017;69(4):761–796.
- [64] Van Ness HCHC, Abbott MM. Classical thermodynamics of nonelectrolyte solutions : with applications to phase equilibria. 1982.
URL <http://catalog.hathitrust.org/api/volumes/oclc/7551672.html>
- [65] Reid RC, Prausnitz JM, Poling BE. *The properties of gases and liquids*. New York: McGraw-Hill. 1987.
- [66] Yokozeki A, Sato H, Watanabe K. Ideal-Gas Heat Capacities and Virial Coefficients of HFC Refrigerants. *International Journal of Thermophysics*. 1998;19(1):89–127.
URL <https://doi.org/10.1023/A:1021499018749>

Appendix A: VLE Modeling using the Gamma-Phi Approach

VLE expression for component k in an N component system can be expressed by the following equation⁶⁴:

$$y_k P \Phi_k = \tilde{x}_k \gamma_k P_k^s \quad (i = 1, \dots, N) \quad (38)$$

where, y_k is the vapor phase composition of component k , P is the total pressure of the system, Φ_k is the vapor phase fugacity coefficient of component k , $\tilde{x}_k$ is the equilibrium liquid phase composition of component k , γ_k is the liquid phase activity coefficient of component k and P_k^s is the saturated vapor pressure of component k at the system temperature T .

The saturated vapor pressure can be expressed by an Antoine type equation as follows:

$$\ln P_k^s = A_k^s - \frac{B_k^s}{T + C_k^s} \quad (39)$$

where, A_k^s , B_k^s , and C_k^s are Antoine parameters of component k .

The vapor phase nonideality can be approximated⁶⁴ by the following fugacity coefficient expression:

$$\Phi_k = \exp \frac{(B_k - \tilde{V}_k)(P - P_k^s)}{RT} \quad (40)$$

where, B_k is the second virial coefficient of component k at system temperature T , $\tilde{V}_k$ is the saturated molar liquid volume of component k at system temperature T , and R is universal gas constant.

Margules and Non Random Two Liquid (NRTL) are well known two-parameter activity coefficient models⁶⁵ to analyze experimental solubility data. For a binary mixture, these models can be used to calculate the activity coefficient of each of the components by the following equations:

$$\ln \gamma_1 = \left\{ \begin{array}{ll} [A + 2(B - A)\tilde{x}_1]\tilde{x}_2^2, & \text{Margules} \\ \tilde{x}_2^2 [\tau_{21}(\frac{G_{21}}{\tilde{x}_1 + \tilde{x}_2 G_{21}})^2 + \frac{\tau_{12} G_{12}}{(\tilde{x}_2 + \tilde{x}_1 G_{12})^2}], & \text{NRTL} \end{array} \right\} \quad (41)$$

$$\ln \gamma_2 = \left\{ \begin{array}{ll} [B + 2(A - B)\tilde{x}_2]\tilde{x}_1^2, & \text{Margules} \\ \tilde{x}_1^2 [\tau_{12}(\frac{G_{12}}{\tilde{x}_2 + \tilde{x}_1 G_{12}})^2 + \frac{\tau_{21} G_{21}}{(\tilde{x}_1 + \tilde{x}_2 G_{21})^2}], & \text{NRTL} \end{array} \right\} \quad (42)$$

where, $A, B, \tau_{12}, \tau_{21}$ are binary interaction parameters, $G_{12} = \exp(-\alpha\tau_{12})$, $G_{21} = \exp(-\alpha\tau_{21})$, and α is binary system specific NRTL constant.⁵²

To this end, we use the above described Gamma-Phi approach to predict the amount of HFC absorbed in IL. It is assumed that, for HFC(1)/IL(2) binary system, the vapor phase is pure HFC ($y_1 = 1$, no IL is present, $P_2^S = 0$).⁵² Under this assumption, the Gamma-Phi thermodynamic model simplifies to the following form⁵²:

$$f(P, T, \tilde{x}_1) = \ln P + \frac{(P - P_1^s)(B_1 - \tilde{V}_1)}{RT} - \ln P_1^s - \ln \tilde{x}_1 - \ln \gamma_1 \quad (43)$$

