Outer Approximation Algorithm with Physical Domain Reduction for Computer-Aided Molecular and Separation Process Design

Smitha Gopinath, George Jackson, Amparo Galindo, and Claire S. Adjiman
Dept. of Chemical Engineering, Centre for Process Systems Engineering, Imperial College London, South Kensington Campus, London SW7 2AZ, U.K

DOI 10.1002/aic.15411
Published online August 4, 2016 in Wiley Online Library (wileyonlinelibrary.com)

Integrated approaches to the design of separation systems based on computer-aided molecular and process design (CAMPD) can yield an optimal solvent structure and process conditions. The underlying design problem, however, is a challenging mixed integer nonlinear problem, prone to convergence failure as a result of the strong and nonlinear interactions between solvent and process. To facilitate the solution of this problem, a modified outer-approximation (OA) algorithm is proposed. Tests that remove infeasible regions from both the process and molecular domains are embedded within the OA framework. Four tests are developed to remove subdomains where constraints on phase behavior that are implicit in process models or explicit process (design) constraints are violated. The algorithm is applied to three case studies relating to the separation of methane and carbon dioxide at high pressure. The process model is highly nonlinear, and includes mass and energy balances as well as phase equilibrium relations and physical property models based on a group-contribution version of the statistical associating fluid theory (SAFT-γ Mie) and on the GC$^+$ group contribution method for some pure component properties. A fully automated implementation of the proposed approach is found to converge successfully to a local solution in 30 problem instances. The results highlight the extent to which optimal solvent and process conditions are interrelated and dependent on process specifications and constraints. The robustness of the CAMPD algorithm makes it possible to adopt higher-fidelity nonlinear models in molecular and process design. © 2016 The Authors AIChE Journal published by Wiley Periodicals, Inc. on behalf of American Institute of Chemical Engineers *AIChE J*, 62: 3484–3504, 2016
Keywords: mixed-integer optimization, molecular design, absorption, carbon dioxide capture, SAFT equation of state

Introduction

The transformation of feedstocks to desired products in chemical processes involves the use of a large variety of processing materials[1] such as solvents, adsorbents, catalysts, and heat transfer fluids. Traditionally, the selection of processing materials and the design of the process (flow sheet, unit sizes, operating conditions) have been approached sequentially,[2] although, material and process decisions are in fact interdependent.[2,3] Choosing a processing material based on a few desirable physicochemical properties, in isolation from process performance considerations, can thus lead to poor decisions: for example, a solvent that exhibits high solubility and selectivity for the solute of interest may be too expensive to regenerate, compromising the economic viability of the process. Instead, a process-wide evaluation of the material is essential to identify choices that lead to better, or even optimal, process performance metrics such as reduced cost and environmental impact.[4] Given the potential benefits that can be derived from an integrated approach to material and process design, there has been growing interest in addressing computer-aided molecular and process design (CAMPD) problems,[5] in which the design of the processing materials or molecules and that of the process are considered simultaneously.

In general, a CAMPD problem can be posed as a mixed-integer nonlinear optimization problem (MINLP), provided that predictive algebraic models are available to capture the impact of material/molecular structure on relevant physicochemical properties, and the effect of these properties on the appropriate unit operations. Discrete variables are used to represent molecular-level decisions such as the number of groups of a given kind (for example, how many hydroxyl groups the optimal molecule contains, if any), with constraints used to specify how the groups can be combined.[6–9] Discrete variables can also be used to represent the connectivity between the groups,[10,11] and the identity of components if the material of interest is a mixture.[12–14]

The CAMPD problem is inherently more complex than the corresponding process design problem with fixed material choices. First, the presence of discrete choices makes the problem combinatorial in nature. Second, the design problem is highly nonlinear: many of the models that relate structural information to physical properties, such as the UNIQUAC functional-group activity coefficients (UNIFAC) model[15] or the group contribution statistical associating fluid theory with

Correspondence concerning this article should be addressed to C. S. Adjiman at c.adjiman@imperial.ac.uk.

a Mie potential (SAFT-γ Mie) equation of state,[16] are nonconvex, making it more challenging for local solvers to converge to the global minimum or even a good solution. This is compounded by the fact that the identification of a feasible point for the process model for given values of the design variables can be challenging from a numerical perspective in the absence of a good initial guess. Third, there usually exist combinations of the discrete variables that satisfy all molecular design constraints but that make the process model infeasible, because many implicit phase-behavior constraints must be satisfied for the successful solution of a process model. For instance, in the case of a solvent-based gas separation process, the process is infeasible if the discrete variables represent a solvent that is in the vapor phase at inlet conditions (temperature and pressure). A more challenging implicit constraint is that both the vapor and liquid phases must coexist at equilibrium across the entire set of operating conditions of the separation unit. Process models are usually derived assuming that this behavior holds, a reasonable assumption when all materials are fixed. In the context of CAMPD, however, the violation of these implicit constraints on fluid-phase behavior is likely to occur and often leads to numerical failure. Even if the nonlinear equation solver converges, the solution is usually physically meaningless in such cases. Furthermore, constraints on phase behavior are inherently discontinuous,[17] and can thus result in the failure of the optimization solver unless they are handled specifically. One strategy to address these discontinuities is to incorporate them explicitly in the process model through the use of disjunctions[18] or through the use of complementarity constraints.[19] Such formulations, however, can require a greater number of discrete variables and can increase the complexity of process models. Another recently proposed strategy to deal with model discontinuities arising from a change in the number of phases is to carry out phase stability and equilibrium calculations for each stage via an external function.[20] The effective handling of these implicit constraints remains an active area of research.

Given these significant challenges, several methodologies have been proposed for the solution of CAMPD problems. One approach is the reformulation of the problem as a continuous NLP. This can sometimes be achieved by placing restrictions on the types of materials that can be designed. For instance, Pereira et al.[5] considered the simultaneous design of a blend of *n*-alkanes and the corresponding absorption process for the removal of carbon dioxide from a methane stream. Another way to develop a continuous optimization formulation at the process level is to optimize process performance in the space of molecular properties or descriptors in a first stage, leaving the identification of the optimal molecule or molecular structure for a second stage. In this vein, Eden et al.[2] proposed the formulation of a continuous process design problem to identify physical property targets, that is, the values of the properties that give the best process performance. The concept of a "property cluster" was used to reduce the dimensionality of the problem. These targets were then used in a computer-aided molecular design (CAMD) approach to find molecules that (nearly) achieve these targets. Within this class of approaches, the CAMD step can be performed via the use of "molecular property clusters,"[21] by using algebraic methods,[22] or molecular signatures.[23] In another method first proposed by Bardow et al.,[24,25] continuous molecular targeting (CoMT), the continuous descriptors of an optimal (hypothetical) solvent, representing the parameters of the PC-SAFT equation of state,[26] were first determined based on process performance. This was then used to identify an optimal molecule with similar descriptors, from a list of compounds[25,27] or more recently by deploying CAMD techniques to derive a CoMT-CAMD methodology.[28] These two-stage approaches can be seen as top-down strategies, as optimal process performance is sought first, and an appropriate molecular structure is then derived from this.

Other two-stage approaches can be viewed as bottom-up approaches in that they start from a molecular perspective and build up to an optimal process. The central idea is to reduce the combinatorial complexity of molecular design by first screening molecules from a wide design space, often using relatively simple property models and user-defined property targets, before using more demanding property and process models to evaluate the remaining options. This general methodology has been explored by several groups. The work of Karunanithi et al.[29] falls within this category, for example. In addition to screening based on property targets, Hostrup et al.[3] have used an analysis of phase diagrams, along with a metric of the driving force required for vapor-liquid separation to screen both molecular and separation process alternatives. Such a framework has been applied more recently to the design of ionic liquid entrainers for extractive distillation.[30] Approximate process models have also been used in the screening stage, for example by using targets on solvent selectivity and on process energy demand, as predicted with a shortcut model, to screen for entrainers.[31] The use of explicit property targets that are set based on prior knowledge or heuristics can, however, lead to the elimination of optimal solutions, just as the use of approximate models can. An alternative to specified property targets is to set targets based on the preferred "direction" of each property, that is, whether the property value should maximized or minimized. Multi objective optimization (MOO) techniques have been applied in this context, to identify molecules that lie on a Pareto front of physical property targets set by the designer based on insights into the process of interest. This smaller space of molecules, consisting of molecules in the Pareto set of solutions, can then be assessed further based on performance in the process[32,33] or by using clustering of molecules to reduce the number of options.[32,34–36]An underlying assumption in such methods is that the optimal solution of the CAMPD problem lies on the Pareto front. However, this may not be the case if the objective function of the CAMPD problem does not vary monotonically with respect to each property or if the constraints of the CAMPD problem make some Pareto points infeasible. Another decomposition approach has been to optimize the structure of the molecule using a stochastic algorithm, whilst solving the process design problem with a gradient-based algorithm for each structure generated.[37]

In both top-down and bottom-up two-stage methods, the solution obtained may differ from the solution of the fully integrated CAMPD problem. We note that in principle MOO-based approaches offer a greater likelihood of identifying the solution than other approaches due to the absence of weights on the properties. In decomposing the problem, the strong interdependence between the process and molecular scales is represented in a simplified manner. In reality, several properties of the molecules or materials being designed play a role in determining the performance of the process and they do so in a nonlinear way, with unknown or indeed variable relative importance.[25,35] Furthermore, many of the molecular/mixture

properties vary with operating conditions, that is, they are secondary properties in the sense discussed by Jaksland et al.[38] The optimal operating conditions of the process, in turn, depend on the material that is chosen, as well as the process constraints and specifications. Even the feasible operating region depends on the material chosen: the range of temperatures and pressures at which a solvent is in the liquid state depends on its molecular structure; some choices of molecular structure may lead to the appearance of new phases, perhaps due to immiscibility or partial miscibility between the various components in the process. The optimal solution of the full CAMPD problem, therefore, corresponds to a trade-off between different properties and process variables. In this closely interlinked multidimensional problem, the sequential design of a system consisting of the process and the processing materials, or molecules, may be suboptimal.[1]

To address this issue, several solution methodologies for the integrated molecular and process design problem (the "full MINLP") have been proposed. The main challenge arises from the highly nonlinear nature of the MINLP formulation that represents the integrated design problem. In the approach of Pereira et al.,[5] mentioned previously, the SAFT-VR SW equation of state[39,40] was used as a reliable and predictive model of the relevant thermodynamics. Although this model is highly nonlinear, the tractability of the problem was ensured by considering a continuous molecular design space. The direct solution of the MINLP arising from the CAMPD problem was adopted by Zhou et al.[41] to design a reactive process and the corresponding reaction solvent, including the recovery of the solvent from the reaction products by distillation. The complexity of the process model was tailored to make the problem tractable. In particular, the distillation column was modelled via a shortcut model and by assuming ideal vapor and liquid phases. The full CAMPD problem was also solved to design an extractive fermentation process and solvent,[42] based on a mixed-integer quadratic formulation. Initial guesses for the solution of the CAMPD using mixed integer sequential quadractic programming[43] algorithm were obtained by applying an evolutionary algorithm to solve the CAMPD. The integrated design of an organic Rankine cycle process conditions and working fluid was also solved as a full MINLP in recent work, facilitated by the fact that only pure component phase behavior is of relevance in such a case.[44]

To handle more general design problems, one can adopt the approach of Buxton et al.[8] who modified the generalized Benders decomposition (GBD) algorithm[45]: they introduced several steps prior to the solution of the primal problem, including a series of property tests that form a subset of the CAMD problem constraints, the initialization of various sets of equations in the process model, and mass-transfer feasibility tests, in which the process operating conditions were assumed to be fixed *a priori*. This approach was extended to tackle mixed-integer dynamic optimization problems,[46] to enable the simultaneous design of a batch process and the associated solvent. In these studies, the highly-nonlinear UNIFAC model[15] was combined with the ideal gas equation to represent the relevant phase equilibria. The full solution of the CAMPD problem was also achieved by Burger et al.[47] based on a hierarchical optimization approach (HiOpt). In this case, simplified models of the process units were combined with rigorous thermodynamics using the SAFT-γ Mie equation of state[16] and were used to optimize several performance metrics derived from the simplified process model. A multiobjective optimization algorithm was used to generate solutions that approximate the Pareto front of the MOO problem. These were then used as initial guesses for the solution of the full MINLP, which included detailed process and thermodynamic models. Thus, whereas MOO has been embedded in other approaches as a screening tool to reduce the size of the solution space, Burger et al.[47] used MOO to generate high-quality starting points to help overcome the inherent nonconvexity of the problem, albeit without guarantee of global optimality. We note, however, that despite the useful initialization data that were produced by the solution of the MOO, the local solution of the full MINLP remained prone to initialization and convergence failures.

In our current contribution, we build on recent work[48] to propose a robust algorithm for the solution of the full MINLP. Several novel tests are embedded within a modified outer-approximation (OA)[49,50] algorithm to solve the MINLP, akin to the general principle of integrating tests into a modified GBD algorithm deployed by Buxton et al.[8] and Giovanoglou et al.[46] The tests we develop differ from these earlier approaches, however, as the feasibility of the process is assessed for combinations of the values of the process and molecular variables, rather than for values of the molecular variables alone. When a new solvent is generated at a major iteration of the modified OA algorithm, the tests help to ascertain the feasibility of using the solvent in the process, before solving the process optimization problem (primal problem) for the fixed solvent. The aims of the tests are twofold: to determine *a priori* if a solvent is feasible in the process and, if it is, the ranges of values of the process variables for which it may be feasible. If the solvent is found to be infeasible throughout the process domain, it is removed from the search space without the need to solve the primal problem. If it is found to be feasible for some ranges of the variables only, these ranges define the "reduced process domain." Through the tests we thus recognize that the feasible process domain varies with the choice of solvent. In the screening methodology proposed here, unlike in previous work, molecules do not have to be screened at arbitrarily fixed operating conditions, but their feasibility may be evaluated across the process domain. A further useful output of the tests comes in the form of initial guesses for the optimization of the primal problem that lie in the reduced process domain. This is complemented by an initialization strategy that contributes to the overall robustness of the algorithm. In our current contribution, the tests are developed with a specific focus on solvent-based absorption processes; a similar approach can be followed for other separation processes, for example, liquid-liquid extraction.

The article is organized as follows. In the next section, a motivating example is introduced to highlight more specifically the difficulties that must be overcome to solve CAMPD problems. The proposed tests are then developed in the methodology section and their integration into the modified OA algorithm is discussed in the proposed CAMPD algorithm section. The application of the algorithm to several variants of the motivating example is investigated in the case studies section, where the effectiveness of the tests and the robustness of the algorithm are analyzed. Conclusions and perspectives are discussed in the final section of the article.

Motivating Example

To illustrate the challenges inherent in CAMPD, we consider the following gas absorption design problem previously studied by Burger et al.[47]: Given a flow sheet configuration for an

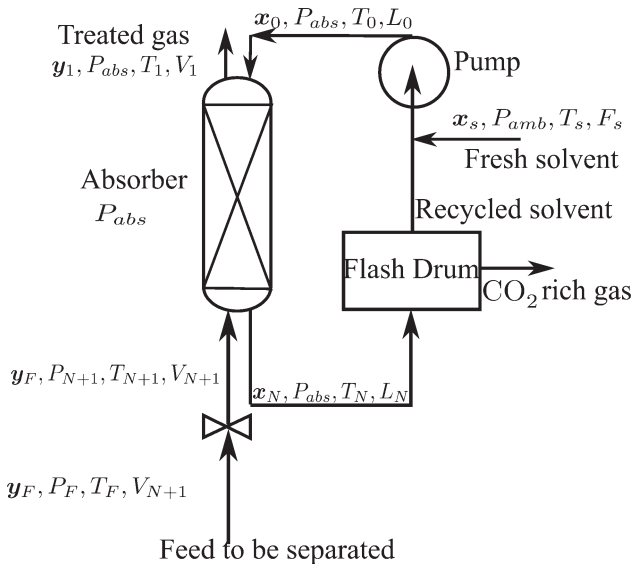

Figure 1. A flow sheet for the removal of carbon dioxide from a natural gas stream via absorption, as considered in Burger et al.[47]

absorption process, the composition $\mathbf{y}_F$, temperature T_F, and pressure P_F of the gaseous feed to be separated, and performance objectives and constraints, find the optimal values of the pressure in the absorber P_{abs}, the recycle flow rate of the solvent L_0, and the vector $\boldsymbol{n}$ of numbers of groups of each type in the solvent.

The flow sheet is shown in Figure 1. The feed to be separated comprises carbon dioxide and methane (as a simplification of a natural gas stream). The feed passes through an expansion valve and is contacted with a solvent in a counter current absorber with 10 stages. The treated gas leaving at the top of the absorber is required to have a methane purity of at least y_p. The spent solvent is regenerated at $P_{\text{flash}} = 0.1$ MPa. The regenerated solvent is mixed with a pure solvent at temperature $T_s = 298$ K to make up for solvent losses. The resulting solvent stream is then pumped back into the absorber at a flow rate L_0. The objective is to maximize the net present value of the process over a 15 year lifetime.

The models chosen to represent the thermodynamics of the mixtures in the process play an important role in determining the validity of the solutions obtained. In Burger et al.,[47] most thermodynamic properties were predicted using SAFT-γ Mie,[16] a group contribution equation of state (EoS) that belongs to the family of SAFT EoSs.[51–54] A group contribution EoS, such as SAFT-γ Mie, offers a computationally tractable way of predicting the properties of molecules based on their chemical composition as described by the number of occurrences of each type of group in the molecule. Although it is common in molecular design work to describe liquid phases with the well-established UNIFAC model,[15] SAFT -γ Mie allows one to consider a continuous and consistent description of thermodynamic properties for the entire fluid region (i.e., gas and liquid) and provides accurate predictions of fluid phase behavior at the high pressures relevant to this case study. In addition, the GC$^+$ method[55] was used to predict melting points and flash points, while the viscosity and surface tension were estimated using correlations.[56,57] More details of the process model and property models may be found in the papers by Burger et al.[47] and Pereira et al.[5]

Although the flow sheet considered is relatively simple, the solution of the CAMPD problem is challenging. The implementation of the process model presented by Burger et al.[47] comprised 548 equations, excluding the equations related to the evaluation of thermodynamic functions (e.g., enthalpy, chemical potential) with the SAFT-γ Mie EoS. The phase-equilibrium equations, namely the equality of the chemical potentials of each component across all phases and the equality of pressure across all phases (and trivially the equality of temperature), were included explicitly in the model for each stage as no flash algorithm was available for use with SAFT-γ Mie at the time the work was conducted. An added complication in this model is that the EoS is explicit in the space of temperature T-volume V- mole fraction $\boldsymbol{x}$ coordinates, whereas the process model is implemented in T-P-$\boldsymbol{x}$ coordinates, where P is the pressure. This nonlinear subset of equations may have several roots, the number of which is not known *a priori*. Hence, the initialization of the EoS with a good guess for the volume was often necessary to obtain a solution that satisfies phase equilibrium. In the HiOpt approach,[47] initial guesses were generated by solving a simplified formulation of the full CAMPD as a MOO problem, providing solvent candidates judged to be of high quality on the basis of the MOO criteria. Each solution was then used as a starting point to solve the full CAMPD with the default OA-based MINLP solver in gPROMS.[58] Of the six starting points generated, the full CAMPD problem was solved for only three starting points. Difficulties arise in particular when the nonlinear solver fails to find a feasible point during the solution of the primal problem. While it is expected for infeasible primal problems to be encountered, it may be that a feasible point exists but is not found due to nonlinearities. Furthermore, in the gPROMS modelling environment used in this and our current work, a sequential solution approach is adopted in solving the primal problem so that the optimization takes place in the space of degrees of freedom only. It is then important to find a feasible point for the process model equations (i.e., a square system of nonlinear equations) and to obtain the gradients of the constraints with respect to the degrees of freedom. Failed evaluations of the process model or its gradients were found to occur during the course of optimizations from three starting points and led to convergence failure. Thus, while the HiOpt approach yielded high-performance solvent/process combinations, there is significant scope for further enhancement of robustness and efficiency, which may in turn lead to improved local solutions of the CAMPD problem.

Proposed Methodology

The general CAMPD problem may be formulated as follows

$$\begin{aligned} \min_{\boldsymbol{x},\boldsymbol{n}} \quad & f(\boldsymbol{x},\boldsymbol{n}) \\ \text{s.t.} \quad & \boldsymbol{h}(\boldsymbol{x},\boldsymbol{n}) = 0 \\ & \boldsymbol{g}(\boldsymbol{x},\boldsymbol{n}) \le 0 \\ & \boldsymbol{C}\boldsymbol{n} \le \boldsymbol{d} \\ & \boldsymbol{x} \in \boldsymbol{X} \\ & \boldsymbol{n} \in \boldsymbol{N} \end{aligned} \tag{P}$$

where $\boldsymbol{x} \in \boldsymbol{X} \subset \mathbb{R}^c$ is a c-dimensional vector of continuous variables, and $\boldsymbol{n} \in \boldsymbol{N} \subset \mathbb{Z}^{+q}$ is a q-dimensional vector of non-negative integer variables, where n_i represents the number of

occurrences of group i in the molecule. The set of equations $\boldsymbol{h}: \boldsymbol{X}\times\boldsymbol{N} \rightarrow \mathbb{R}^e$ represents the process and property models. $\boldsymbol{g}: \boldsymbol{X}\times\boldsymbol{N} \rightarrow \mathbb{R}^a$ represents design constraints. $f: \boldsymbol{X}\times\boldsymbol{N} \rightarrow \mathbb{R}$ is the design objective. The set of linear equations $\boldsymbol{Cn} \leq \boldsymbol{d}$ represents molecular feasibility constraints and bounds on the vector $\boldsymbol{n}$.