Here we consider two HFC/IL binary systems, namely R-134a/[bmim][PF₆] and R-32/[bmim][PF₆]. The experimental data for the second virial coefficient, $B_1[\frac{cm^3}{mol}]$ of the HFCs are obtained from Yokozeki et al.⁶⁶ and the following quadratic fitted ($R^2 > 0.99$) expressions are used for modeling purpose:

$$B_1(T) = \left\{ \begin{array}{ll} -0.02T^2 + 19.33T - 4046.9, & \text{R-134a/[bmim][PF}_6\text{]} \\ -0.01T^2 + 11.51T - 2432.2, & \text{R-32/[bmim][PF}_6\text{]} \end{array} \right\} \quad (44)$$

The expressions for $\tilde{V}_1[\frac{cm^3}{mol}]$ and saturated vapor pressure, $P_1^s[MPa]$ are obtained from Shiflett and Yokozeki.⁵²

List of Figures

1	(a) $g(x)$ along with $L(x)$ generated at $x^{Int}=1$ (the interior sample point). The linear facets underestimating $L(x)$ are shown in solid lines. (b) Multiple interior simulated points result in multiple underestimators. The blue crosses denote the points where the value of the underestimators are known. Note: The underestimators generated at the bound simulation points are not shown here.	26
2	$g(x)$ along with its closest underestimating linear facets (shown in purple). The underestimators generated for the 5 interior points are shown in dotted grey lines.	27
3	$g(x)$ along with its closest underestimating linear facets (shown in purple) and overestimating linear facets (shown in orange). Five simulation points are used.	28
4	The MILP formulation results in the objective value attaining a minimum at the brown square. The approximate solution to the problem $f(x)=0$ is the x value where this minimum occurs.	29
5	Two phase/flash separator representation; (a) conventional representation and (b) building block-based representation.	30
6	Solubility isotherms of R-134a with (a) Margules activity coefficient model and (b) NRTL activity coefficient model. The lines represent the solubility isotherms by the Gamma-Phi based method and the symbols (o) represents the solubility predicted by our MILP-based surrogate approximation with 101 interior points.	31
7	Solubility isotherms of R-32 with (a) Margules activity coefficient model and (b) NRTL activity coefficient model. The lines represents the solubility isotherms by the Gamma-Phi based method and the symbols (o) represents the solubility predicted by the developed surrogate model with 101 interior points.	32
8	CPU times with varying number of interior points.	33