In MINLP solution algorithms such as the OA[49] and the GBD[45] algorithms, a new combination of the integer variables is generated at each major iteration by solving a mixed integer linear program (MILP), the master problem. In conventional implementations, this combination is used to formulate a nonlinear problem (NLP), the primal problem, by fixing all integer variables to their values at the solution of the master problem. Thus, the primal problem for CAMPD is a nonlinear process-design problem (for a fixed solvent), whose solution is nontrivial. In our study, a modified OA algorithm is proposed, whereby each integer variable combination, corresponding to a different candidate solvent, is subjected to a series of tests prior to the solution of the primal problem, with the aim to facilitate its solution by removing infeasible points from the search space and by providing good initial values for key problem variables.

For a solvent $\boldsymbol{n}^{(k)}$ generated at major iteration k of the OA algorithm, we denote the feasible region of the corresponding primal problem by $\boldsymbol{X}^{FR(k)}=\{\boldsymbol{x}\in\boldsymbol{X}: \boldsymbol{h}(\boldsymbol{x},\boldsymbol{n}^{(k)})=0, \boldsymbol{g}(\boldsymbol{x},\boldsymbol{n}^{(k)})\leq 0\}$. The identification of the exact feasible region, $\boldsymbol{X}^{FR(k)}$, for each candidate solvent is a difficult problem in its own right, and the focus is placed on identifying a reduced process domain, $\boldsymbol{X}^{R(k)}$, such that $\boldsymbol{X}^{FR(k)}\subseteq\boldsymbol{X}^{R(k)}\subseteq\boldsymbol{X}$, by applying a series of tests. Thus, the tests are designed to overestimate the feasible region in order to avoid eliminating potential solutions. Only regions that can be detected *a priori* to be infeasible with respect to implicit and explicit process constraints are removed. This not only reduces the optimization search space, but also enhances the convergence of the solver during the solution of the primal problem. Furthermore, when $\boldsymbol{X}^{R(k)}=\varnothing$ for a molecule it is removed from the search space using an integer cut, and the solution of the primal problem for this candidate solvent is avoided.

Four tests are used in our current study to identify (and thus exclude) infeasible regions in the domain. Test 0 is used to identify a subdomain in which the feed is in the desired phase and to tighten user-provided bounds on the process domain. This test is independent of the solvent and only needs to be applied once at the beginning of the algorithm. The three other tests are applied at each iteration. Test 1 is used to determine whether the properties of the pure candidate solvent make it suitable for separation, that is, whether the solvent is a liquid at process temperatures, is safe and is feasible to handle. Test 2 is used to eliminate pressures at which the solvent and feed fail to form a two-phase mixture. Test 3 is used to eliminate pressures at which the treated gas leaving the absorber cannot be obtained at the required purity. If any of Tests 1, 2, or 3 are infeasible, the solvent is eliminated from the search. Tests 2 and 3 are posed as continuous NLPs. If these problems are feasible for the current solvent, they provide bounds as well as initial guesses for the solution of the primal problem. The information gained through the solution of the primal problem and the tests is used to formulate the next master problem and to generate a new solvent.

Test 0: Inlet stream phase stability after isenthalpic expansion

Feed streams in separation systems often undergo adjustments in conditions before entering a separation unit through temperature-change or pressure-change equipment. The aim of Test 0 is to identify the impact of these units on feasible conditions. Test 0 is described here by considering a gas stream undergoing an expansion. Consider a feed stream at pressure P_F, temperature T_F and composition $\mathbf{y}_F$ from which one component must be separated. The feed is expanded with an isenthalpic valve before entering an absorber with N stages. The pressure is thus reduced from P_F to the pressure P_{N+1} at the absorber inlet, as shown in Figure 1. User-defined ranges of allowable pressures and temperatures in the absorber are given by $[P^L_{N+1}, P^U_{N+1}]$ and $[T^L_{N+1}, T^U_{N+1}]$. In Test 0, the aim is to find a subdomain in the space defined by these ranges over which the inlet stream to the absorber is stable. The test is applicable to mixtures with a positive Joule-Thomson coefficient under the relevant conditions, a requirement which commonly holds for gases at ambient temperatures. For instance, the Joule-Thomson coefficient of methane and carbon dioxide is positive at room temperature over a wide range of pressures. The test is based entirely on a thermodynamic analysis of the feed stream alone and it is thus independent of the solvent.

A constraint implicit in most models of absorption columns is enforced in Test 0, namely that the stream to be separated must enter the absorber in the vapor phase. This is indeed necessary in practice for the feasible operation of the process. An evaluation of the process model may fail to converge when the stream at the vapor inlet of the absorber is in a two-phase state or is a liquid, which can imply there is no two-phase solution to the subset of equations that enforce vapor-liquid equilibrium in the column. Even if such an evaluation converges to the trivial solution of the phase-equilibrium equations, a change in the number of phases in the feed as the operating pressure changes during the solution of the primal problem introduces a discontinuity that usually causes the optimizer to fail to converge. Such discontinuities are averted by using Test 0 as it identifies *a priori* the region of the process domain where the feed is in the gas phase.

To illustrate the development of Test 0, the dew point curve for a binary mixture of CO_2 and methane at fixed mole fraction of CO_2 of 0.8 is shown in Figure 2. The maximum temperature at which two phases can occur for a stream of fixed composition, which is referred to as the cricondentherm, T_{cr}, is indicated by

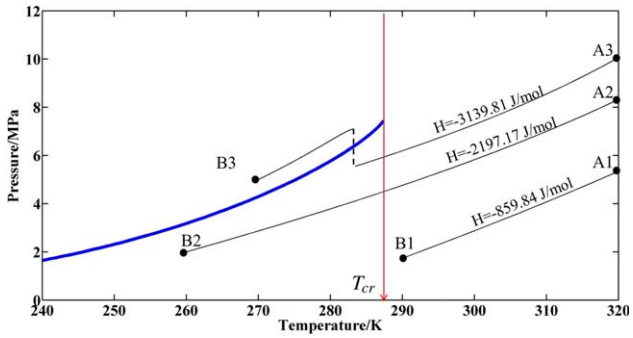

Figure 2. A diagram illustrating isenthalpic expansion for a mixture of methane and CO_2 with a constant total mole fraction of CO_2 of $y_{F_{CO_2}}=0.8$.

The arrow denotes the cricondentherm, T_{cr}. The thick solid curve denotes the dew pressure as a function of temperature, as calculated using the SAFT-γ Mie equation of state.[47] Isenthalpic curves (thin solid curves) denote adiabatic expansions from three points A*i*, $i=1,2,3$ to three points B*i*, $i=1,2,3$. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

the vertical arrow. We note that an isenthalpic expansion of a mixture with a positive Joule-Thomson coefficient, such as the mixture in Figure 2, results in a decrease in both pressure and temperature. Thus, when $P_{N+1} < P_F$, $T_{N+1} < T_F$ must hold.

The phases that can exist in the valve outlet stream (points B in Figure 2), which corresponds to the inlet to the absorber, depend on the value of its temperature, T_{N+1}, relative to T_{cr}. If T_{N+1} is greater than T_{cr}, the inlet to the absorber is in the gas phase. This is illustrated in Figure 2 for an isenthalpic expansion from A1 to B1. However, when $T_{N+1} \leq T_{cr}$, two situations can occur depending on the value of the dew point pressure, P_D, relative to the stream pressure, P_{N+1}: if $P_{N+1} < P_D(T_{N+1}, \boldsymbol{y}_F)$, the inlet stream is in the vapor region (expansion A2-B2 in Figure 2); if $P_{N+1} \geq P_D(T_{N+1}, \boldsymbol{y}_F)$, the expanded stream is either in the two-phase or the liquid region (expansion A3-B3 in Figure 2).

In order to avoid the two-phase region altogether, in Test 0 we use T_{cr} to set a lower bound on the temperature of the absorber inlet stream as:

$$T_{N+1}^{L0} = \max\left(T_{cr}, T_{N+1}^{L}\right), \tag{1}$$

where the value of T_{cr} may be obtained from the iterative solution of an isothermal flash problem at $\boldsymbol{y}_F$ until a temperature is found for which no dew pressure exists. This lower bound on the temperature is then used to derive a lower bound on the absorber pressure by considering an isenthalpic expansion from (P_F, T_F) to temperature T_{N+1}^{L0}. The pressure P_H following the expansion is obtained by equating the enthalpies at the inlet and outlet of the expansion valve

$$H(P_F, T_F, \boldsymbol{y}_F) = H(P_H, T_{N+1}^{L0}, \boldsymbol{y}_F). \tag{2}$$

The minimum allowable pressure in the absorber may then be found as

$$P_{N+1}^{L0} = \max\left(P_H, P_{N+1}^{L}\right). \tag{3}$$

A summary of Test 0 is given in Table 1. Unlike subsequent tests that depend on the solvent candidate $\boldsymbol{n}^{(k)}$, Test 0 is conservative in that it may remove some solutions at which $T_{N+1} \leq T_{cr}$ and the stream is nonetheless in the vapor phase. If the solution of the CAMPD problem is found to be at the lower bound on temperature or pressure, these conservative bounds can be relaxed.

Table 1. Procedure for Test 0

1. Calculate T_{cr} at composition $\boldsymbol{y}_F$.
2. Calculate T_{N+1}^{L0} using Eq. 1.
3. Calculate P_H using Eq. 2.
4. Calculate P_{N+1}^{L0} using Eq. 3.
5. Set $T_{N+1}^{L} = T_{N+1}^{L0}$ and $P_{N+1}^{L} = P_{N+1}^{L0}$.

Test 1: Solvent handling feasibility test

The feasibility of employing a given molecule as the solvent in an absorption process is evaluated in Test 1 based on pure-component properties independently of the process under consideration. The properties that are evaluated in this test are "essential properties," as previously defined by Harper et al.[59] These constraints form part of the overall design problem (P) and are an s-dimensional subset $\boldsymbol{g}_1(\boldsymbol{n})$ of $\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{n})$ such that $\boldsymbol{g}_1 : \boldsymbol{N} \rightarrow \mathbb{R}^s$. If they are linear, they are included in the master problem and, therefore, satisfied by the candidate solvent, but otherwise only an approximation is included in the master problem and the constraints may be violated by the candidate solvent. Because these constraints are independent of the process unknowns, they can readily be tested for feasibility before solving the primal problem. Four constraints are described in our current work: failure to meet any of these results in the elimination of the candidate molecule. Other process-independent nonlinear pure-component property constraints can readily be included in Test 1.

Prior to Test 1, the user specifies a solvent inlet temperature T_s, corresponding to the temperature at which fresh solvent enters the process, and a desired temperature handling range, $[T_{sh}^{L}, T_{sh}^{U}]$, corresponding to the temperatures at which the solvent may be stored or transported, and which may depend on ambient conditions. For solvent handling to be feasible, it is imperative for the solvent to be in the liquid state over the range of temperatures

$$[T_s^{L}, T_s^{U}] = [\min\left(T_{sh}^{L}, T_s\right), \max\left(T_{sh}^{U}, T_s\right)]. \tag{4}$$

It is generally expected that $T_s \in [T_{sh}^{L}, T_{sh}^{U}]$, but Eq. 4 ensures that the most appropriate bounds are set if this is not the case. Given the monotonic dependence of saturated-vapor pressure on temperature and the limited dependence of the melting line on pressure, a solvent that remains liquid over $[T_s^{L}, T_s^{U}]$ at atmospheric pressure can be assumed to remain liquid at higher pressures (unless of course very high pressures are considered).

Hence, the first constraint in Test 1 is that the normal melting point T_{mp} of the solvent is lower than T_s^{L}

$$T_{\mathrm{mp}}(P = 1\ \mathrm{atm}, \boldsymbol{n}^{(k)}) \leq T_s^{L}. \tag{5}$$

Furthermore, the normal boiling point T_{bp} of the solvent must be greater than T_s^{U}, and this is enforced by the second property constraint

$$T_s^{U} - T_{\mathrm{bp}}(P = 1\ \mathrm{atm}, \boldsymbol{n}^{(k)}) \leq 0. \tag{6}$$

In addition to verifying the liquid range of the solvent, the safety of handling the solvent is evaluated using its flash point T_{fp} at atmospheric pressure. The flash point must be greater than T_s^{U}, as expressed by the third constraint in Test 1

$$T_s^{U} - T_{\mathrm{fp}}(P = 1\ \mathrm{atm}, \boldsymbol{n}^{(k)}) \leq 0. \tag{7}$$

Finally, the last pure-component property criterion applied in this work is that the viscosity v of the solvent must not exceed v^{U}, the maximum viscosity that can be handled by the pump in the absorption plant. Assuming that the viscosity increases monotonically with decreasing temperature, it attains its maximum value at T_s^{L} for temperatures in the range $[T_s^{L}, T_s^{U}]$. Thus, the viscosity is evaluated at T_s^{L} in the fourth constraint in Test 1

$$v(T_s^{L}, P = 1\ \mathrm{atm}, \boldsymbol{n}^{(k)}) - v^{U} \leq 0. \tag{8}$$

In summary, Test 1 is an evaluation of problem (P1)

$$\begin{aligned} T_{\mathrm{mp}}(P = 1\ \mathrm{atm}, \boldsymbol{n}^{(k)}) - T_s^{L} &\leq 0 \\ T_s^{U} - T_{\mathrm{bp}}(P = 1\ \mathrm{atm}, \boldsymbol{n}^{(k)}) &\leq 0 \\ T_s^{U} - T_{\mathrm{fp}}(P = 1\ \mathrm{atm}, \boldsymbol{n}^{(k)}) &\leq 0 \\ v(T_s^{L}, P = 1\ \mathrm{atm}, \boldsymbol{n}^{(k)}) - v^{U} &\leq 0 \end{aligned} \tag{P1}$$

Test 2: Separation feasibility test

Test 2 is introduced to reduce the size of the process domain or of the molecular domain based on the thermodynamic

feasibility of the separation. No purity target is imposed, other than the implicit constraint that the desired product leaves the absorber in the gas stream. Test 2 can be formulated for gas-liquid or liquid-liquid separations and is presented here in the context of gas absorption. For a given solvent $\boldsymbol{n}^{(k)}$, the test can be used to identify a value of the pressure on the bottom stage of the absorber above which separation is not feasible. If no such value can be found above the lower bound on absorber pressure, the solvent can be eliminated. The test is based on the fact that the coexistence of two phases on stage N is a necessary condition to effect any separation, as the inlet stream to be separated enters the absorber at stage N and the loaded solvent leaves from stage N. Furthermore, from a modeling perspective, the presence of only one phase on stage N (or on any stage of the absorber) results in a discontinuity that can lead to numerical difficulties and it is therefore desirable to avoid carrying out process optimization in such cases.

Consider a counter-current absorption column with N stages as shown in Figure 1. Let the composition, flow rate, and temperature of the vapor stream that leaves from a given stage j be represented by $\mathbf{y}_j$, V_j and T_j, respectively and the composition, flow rate, and temperature of the liquid stream that leaves any stage j be represented by $\boldsymbol{x}_j$, L_j and T_j, respectively. The liquid stream entering the absorber on stage 1 is denoted by the subscript "0." The following simplifying assumptions are made to develop the test:

1. The composition of the solvent stream entering the counter-current column (at stage 1) is assumed to be known. In a process with solvent recycle, the exact composition of the solvent that enters the absorber is unknown. However, by assuming that the regeneration step leaves only small quantities of non-solvent components dissolved in the solvent, the composition of the solvent is set equal to that of a pure solvent for the purpose of this test alone. One may also argue that it must be feasible to operate the process with a pure solvent stream at plant start-up.
2. The feed to be treated is assumed to consist of two components only: the component to be removed is referred to as the "solute" and the component to be purified as the "product." If the feed stream consists of more than two components, the proposed test can be applied based on the two main components to be separated.
3. Stage N of the absorber is assumed to be an equilibrium stage.

We note that it is not necessary to assume that the two-phase region at given temperature and pressure is convex. The concepts of operating lines and difference points, developed for the design of ternary extraction systems by Hunter and Nash,[60] and discussed in Henley et al.,[61] are used in Test 2 to infer the conditions at which the separation is feasible. The difference point is a hypothetical stream[60] with "flowrate" $\triangle$ and "composition" $\boldsymbol{d}$ that can be defined with respect to any stage j in the column based on the vapor stream entering stage j and the liquid stream leaving that stage. It can be shown through overall and component mass balances that $\triangle$ and $\boldsymbol{d}$ are independent of j. These variables are defined by the following equations

$$\triangle = V_{j+1} - L_j, \quad \forall j = 0, \ldots, N \tag{9}$$

$$\triangle d_i = V_{j+1} y_{j+1,i} - L_j x_{j,i}, \quad \forall j = 0, \ldots, N, \quad \forall i = 1 \text{ to } NC \tag{10}$$

where NC is the total number of components. Note that a hypothetical stage corresponding to $j=0$ has been defined in these equations to represent the vapor stream leaving the column as V_1 and the clean solvent stream entering the column as L_0. Combining Eqs. 9 and 10, with $j=0$, to eliminate $\triangle$, and using assumption 1 to set $\boldsymbol{x}_0$ to the pure solvent composition, $\boldsymbol{x}_s$, one can derive the following relation

$$y_{1,i} = x_{s,i} L_0/V_1 + d_i(1 - L_0/V_1), \quad \forall i = 1, \ldots, NC. \tag{11}$$

Equation 11 indicates that $\mathbf{y}_1$ lies on the line joining $\boldsymbol{x}_s$ and $\boldsymbol{d}$. This is illustrated in Figure 3 for a representative ternary phase diagram for CO_2, methane and propyl-methyl ether as a solvent. In addition to $\boldsymbol{x}_s$ and $\mathbf{y}_1$, points $\boldsymbol{d}'$ and $\boldsymbol{d}''$ are placed for convenience on line $\overleftrightarrow{\mathbf{y}_1 \boldsymbol{x}_s}$, on either side of the ternary diagram. To further analyze the locus of difference points $\boldsymbol{d}$, two cases, shown as dashed lines in Figure 3, can be distinguished:

- When V_1 is greater than L_0, $\triangle$ is non-negative and the ratio L_0/V_1 is less than one. Thus, $\mathbf{y}_1$ must lie between $\boldsymbol{d}$ and $\boldsymbol{x}_s$, or equivalently, $\boldsymbol{d}$ must lie on the open ray $\overrightarrow{\mathbf{y}_1 \boldsymbol{d}''}$. The open ray is used to specify that the point $\mathbf{y}_1$ itself does not lie within the feasible locus of $\boldsymbol{d}$.
- Similarly, when V_1 is less than L_0, then $\triangle$ is negative and the locus of $\boldsymbol{d}$ is the open ray $\overrightarrow{\boldsymbol{x}_s \boldsymbol{d}'}$. This may be inferred by rearranging Eq. 11 as

$$x_{s,i} = y_{1,i} V_1/L_0 + d_i(1 - V_1/L_0), \quad \forall i = 1, \ldots, NC. \tag{12}$$

Thus, the locus of $\boldsymbol{d}$ consists of the two disconnected rays defined by excluding the closed line segment $\overline{\mathbf{y}_1 \boldsymbol{x}_s}$ from line $\overleftrightarrow{\mathbf{y}_1 \boldsymbol{x}_s}$.

The operating line for stage N may also be specified as

$$V_{N+1} y_{F,i} - L_N x_{N,i} = \triangle d_i, \quad \forall i = 1 \ldots NC. \tag{13}$$

It is apparent from Eq. 13 that $\mathbf{y}_F$, $\boldsymbol{x}_N$, and $\boldsymbol{d}$ are collinear. As $\boldsymbol{x}_N$ is the composition of the liquid stream leaving stage N, from assumption 3, it must be a point on the saturated-liquid curve and therefore line $\overleftrightarrow{\mathbf{y}_F \boldsymbol{d}}$ must intersect with the saturated-liquid curve. Thus, a necessary condition for absorption to be feasible is that the two-phase region should be large enough for a point $\boldsymbol{d}$ to exist such that $\overleftrightarrow{\mathbf{y}_F \boldsymbol{d}}$ intersects the saturated liquid curve at the stage pressure P_N and temperature T_N.

To define further the condition for which separation is feasible, consider the situation when $\overline{\mathbf{y}_F \boldsymbol{x}_s}$ is tangential to the saturated-liquid curve at $\boldsymbol{x}_N$, and does not intersect the saturated-liquid curve at any other point on the curve. θ is the angle the tangent makes with the horizontal in the clockwise direction (as shown Figure 3). Line $\overline{\mathbf{y}_F \boldsymbol{x}_s}$ is an infeasible operating line as $\boldsymbol{x}_s$ does not lie within the feasible locus of $\boldsymbol{d}$. Consider any other point on the saturated-liquid curve, $\boldsymbol{x}_N''$. Let $\overline{\boldsymbol{y}_F \boldsymbol{x}_N''}$ make an angle θ'' with the horizontal in the clockwise direction. It is easy to visualize and infer that $\theta'' > \theta$ and that $\overline{\boldsymbol{y}_F \boldsymbol{x}_N''}$ intersects the line segment $\overline{\mathbf{y}_1 \boldsymbol{x}_s}$ (at point $\boldsymbol{o}$ in Figure 3). However, such an operating line is infeasible as the feasible locus of the difference point excludes the line segment $\overline{\mathbf{y}_1 \boldsymbol{x}_s}$. Consider any operating line drawn with $\theta' < \theta$. Such an operating line is infeasible as it fails to intersect the two-phase region. Thus, separation becomes infeasible if $\overline{\mathbf{y}_F \boldsymbol{x}_s}$ is tangential to the two-phase region. Using the arguments outlined above, separation is also infeasible when the line $\overline{\mathbf{y}_F \boldsymbol{x}_s}$ falls above the two-phase region, that is, it does not intersect (and is not even tangential) to the two-phase region. This analysis holds for different types of phase diagrams and this is illustrated in Appendix A. Thus, there exists an operating line that connects to the locus of feasible difference points and that