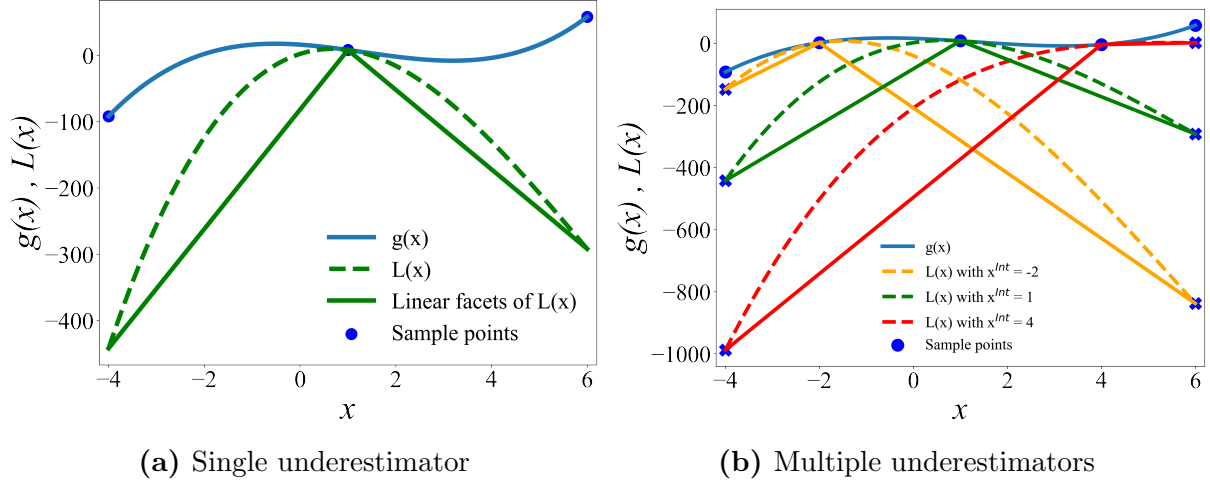


Figure 1: (a) $g(x)$ along with $L(x)$ generated at $x^{Int}=1$ (the interior sample point). The linear facets underestimating $L(x)$ are shown in solid lines. (b) Multiple interior simulated points result in multiple underestimators. The blue crosses denote the points where the value of the underestimators are known. Note: The underestimators generated at the bound simulation points are not shown here.

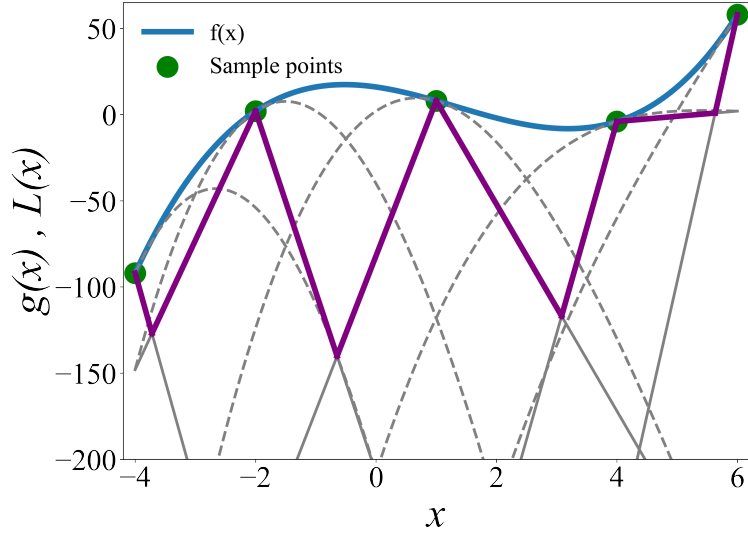


Figure 2: $g(x)$ along with its closest underestimating linear facets (shown in purple). The underestimators generated for the 5 interior points are shown in dotted grey lines.

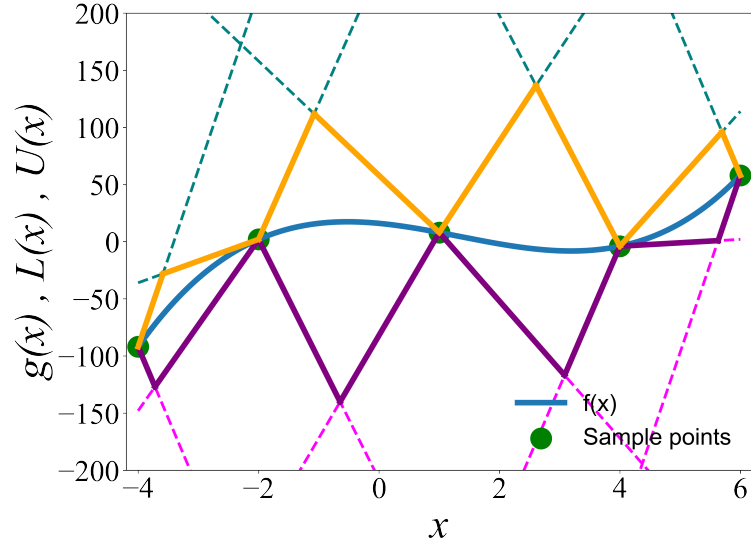


Figure 3: $g(x)$ along with its closest underestimating linear facets (shown in purple) and overestimating linear facets (shown in orange). Five simulation points are used.