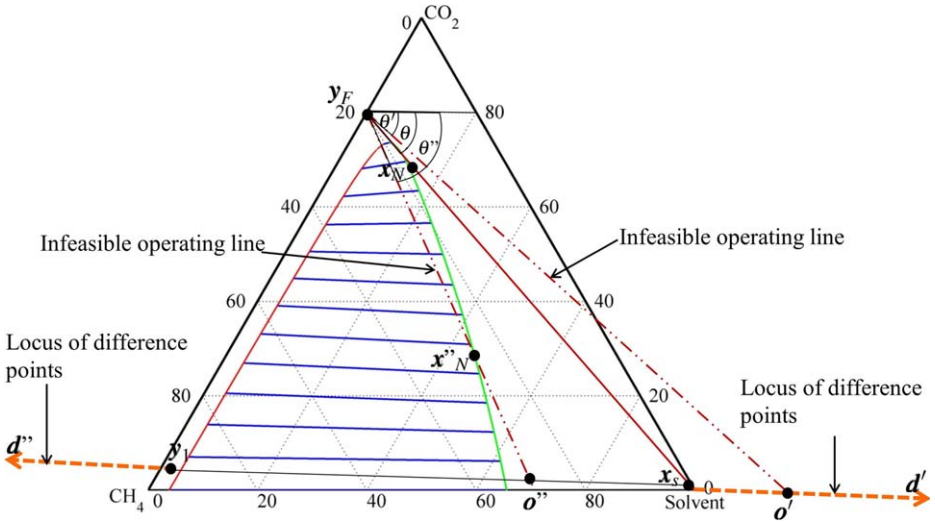

Figure 3. A phase diagram for CO_2-methane-solvent (propyl-methyl ether) at T_N= 304.4 K, P_{abs}= 9.0 MPa, illustrating the locus of difference points (dashed lines, points d', d'', o', and o'') and infeasible operating lines as discussed in the text.

y_1 is the composition of the gas stream leaving the absorber, y_F the composition of the feed stream entering the absorber and x_s the composition of the pure solvent stream entering the absorber. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

intersects the two-phase region if and only if the segment $\overline{y_F x_s}$ cuts through the two-phase region. Test 2 is based on searching for pressures at which this requirement is met.

Based on the analysis of difference points, Test 2 is formulated as a search for a maximum pressure $P_N^{U(k)}$ at which the line connecting the feed composition $\mathbf{y}_F$ and the pure solvent $\mathbf{x}_s$ intersects the two-phase region. If there is no such pressure, the separation is infeasible at all pressures and the solvent is removed from the search space. The optimization problem is given by

$$
\begin{aligned}
P_N^{U(k)} = \max_{P_N, T_N, \mathbf{y}_N, \mathbf{x}_N} \quad & P_N \\
\text{s.t.} \quad & \frac{y_{F,1}-x_{N,1}}{y_{F,2}-x_{N,2}} = \frac{y_{F,1}-x_{s,1}}{y_{F,2}-x_{s,2}} \\
& \mu_i^V(\mathbf{y}_N, T_N, P_N, \boldsymbol{n}^{(k)}) = \mu_i^L(\mathbf{x}_N, T_N, P_N, \boldsymbol{n}^{(k)}) \;\; \forall i = 1 \ldots NC \\
& \| \mathbf{y}_N - \mathbf{x}_N \|_2 \geq \epsilon \\
& \sum_{i=1}^{NC} x_{N,i} = 1 \\
& \sum_{i=1}^{NC} y_{N,i} = 1 \\
& P_{N+1}^L - \Delta P \leq P_N \leq P_{N+1}^U \\
& \max\left(T_{\text{mp}}(\boldsymbol{n}^{(k)}) + 10, T_N^L\right) \leq T_N \leq \min\left(T_F + 20, T_N^U\right) \\
& 0 \leq \mathbf{x}_N \leq 1 \\
& 0 \leq \mathbf{y}_N \leq 1
\end{aligned}
\tag{P2}
$$

where the first constraint defines a point $\mathbf{x}_N$ on the segment $\overline{y_F x_s}$, and the second constraint ensures this point is in equilibrium with a point $\mathbf{y}_N$, thereby lying on the two-phase boundary. μ_i^L and μ_i^V are the chemical potentials of component i in the liquid and vapor phases, respectively. The third constraint ensures that the composition vectors $\mathbf{x}_N$ and $\mathbf{y}_N$ that are obtained are not trivial solutions to the phase-equilibrium equations by setting ϵ to be a small positive number. In posing Problem (P2), bounds are imposed on the pressure P_N and temperature T_N. If the maximum pressure drop across stage N is ΔP, the lowest allowable value of the stage pressure P_N is $P_{N+1}^L - \Delta P$, where P_{N+1}^L is inherited from Test 0. In addition, an upper bound on P_N is given by $P_N < P_{N+1} \leq P_{N+1}^U$. The bounds on temperature can be set by the user as T_N^L and T_N^U, but constraints are also included to ensure that the lower bound is at least 10 K greater than the normal melting point of the solvent and the upper bound is at most 20 K greater than the feed temperature T_F. Problem

(P2) can be challenging to solve because it is infeasible for some pressures and it may thus be difficult to find a feasible direction from an infeasible point due to the high degree of nonlinearity of the problem. A more tractable reformulation of the problem is presented in Appendix B.

As Test 2 is based on thermodynamic feasibility only, Problem (P2) does not require the composition $\mathbf{y}_1$ of the treated gas stream to be specified and is based entirely on the feed specification. Furthermore, the condition of separation feasibility that is used here for a counter-current column is exactly the same as that for a single-stage separation unit. Separation in a single stage is possible when the total composition of a mixture formed by combining the feed and the solvent lies within the two-phase region.

Test 3: Purification feasibility

Most separation processes are designed with a constraint on the required purity of the treated stream and in Test 3 a thermodynamic analysis is used to eliminate conditions and solvents for which this constraint cannot be met. In the context of gas absorption, Test 3 can be used to find a lower bound $P_1^{L(k)}$ on the operating pressure at the top of the absorber that ensures that the separation can yield a vapor stream with the required purity while using solvent $\boldsymbol{n}^{(k)}$. If the purity criterion is found to be infeasible within the known pressure bounds, solvent $\boldsymbol{n}^{(k)}$ can be eliminated from the search space.

The temperature at stage 1 is denoted by T_1. The treated gas leaving the absorber with mole fractions $\mathbf{y}_1$ is required to have a mole fraction $y_{1,1}$ of product (component 1) of at least y_p and is assumed to be in equilibrium with a liquid stream that leaves stage 1. A necessary condition for the purification to be feasible is thus that there exists an equilibrium point $\mathbf{y}^*$ on the two-phase envelope such that the mole fraction of product y_1^* is greater than or equal to y_p at some temperature and pressure. Thus, the feasibility of achieving the required degree of separation is evaluated based on an analysis of the vapor-liquid envelope in relation to a process design constraint on product purity, y_p.

The example of a mixture of CO_2, methane, and propylmethyl ether is used once again in Figures 4a and 4b to illustrate the test. The shaded region in the figures represents the area where the mole fraction of the methane product in the treated gas, $y_{1,1}$, meets or exceeds the minimum acceptable purity of $y_p = 0.97$. At a pressure of 0.1 MPa and a temperature of 270 K (Figure 4a), the vapor-liquid boundary does not intersect the feasible region. When the pressure is increased to 0.610 MPa at the same temperature (Figure 4b), the saturated vapor curve passes through $y_p = 0.97$, indicating that a feasible pressure has been chosen.

In general, the test proposed here may be used to find the range of pressures over which the required purity criterion may be met. In our work, only a lower bound on pressure is sought by assuming the mole fraction of product (the purity) increases monotonically with pressure, that is, if a pressure is found at which the purity criterion is satisfied, then it is assumed to be satisfied at all higher pressures. However, if at higher pressures the purity constraint cannot be met (see e.g., Figure 6b in Appendix A where the maximum purity that can be obtained decreases with an increase in pressure) the test overestimates the feasible region. The test is formulated as follows

$$
\begin{aligned}
P_1^{L(k)} = \min_{P_1, T_1, \mathbf{y}_1, \mathbf{x}_1} \quad & P_1 \\
\text{s.t.} \quad & \mu_i^V(\mathbf{y}_1, T_1, P_1, \boldsymbol{n}^{(k)}) = \mu_i^L(\mathbf{x}_1, T_1, P_1, \boldsymbol{n}^{(k)}) \;\; \forall\, i = 1, \ldots, NC \\
& \| \mathbf{y}_1 - \mathbf{x}_1 \|_2 > \epsilon \\
& \sum_{i=1}^{NC} x_{1,i} = 1 \\
& \sum_{i=1}^{NC} y_{1,i} = 1 \\
& y_{1,1} \geq y_p \\
& 0 \leq \mathbf{x}_1 \leq 1 \\
& 0 \leq \mathbf{y}_1 \leq 1 \\
& P_1^L \leq P_1 \leq P_1^U \\
& \max\left(T_{\text{mp}}(\boldsymbol{n}^{(k)}) + 10, T_1^L\right) \leq T_1 \leq \min\left(T_1^U, T_F + 20\right)
\end{aligned}
\tag{P3}
$$

where $\mathbf{x}_1$ represents the composition of the liquid in equilibrium with a gas of composition $\mathbf{y}_1$ and μ_i^V and μ_i^L represent the chemical potentials in the vapor and liquid phases, respectively. The first constraint ensures that the two compositions lie on the vapor-liquid envelope. The second constraint is used as in Problem (P2) to ensure that $\mathbf{x}_1$ and $\mathbf{y}_1$ are distinct compositions at equilibrium rather than a trivial solution to the phase equilibrium equations. The bounds on pressure P_1 can be derived from Test 2 based on the pressure drop model adopted. The bounds on temperature are set in a similar manner to those in Test 2. Convergence to the solution of Problem (P3), which is highly nonlinear, can be achieved by using an initial guess within the feasible region for the problem. At the solution, the purity constraint, $y_{1,1} \geq y_p$, is typically active, and the constraint ensuring a minimum separation between the two equilibrium points is inactive. Hence, an initial guess within the feasible region can

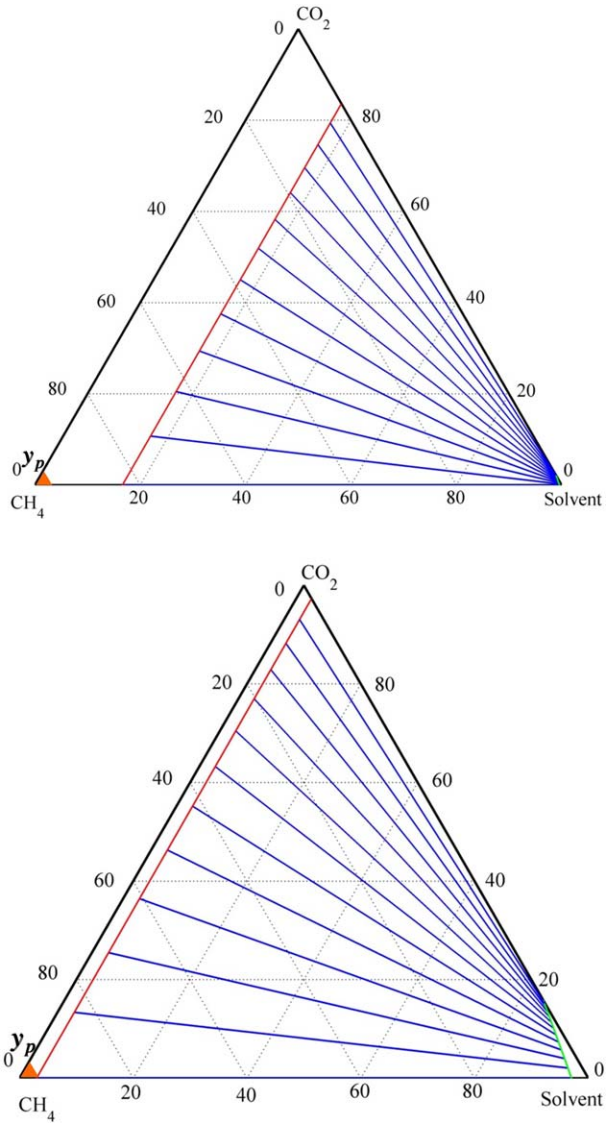

Figure 4. Phase diagram for CO_2-methane-solvent (propyl-methyl ether) at $T_1 = 270$ K and pressure P_1.

(a) $P_1 = 0.1$ MPa. (b) $P_1 = 0.610$ MPa. The shaded region represents $y_{1,1} \geq y_p = 0.97$. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

circumvent difficulties arising from the high degree of nonlinearity of the problem. Another strategy for solving this problem is to reformulate it as shown in Appendix C.

Finally, we note that the composition of the treated gas stream could of course be determined by solving the MESH equations for the N stages of the absorber. However, the use of Test 3 prior to such an evaluation allows a check to be performed based on the underlying phase-equilibrium model only, and it can lead to the *a priori* removal of regions of the domain where the purity constraint of the process cannot be met, without resorting to evaluating a more complex model.

Proposed CAMPD Algorithm

Overview of the algorithm

The outer-approximation algorithm[49,50] is modified to embed the tests presented in the previous sections. As in a standard OA framework, the primal and master problems are solved alternately. In the context of the general CAMPD problem (P), the primal problem at some iteration k consists of a process design problem for a fixed solvent $\boldsymbol{n}^{(k)}$. It is a continuous NLP that produces an upper bound on the optimal value of the objective function as well as information (optimal variable values, gradients and function values at the solution) that can be used to construct the master problem, a MILP. The solution of the master problem provides a lower bound on the optimal objective function and also yields a candidate solvent, $\boldsymbol{n}^{(k+1)}$, for the next iteration.

Primal Problem. In the proposed algorithm, as shown in Figure 5, Test 0 is applied once at the start of the algorithm, yielding updated lower bounds on P_{N+1} and T_{N+1}. These bounds are used throughout the algorithm. At each major iteration k of the algorithm, Tests 1 to 3 are solved sequentially. If any of these tests is infeasible, the algorithm proceeds directly to the solution of the master problem, which is formulated to embed some information from the failed test as described in detail in the next section on the master problem. If Test 1, Test 2, and Test 3 are all feasible for solvent $\boldsymbol{n}^{(k)}$, the primal problem is solved by following a two-step procedure which consists of initialization and solution. The initialization procedure is described further in Appendix D. Variable bounds for the primal problem are inherited from the solutions of Test 0, 2, and 3.

Before presenting the formulation of the primal problem, we note that the process and physical property models, as represented by equalities $\boldsymbol{h}(\mathbf{x}, \boldsymbol{n}) = 0$ in Problem (P), are treated as implicit constraints in the solution approach developed here. Hence, the variable set $\mathbf{x}$ is partitioned into a set of independent (decision) variables $\boldsymbol{u}$ and a set of dependent variables $\mathbf{x}^d$ so that $\mathbf{x} = (\boldsymbol{u}, \mathbf{x}^d)^T$. For fixed $\boldsymbol{u}$ and $\boldsymbol{n}$, $\boldsymbol{h}(\boldsymbol{u}, \mathbf{x}^d, \boldsymbol{n}) = 0$ thus represents a square system of equations of dimension $e \times e$ [cf., the definition of Problem (P)] that can equivalently be written as $\mathbf{x}^d = \mathbf{x}^d(\boldsymbol{u}, \boldsymbol{n})$.

This leads to the following formulation of the primal problem

$$
\begin{aligned}
f^k = \min_{\boldsymbol{u}} \quad & f(\boldsymbol{u}, \boldsymbol{n}^{(k)}) \\
\text{s.t.} \quad & \boldsymbol{g}_2(\boldsymbol{u}, \mathbf{x}^d(\boldsymbol{u}, \boldsymbol{n}^{(k)}), \boldsymbol{n}^{(k)}) \leq 0 \\
& \mathbf{x}^{dL(k)} \leq \mathbf{x}^d(\boldsymbol{u}, \boldsymbol{n}^{(k)}) \leq \mathbf{x}^{dU(k)} \\
& \boldsymbol{u}^{L(k)} \leq \boldsymbol{u} \leq \boldsymbol{u}^{U(k)}
\end{aligned}
\tag{P4}
$$

where $\boldsymbol{g}_2 \leq 0$ is the subset of inequality constraints obtained by removing the constraints used in Test 1 ($\boldsymbol{g}_1 \leq 0$) from the overall set of inequality constraints $\boldsymbol{g} \leq 0$ in Problem (P), and the superscripts L and U denote lower and upper bounds, respectively. The variable bounds may be specified by the user or inherited from the tests.

Furthermore, in the proposed formulation, the discrete choices corresponding to the number of groups of each type are represented by general integer variables rather than by binary variables as is common in the OA literature, with the exception of Fletcher and Leyffer.[50] This leads to a smaller number of variables in the problem: the number of discrete variables is reduced because it is not necessary to express each integer variable as a function of several binary variables and there is no need to introduce additional continuous variables and equations to represent the number of groups as a function of the relevant binary variables. Consequently, fewer gradients

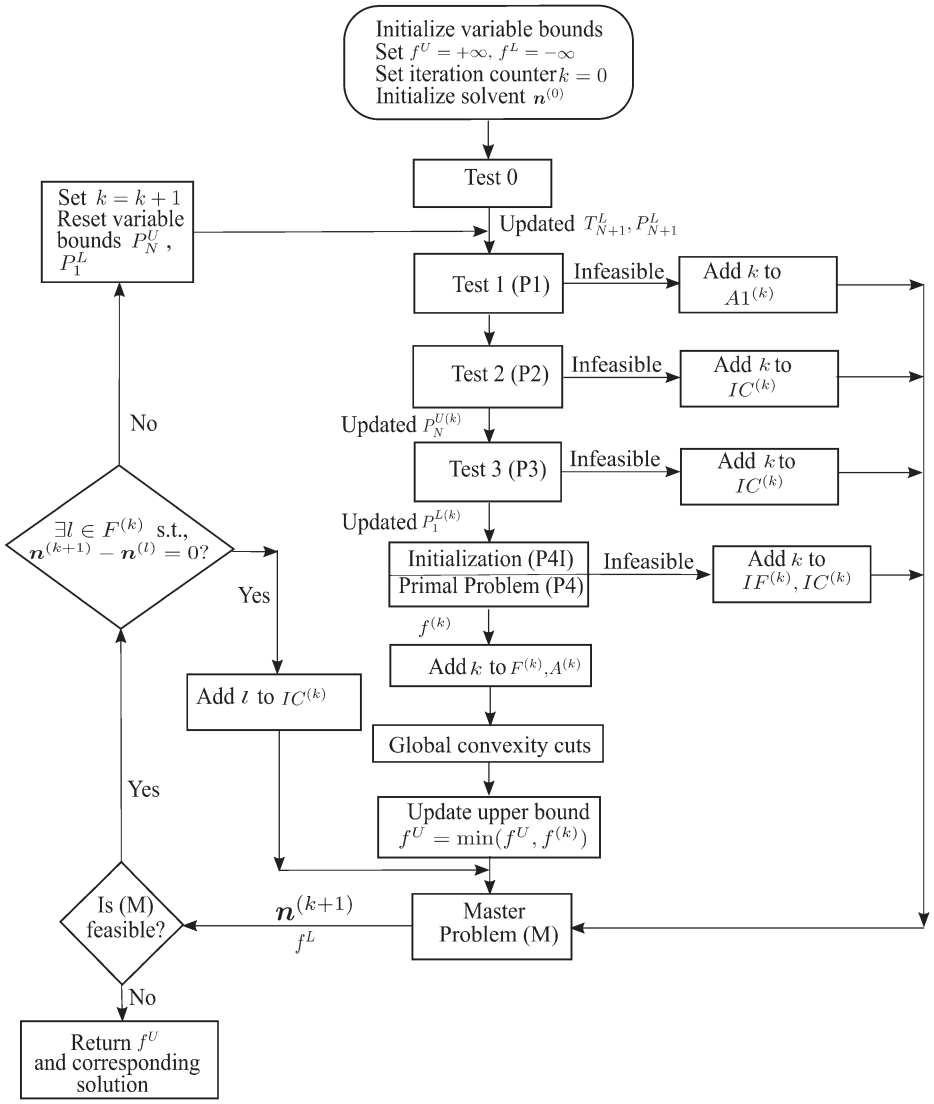

Figure 5. A flowchart of the proposed algorithm.

need to be evaluated when solving the primal problem and deriving linearized constraints for the master problem.

Master Problem. The exact formulation of the master problem depends on the outcome of Tests 1 to 3 and of (P4). If Tests 1, 2, and 3, and Problem (P4) are feasible, linearizations of the objective function and inequality constraints around the solution of the primal problem are added to the master problem. Several sets are defined in order to do so. $A1^{(k)}$ is a set used to keep track of all active and violated constraints in Test 1. It contains pairs of indices (l, j), where each j is the index of an active or violated constraint in (P1) at major iteration l, where $l \leq k$. A set $F^{(k)}$ is also defined such that each $l \in F^{(k)}$ is the index of a major iteration $l \in \{1, \ldots, k\}$ at which the primal was found to be feasible. For each $l \in F^{(k)}$, the value of $\boldsymbol{u}$ at the solution of the primal problem is denoted by $\boldsymbol{u}^{(l)}$, that is, $\boldsymbol{u}^{(l)} = \operatorname{argmin}_{\boldsymbol{u}} f(\boldsymbol{u}, \boldsymbol{n}^{(l)})$. Furthermore, the set $A^{(k)}$ contains pairs of indices (l, m) where $l \in F^{(k)}$ and each m is the index of an active constraint in $\boldsymbol{g}_2$ at the solution of Problem (P4) at major iteration l, thereby keeping track of all active constraints in (P4) at successful solutions of the primal problem. A constraint is declared active if $g_{2,m}(\boldsymbol{u}^{(l)}, \boldsymbol{n}^{(l)}) \geq \epsilon_a$, where ϵ_a is a small negative number (the absolute value of which is less than or equal to the feasibility tolerance). Finally, for each $l \in F^{(k)}$, where $l > 0$, once the primal problem is solved, global convexity tests[62] are employed. The constraints in the set $A^{(k)}$ of the master problem are evaluated with the integer variables fixed to $\boldsymbol{n}^{(l)}$ and the continuous variables to the solution $\boldsymbol{u}^{(l)}$. If any of the constraints are violated, this indicates it is an invalid underestimator of the non-convex feasible region,[62] and hence it is removed from set $A^{(k)}$.