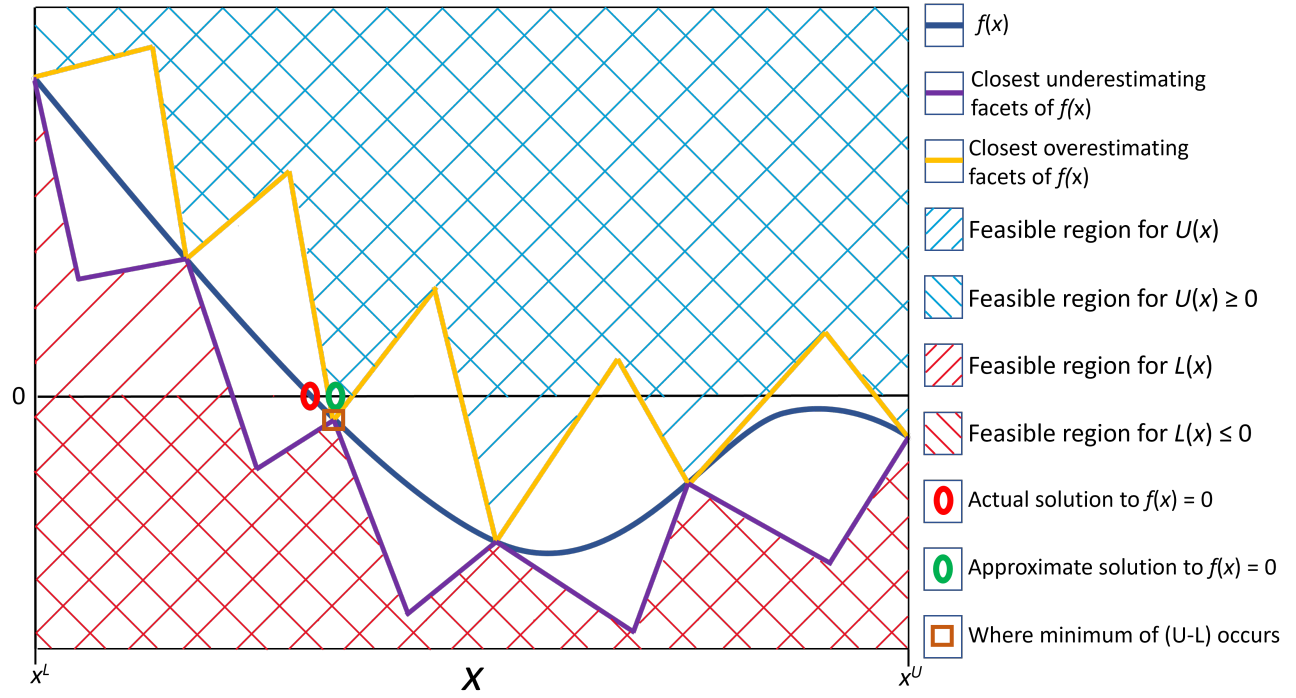


Figure 4: The MILP formulation results in the objective value attaining a minimum at the brown square. The approximate solution to the problem $f(x)=0$ is the x value where this minimum occurs.

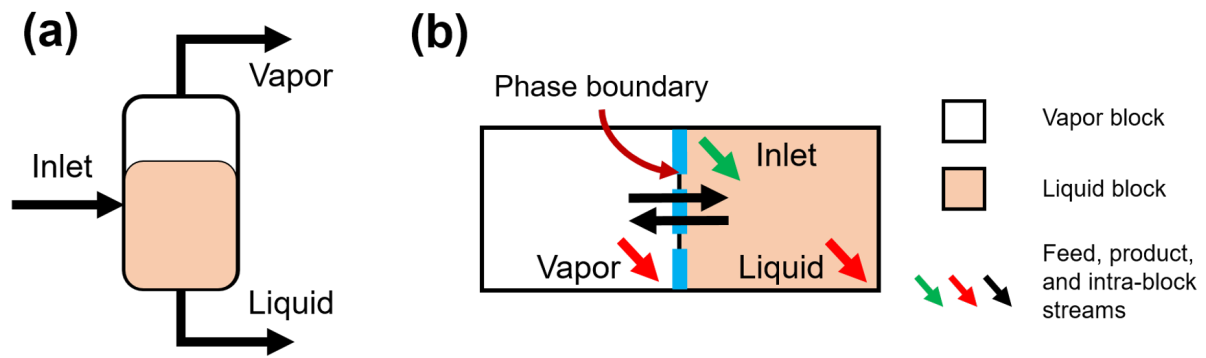


Figure 5: Two phase/flash separator representation; (a) conventional representation and (b) building block-based representation.