If one of the tests or the primal problem is infeasible, the recurrence of the infeasible candidate solvent is prevented by introducing an integer cut in the master problem. There are several ways to formulate such an integer cut. A commonly used approach in the MINLP literature is the constraint proposed by Duran and Grossmann,[49] which applies to binary variables only, and therefore, cannot be applied to our formulation. A more general approach to integer cuts, which does not require the discrete variables to be binary, has been developed by Fletcher and Leyffer.[50] It involves the solution of a feasibility problem, a continuous optimization problem in which the discrete variables are fixed to the values corresponding to the infeasible combination. In the feasibility problem, the objective to be minimized is the

violation of the infeasible constraints, subject to the feasible constraints of the problem. Linearizations of the violated constraints at the solution of the feasibility problem may then be added to the master problem, to prevent recurrence of an infeasible combination. While this approach is general, the need to solve an additional optimization problem increases the computational cost. Thus, this is only applied to the constraints in Test 1, when Test 1 is infeasible. Since there are no continuous decision variables in (P1), linearizations of the violated constraints in Test 1 with respect to the integer variables can be added to the master problem without having to solve an NLP.

If one of Test 2, Test 3, or (P4) is infeasible, however, an integer cut based on the "Big-M" approach is added to the master problem for all subsequent iterations to remove the infeasible solvent. An integer cut is also applied if the master problem generates an integer combination that has previously been found to be feasible in problem (P4). Such a cut is not added at every iteration to prevent an unnecessary increase in the number of constraints and auxiliary variables. To formulate the integer cut, the set $IC^{(k)}$ is introduced to keep track of all major iterations l at which Test 2, Test 3 or the primal (P4) is infeasible ($l \leq k$), or at which the solution $\boldsymbol{n}^{(l'+1)}$ of the master problem is a repetition of a previously generated (feasible) integer combination, that is, there exists l' such that $\boldsymbol{n}^{(l')}=\boldsymbol{n}^{(l+1)}$, $l' \leq l \leq k$. The set of integer cuts takes the following form

$$M^L(1-y_l)+\epsilon_c \leq \sum_{i=1}^{q}\left(b^{i-1}\left(n_i-n_i^{(l)}\right)\right) \leq M^U y_l-\epsilon_c,\ \forall l \in IC^{(k)}. \tag{14}$$

where the vector $\boldsymbol{n}$, with elements n_i, describes the solvent being sought in the master problem, M^L is a large negative number, M^U a large positive number, b is a constant set such that $b > \max_{i=1,\ldots,q}(n_i^U)$, ϵ_c is a small positive number, and y_l is a binary variable introduced for iteration l, which ensures the central term is strictly positive or negative, but not equal to zero. The integer cut is applicable when each of the variables n_i is non-negative. We have used the sum of the products of n_i with powers of b to distinguish between two integer combinations. Alternatively, the sum of products of n_i with logarithms of prime numbers may be used as an integer cut.[63] Care must be taken in choosing M^L and M^U to be of sufficiently large magnitude to prevent spurious infeasibilities, while ensuring that the MILP can be solved successfully.

Additional constraints are constructed if problem (P4) is found to be infeasible starting from $\boldsymbol{u}_0^{(k)}$, the initial guess to Problem (P4) at iteration k. First, an integer cut is constructed for $\boldsymbol{n}^{(k)}$ and added to the master problem. Furthermore, when a feasible solution $\left(\boldsymbol{u}_0^{(k)}, \mathbf{x}^{d(k)}\right)$ to the set of equations $h(\boldsymbol{u}_0^{(k)}, \mathbf{x}^d, \boldsymbol{n}^{(k)})=0$ has been found, the objective function is linearized around $\left(\boldsymbol{u}_0^{(k)}, \boldsymbol{n}^{(k)}\right)$ and added to the master problem. Conversely, when no feasible solution to $h(\boldsymbol{u}_0^{(k)}, \mathbf{x}^d, \boldsymbol{n}^{(k)})=0$ is found, no further information is included in the master problem. This approach ensures that the algorithm proceeds despite a failure to solve the process model, without compromising convergence. The set $IF^{(k)}$ is defined as the set of iterations numbers $l \leq k$ at which the primal (P4) was found to be infeasible, but where a feasible solution to the set of equality constraints was found for solvent $\boldsymbol{n}^{(l)}$.

Finally, the master problem contains bounds on the molecular and process variables. Constraints $C\boldsymbol{n} \leq \boldsymbol{d}$ also ensure that molecular feasibility rules such as the octet rule[6] are satisfied by the molecule. The set of constraints that are required to be present is often dictated by the representation of the solvent-design space in the property prediction models. Molecular feasibility constraints are given in the case study section.

The formulation of the master problem at iteration k is given by

$$\begin{aligned}
\eta^{(k)} = \min_{u,n,\eta,y}\ & \eta \\
\text{s.t.}\quad & f(\boldsymbol{u}^{(l)}, \boldsymbol{n}^{(l)})+\nabla_{\boldsymbol{n}}^T f(\boldsymbol{u}^{(l)}, \boldsymbol{n}^{(l)})[\boldsymbol{n}-\boldsymbol{n}^{(l)}] \\
& +\nabla_{\boldsymbol{u}}^T f(\boldsymbol{u}^{(l)}, \boldsymbol{n}^{(l)})[\boldsymbol{u}-\boldsymbol{u}^{(l)}] \leq \eta,\ \ \forall l \in F^{(k)} \\
& f(\boldsymbol{u}^{0(l)}, \boldsymbol{n}^{(l)})+\nabla_{\boldsymbol{n}}^T f(\boldsymbol{u}^{0(l)}, \boldsymbol{n}^{(l)})[\boldsymbol{n}-\boldsymbol{n}^{(l)}] \\
& +\nabla_{\boldsymbol{u}}^T f(\boldsymbol{u}^{0(l)}, \boldsymbol{n}^{(l)})[\boldsymbol{u}-\boldsymbol{u}^{0(l)}] \leq \eta,\ \ \forall l \in IF^{(k)} \\
& g_{2,m}(\boldsymbol{u}^{(l)}, \boldsymbol{n}^{(l)})+\nabla_{\boldsymbol{n}}^T g_{2,m}(\boldsymbol{u}^{(l)}, \boldsymbol{n}^{(l)})[\boldsymbol{n}-\boldsymbol{n}^{(l)}] \\
& +\nabla_{\boldsymbol{u}}^T g_{2,m}(\boldsymbol{u}^{(l)}, \boldsymbol{n}^{(l)})[\boldsymbol{u}-\boldsymbol{u}^{(l)}] \leq 0,\ \ \forall (l,m) \in A^{(k)} \\
& g_{1,j}(\boldsymbol{n}^{(l)})+\nabla_{\boldsymbol{n}}^T g_{1,j}(\boldsymbol{n}^{(l)})[\boldsymbol{n}-\boldsymbol{n}^{(l)}] \leq 0,\ \ \forall (l,j) \in A1^{(k)} \\
& M^L(1-y_l)+\epsilon_c \leq \sum_{i=1}^{q}\left(b^{i-1}\left(n_i-n_i^{(l)}\right)\right) \leq M^U y_l-\epsilon_c, \\
& \qquad\qquad \forall l \in IC^{(k)} \\
& \eta^{(k-1)} \leq \eta \leq f^U-\epsilon_n \\
& \boldsymbol{u}^L \leq \boldsymbol{u} \leq \boldsymbol{u}^U \\
& C\boldsymbol{n} \leq \boldsymbol{d} \\
& y_l \in \{0,1\},\ \ \forall l \in IC^{(k)}
\end{aligned} \tag{M}$$

where $\eta^{(0)}=-\infty$, ϵ_n is a small positive number, and f^U is the lowest known objective function value.

Implementation

A fully automated implementation of the CAMPD algorithm presented in Figure 5 has been developed in C++, with an interface to gPROMS ModelBuilder 4.1.0[58] for the specification and solution of all the subproblems required to solve the primal problem, that is, Test 0 (Table 1), Problem (P1), Problem (P2a) (the reformulation of Problem (P2) in Appendix B), Problem (P3a) (the reformulation of Problem (P3) in Appendix C), and Problem (P4). The gORUN functionality of ModelBuilder is used to launch the solution of each of the subproblems as needed, using batch files. The solution files for each of the subproblems are read and the required information extracted and transferred to subsequent problems. For instance, the bounds derived by solving Problem (P2a) are embedded in Problems (P3a) and (P4) using the Foreign Object feature in gPROMS. In the same manner, the solution of Problem (P3a) provides a lower bound on the pressure in (P4) via a Foreign Object. The default continuous nonlinear optimizer in gPROMS, which is based on sequential quadratic programming, is used to solve problems (P2a), (P3a), and (P4). No attempt has been made to use a deterministic global optimization solver due to the scale and high degree of nonlinearity of the case studies considered. Problem (M) is formulated within the C++ code and solved using Gurobi 6.1,[64] which

comes with an inbuilt C++ interface. The gradients of the objective function and active inequality constraints with respect to the integer variables are calculated using first-order forward finite differences. At iterations where a feasible solution of the primal problem is found, the gradients of the objective function and active inequality constraints with respect to the continuous variables are obtained from the output of the gPROMS nonlinear optimizer. When the primal problem is infeasible, the gradients of the objective function with respect to the continuous variables are computed using first-order forward finite differences.

Several strategies are used within the implementation to enhance the robustness of the algorithm. While Tests 2 and 3 provide rigorous bounds on pressure if solved to global optimality, Problems (P2a) and (P3a) are nonconvex optimization problems that are solved using a local optimization solver within the current implementation. Hence, the tests may cut off feasible regions of the process and solvent domain. Furthermore, Problem (P4) is also nonconvex and may wrongly be found to be infeasible by a local solver. The likelihood of these issues arising is reduced in a practical way by taking a number of steps in problem formulation and initialization. These are described in more detail in Appendix D, where the initialization problem (P4I), also implemented in gPROMS, may be found. We note that initialization is an important element of robustness but that approaches to initialization other than that proposed in Appendix D may be adopted.