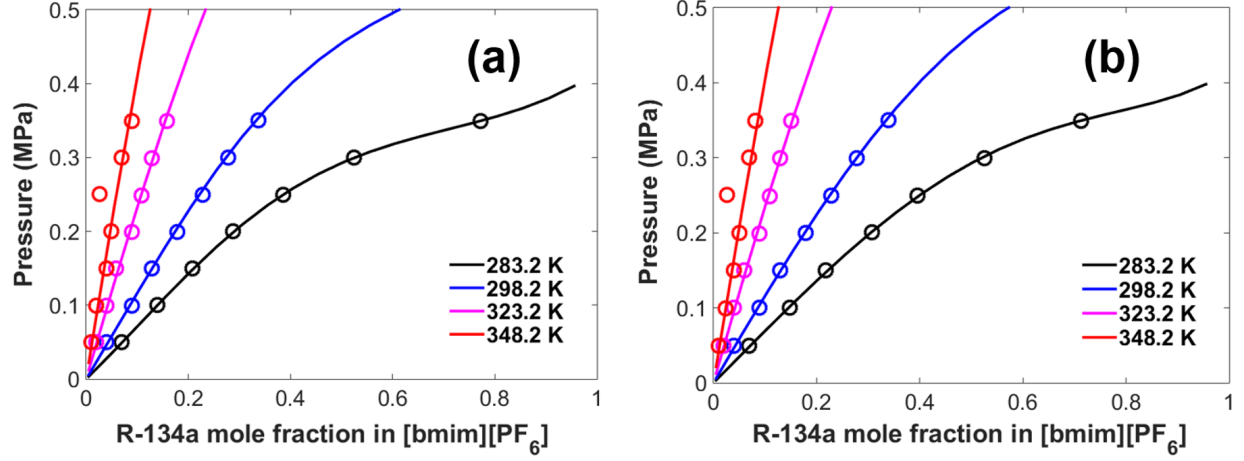


Figure 6: Solubility isotherms of R-134a with (a) Margules activity coefficient model and (b) NRTL activity coefficient model. The lines represent the solubility isotherms by the Gamma-Phi based method and the symbols (o) represents the solubility predicted by our MILP-based surrogate approximation with 101 interior points.

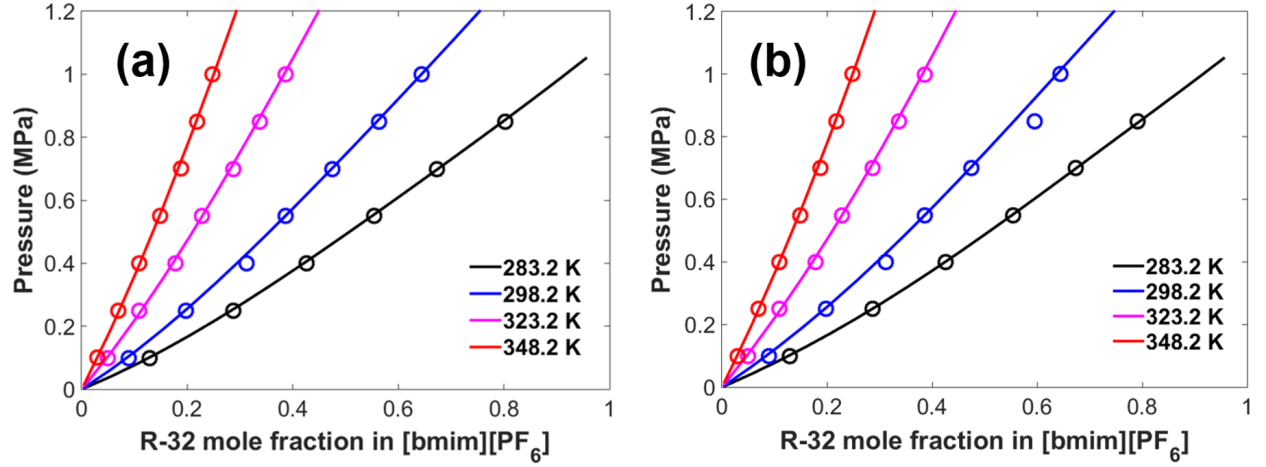


Figure 7: Solubility isotherms of R-32 with (a) Margules activity coefficient model and (b) NRTL activity coefficient model. The lines represents the solubility isotherms by the Gamma-Phi based method and the symbols (o) represents the solubility predicted by the developed surrogate model with 101 interior points.

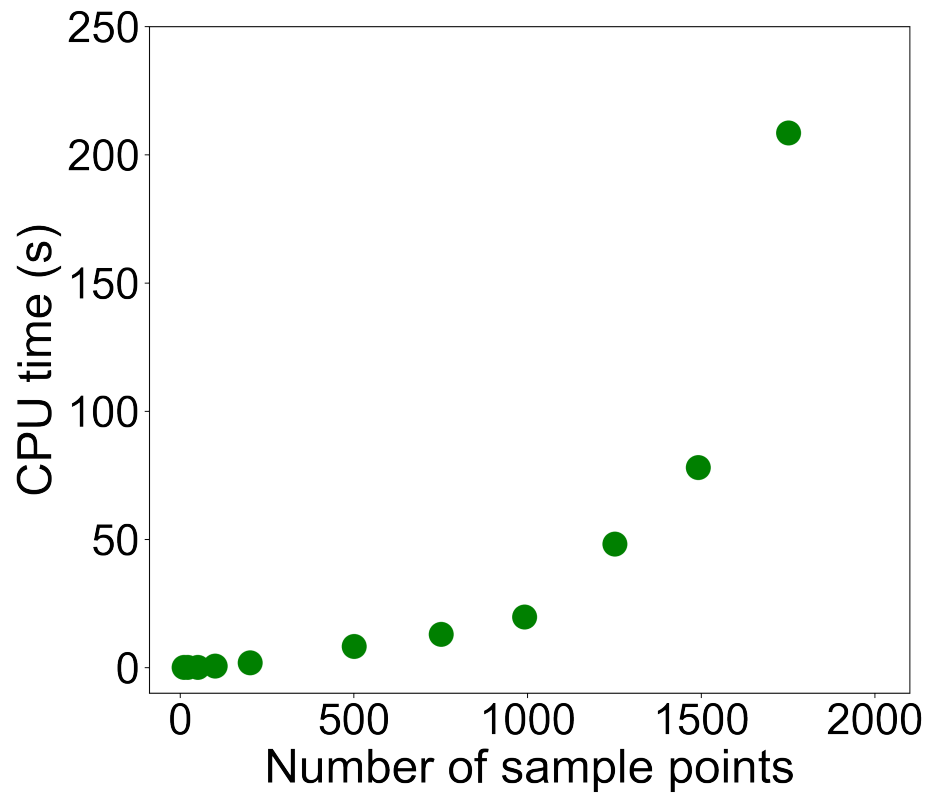


Figure 8: CPU times with varying number of interior points.

Table 1: Maximum separation distance of closest linear underestimating facets from $g(x)$ for different number of samples.

Number of interior points	Maximum separation distance
5	125.19
10	50.02
20	22.47
50	8.53
100	4.16

Table 2: Obtained θ values for the HFC(1)/IL(2) system.

Binary systems (1)/(2)	Activity coefficient model	θ_P	θ_T	$\theta_{\widetilde{x}_1}$
R-134a/[bmim][PF ₆]	Margules	0	0.00029	4999.39
R-134a/[bmim][PF ₆]	NRTL	0	0	5000.71
R-32/[bmim][PF ₆]	Margules	0	0.00005	5000.29
R-32/[bmim][PF ₆]	NRTL	0	0.00009	5000.52

Table 3: Solubility prediction of R-134a with Margules activity coefficient model.

Temperature (K)	Pressure (MPa)	Liquid mol% based on Gamma-Phi approach*	Number of interior points	Liquid mol% based on MILP approximation	Objective function	Prediction error (%)
283.1	0.349	77.5 (72.4)	11	80.0	0	-3.3
			21	75.0	0	3.2
			51	78.0	0	-0.7
			101	77.2	0	0.4
298.1	0.35	33.3 (32.6)	11	30.1	0.1	9.7
			21	35.1	0.1	-5.3
			51	34.1	0	-2.3
			101	33.7	0	-1.1
323.2	0.349	15.4 (15.4)	11	20.1	0.4	-30.3
			21	15.1	0.1	2.1
			51	16.1	0.1	-4.4
			101	15.8	0.1	-2.9
348.2	0.349	8.4 (8.5)	11	10.1	0.3	-19.2
			21	10.1	0.3	-19.2
			51	8.1	0.1	4.0
			101	8.9	0.1	-5.6

*Bracketed values are from experimental results.⁵²

Table 4: Solubility prediction of R-134a with NRTL activity coefficient model.

Temperature (K)	Pressure (MPa)	Liquid mol% based on Gamma-Phi approach*	Number of interior points	Liquid mol% based on MILP approximation	Objective function	Prediction error (%)
283.1	0.349	71.6 (72.4)	11	70.0	0	2.2
			21	70.0	0	2.2
			51	72.0	0	-0.6
			101	71.3	0	0.4
298.1	0.35	33.6 (32.6)	11	30.1	0.2	10.5
			21	35.1	0	-4.4
			51	34.1	0	-1.4
			101	34.0	0	-1.2
323.2	0.349	15.2 (15.4)	11	20.1	0.5	-32.2
			21	15.1	0	0.7
			51	16.1	0	-5.9
			101	15.1	0	0.7
348.2	0.349	8.5 (8.5)	11	10.1	0.3	-18.2
			21	10.1	0.3	-18.2
			51	8.1	0.1	4.7
			101	8.1	0.1	4.7

*Bracketed values are from experimental results.⁵²

Table 5: Solubility prediction of R-32 with Margules activity coefficient model.

Temperature (K)	Pressure (MPa)	Liquid mol% based on Gamma-Phi approach*	Number of interior points	Liquid mol% based on MILP approximation	Objective function	Prediction error (%)
283.2	0.8495	79.8 (81.5)	11	80.2	0.1	-0.5
			21	80.2	0.1	-0.5
			51	80.2	0	-0.5
			101	80.2	0	-0.5
298.1	0.9999	64.4 (62.8)	11	60.4	0.1	6.4
			21	65.3	0	-1.2
			51	64.4	0	0.2
			101	64.4	0	0.2
323.1	0.6996	28.3 (28.5)	11	30.7	0.1	-8.5
			21	30.7	0.2	-8.5
			51	28.7	0	-1.4
			101	28.7	0	-1.4
348.1	0.5498	14.7 (14.6)	11	10.9	0.6	25.8
			21	15.8	0.1	-7.5
			51	14.9	0	-1.4
			101	14.9	0	-1.4

*Bracketed values are from experimental results.⁵²

Table 6: Solubility prediction of R-32 with NRTL activity coefficient model.

Temperature (K)	Pressure (MPa)	Liquid mol% based on Gamma-Phi approach*	Number of interior points	Liquid mol% based on MILP approximation	Objective function	Prediction error (%)
283.2	0.8495	80.7 (81.5)	11	80.2	0	0.6
			21	80.2	0	0.6
			51	80.2	0	0.6
			101	79.2	0	1.9
298.1	0.9999	67.1 (62.8)	11	60.4	0.1	10
			21	65.4	0	2.5
			51	64.4	0	4.0
			101	64.4	0	4.0
323.1	0.6996	28.5 (28.5)	11	30.7	0.2	-7.7
			21	30.7	0.2	-7.7
			51	28.7	0	-0.7
			101	28.7	0	-0.7
348.1	0.5498	14.7 (14.6)	11	10.9	0.6	25.8
			21	15.8	0.1	-7.5
			51	14.9	0	-1.4
			101	14.9	0	-1.4

*Bracketed values are from experimental results.⁵²