Case Studies

The flow sheet for the separation of carbon dioxide from natural gas that was described in the Motivating Example section (cf., Figure 1) is used to develop three case studies (C1, C2, and C3) and to explore the performance of the proposed methodology, in particular in terms of the effectiveness of the proposed tests and the robustness of the overall algorithm. Different specifications of the feed and process constraints are given in each case study, in recognition of the fact that natural gas streams vary with respect to concentration of CO_2, wellhead pressure, and temperature. Commercially-exploited natural gas has a wide range of concentrations of CO_2: from 0 to 90% CO_2.[65] For instance, large natural gas basins have been found in China that contain 80 to 97% of CO_2.[66] Indeed, the CO_2 content of a given field varies as a function of parameters such as drilling time and well depth. Similarly, the pressure P_F of the feed stream sent to the acid gas removal unit is affected by the natural gas processing steps, particularly whether the pressure is lowered for the separation of condensates from the gas. Although the specifications and constraints vary across the three case studies, the optimization variables remain the same, making it possible to investigate the effect of different specifications on the optimal solution.

Some assumptions are made in formulating the case studies. Although natural gas contains other hydrocarbons than methane, methane is used here as the key valuable component, so that the feed is a binary mixture of methane and carbon dioxide. It is further assumed that the pressure drop in the absorber is negligible, so that the absorber operates at a constant pressure P_{abs} with $P_{abs}=P_{N+1}=P_j$, $j=0,\ldots,N$. The process model, which is based on the MESH equations is that described by Pereira et al.[5] and Burger et al.[47] The degrees of freedom that are optimized are the absorber pressure P_{abs} the solvent flow rate L_0 and the solvent structure $\boldsymbol{n}$. The objective to be maximized is the net present value, NPV, of the carbon dioxide removal process over a 15-year period. The variables and specifications for the case studies are shown in Tables 2 and 3, while values assigned to constants that appear in (P2a), (P3a), and (M) are shown in Appendix E.

The space of possible solvents consists of linear compounds containing the groups CH_3, CH_2, eO, and cO, where eO and cO both consist of a single oxygen atom and are distinguished from one another by their position in the molecule. eO (end oxygen) describes an oxygen atom when bonded to one CH_3 and one CH_2 group, and cO (central oxygen) an oxygen atom when bonded to two CH_2 groups.[47] The solvent design space includes groups for which published interaction parameters with the components of the feed (CO_2 and CH_4) are available within the SAFT-γ Mie framework. Based on the bounds used on the molecular variables $\boldsymbol{n}$, a set of 109 molecules can be constructed. While this space is relatively small, this set of groups provides sufficiently varied phase behavior for a proof-of-concept study and makes it possible to enumerate all solutions to investigate the effectiveness of the proposed algorithm. The molecular feasibility constraints for this set of molecules ensure that every central oxygen atom, $n_{cO,}$ is in between two CH_2 groups to prevent the generation of molecules such as peroxides for which existing group parameters are expected to be ill-suited. Thus, the following two requirements need to be met: (1) if cO groups are present, the number of cO groups is less than that of CH_2 groups, (2) when n_{CH_2} is zero, n_{cO} is also zero. These conditions may be written compactly as

$$n_{cO} \leq n_{CH_2} - n_{CH_2}/n^U_{CH_2}. \tag{15}$$

Here, $n^U_{CH_2}$ represents the maximum number of CH_2 groups in the molecule. An end oxygen group can only be present if a CH_2 group is present. This may be written as the following constraint

$$n_{eO} \leq n_{CH_2} n^U_{eO}. \tag{16}$$

In order to carry out a systematic analysis of robustness, the three case studies are solved from the discrete starting points (solvent candidates) listed in Table 4. Initial guess IDs 1-6 were used as starting points for the CAMPD optimization of case study C1 in previous work[47] and have been repeated here for comparison. Solvent IDs 7-10 are introduced in the current work to test more extensively the effect of the initial guess on the solution procedure. Overall, the algorithm is applied to 30 different combinations of initial guesses and specifications.

The effectiveness of the tests is also investigated systematically over the entire solvent design space, and for the different specifications of the three case studies. Test 1 is applied to all solvents in the search space, Test 2 is only applied to solvents that pass Test 1, and Test 3 to solvents that pass Tests 1 and 2, in accordance with the sequential testing protocol used in the algorithm. Test 2 results in updating the upper bound on the absorber pressure for a solvent $\boldsymbol{n}^{(k)}$ only if a value $P^{U(k)}_{abs}$ that is lower than P^U_{abs}, the upper bound on pressure for the case study of interest, is found. Test 3 leads to an update of the pressure lower bound for a solvent $\boldsymbol{n}^{(k)}$ only if a pressure $P^{L(k)}_{abs}$ greater than P^L_{abs}, the lower bound identified in Test 0, is found. The values of the updated pressure bounds taken over all the solvents for which the pressure bounds have been successfully updated following Tests 2 or 3 are analyzed for each case study. Finally, the number of solvents for which a given test is active is investigated; a test is "active" for a given solvent if it can reduce the solvent domain (i.e., eliminate the solvent) or

Table 2. Variable Bounds Common to All Case Studies

Variable	Description	Units	Lower bound	Upper bound
n_{CH_3}	Number of CH_3 groups	–	2	2
n_{CH_2}	Number of CH_2 groups	–	0	8
n_{cO}	Number of cO groups	–	0	7
n_{eO}	Number of eO groups	–	0	2
L_0	Solvent flow rate entering absorber	kmol s^{-1}	0.01	50
V_{N+1}	Feed flow rate entering absorber	kmol s^{-1}	1	1
P_{abs}	Absorber pressure	MPa	0.1	Variable, cf. Table 3
P_{flash}	Pressure in the flash drum	MPa	0.1	0.1
P_{amb}	Pressure at the pure solvent inlet	MPa	0.1	0.1
T_{N+1}	Temperature of gas stream entering absorber	K	230	Feed temperature, T_F
T_N	Temperature on stage N	K	230	Variable, cf. Table 3
T_1	Temperature on stage 1	K	Variable, cf. Table 3	T_F+20
T_{sh}	Solvent handling temperature	K	298	308
T_s	Temperature of the solvent feed to the absorber	K	298	298
v	Solvent viscosity at standard pressure and T_s^L	Pa s	0	0.1

the process domain for that solvent (i.e., identify the process to be infeasible or identify updated pressure bounds).

Application of tests to entire solvent space

Case Study C1. Case study C1 is based on the process model and process constraints that were used in previous work.[47] The treated gas is required at a purity of 97 mol % of methane. The temperature in the flash drum is required to be 10 K above the melting point of the solvent. Other specifications used in case study C1 are summarized in Table 3.

The results of applying the tests to the specifications of this and other case studies are shown in Table 5. For case study C1, Test 0 yields the cricondentherm T_{cr} as 222 K and P_{abs}^L is unaffected at the relevant conditions. The performance of the other tests is shown in Table 5. Test 1 is the only test that results in the elimination of solvents for the specifications in this case study; it removes 19.3% of the search space.

Case Study C2. In case study C2, the separation of methane from a stream that has a high CO_2 content (92 mol %) is considered. The feed is available at a relatively high temperature of 340 K. The feed is assumed to be available through some compression or expansion process, which is outside the system boundary considered for this case study, at the absorber pressure P_{abs} which is an optimization variable. A larger domain for the absorber pressure is considered in this process, namely $0.1\,\text{MPa} \le P_{abs} \le 12.9\,\text{MPa}$. An additional constraint is placed on the process: that the temperature on stage N remains less than or equal to 325 K. Other constraints remain the same as in case study C1 (cf., Table 3).

Test 0 is not relevant to case study C2 as there is no isenthalpic valve. The performance of each of the other tests for case study C2 is reported in Table 5. Tests 1 and 2 are both effective in this case. Test 2 produces an upper bound on pressure such that $P_{abs}^{U(k)} \le P_{abs}^U$ for 20.5% of the solvent search space that passes Test 1 (88 molecules), as shown in Table 5.

Case Study C3. In case study C3, a feed at an intermediate CO_2 concentration (50 mol %) is considered. The purity constraint is tightened: 99 mol % of methane is required in the treated gas stream. An additional temperature constraint is also imposed, in which temperature T_1 is required to be greater than or equal to 298 K. The upper bound on absorber pressure P_{abs}^U is set equal to the pressure P_F of the feed. All other constraints remain the same as in case study C1.

Test 0 yields the cricondentherm T_{cr} as 260 K and P_{abs}^L is not updated by the test at conditions relevant to case study C3. The performance of each of the other tests for case study C3 can be seen in Table 5. Tests 1 and 3 are active in the case study, with Test 3 leading to the elimination of one solvent and an increase in the lower bound on absorber pressure for nine solvents. The improved bound on pressure that is provided by Test 3 results in a small reduction of the domain.

Application of the CAMPD algorithm

Overview of Results over 30 Runs. The proposed algorithm is applied to each of the three case studies from the ten starting points in Table 4. Throughout the discussion, average values relating to the performance are calculated as arithmetic means. All 30 runs converge successfully to locally optimal solutions. The results of applying the proposed algorithm to the case studies from the ten starting points are presented in Table 6 and the performance statistics are reported in Table 7. In case study C1, the best solvent found is $CH_3O(CH_2O)_5CH_3$. The algorithm converges to the same solution from each of the starting points. This a significant improvement in robustness compared to that observed in previous work.[47] Convergence of the algorithm was previously achieved from only 3 out of 6 starting points when attempting to solve CAMPD case study C1 using a standard MINLP algorithm without applying any tests. The solvent design space has been enumerated for case study C1 by solving an NLP for each solvent that passes Test 1. The best solution is found to be identical to that obtained with the proposed algorithm. Even with the relatively small design space, we note that the algorithm is remarkably effective at identifying a good solvent while evaluating only a small fraction (9.3%) of the space. This advantageous computational performance is expected to be even more marked when tackling problems with a larger design space.

The results of applying the algorithm to case study C2 from the 10 solvent starting points are also presented in Table 6. The non-convexity of the space is apparent in applying the

Table 3. Specifications that Vary Depending on the Case Study

Case Study	$y_{F_{CO_2}}$	T_F/K	P_F/MPa	y_p	P_{abs}^U	Valve	Extra constraint	T_N^U/K	T_1^L/K
C1	0.20	301.48	7.961	0.97	7.500	Yes	No	400	230
C2	0.92	340.00	$P_F=P_{abs}$	0.97	12.900	No	$T_N \le T_N^U$	325	230
C3	0.50	325.00	9.800	0.99	9.800	Yes	$T_1 \ge T_1^L$	400	298

Table 4. Initial Guesses Used to Solve Problems C1, C2, C3

Initial guess ID	$n^0_{CH_3}$	$n^0_{CH_2}$	n^0_{eO}	n^0_{cO}	L^0_0/ kmol s^{-1}	P^0_{abs}/ MPa
1	2	4	2	3	0.619	7.5
2	2	5	2	4	0.619	7.5
3	2	3	2	2	0.619	7.5
4	2	8	2	4	0.619	7.5
5	2	2	2	1	0.619	7.5
6	2	1	2	0	0.619	7.5
7	2	2	1	0	0.619	7.5
8	2	7	2	2	0.619	7.5
9	2	8	0	0	0.619	7.5
10	2	0	0	0	0.619	7.5

algorithm to case study C2. The chemical composition of the best solvent found is $CH_3O(CH_2)_7OCH_2OCH_3$. The algorithm converges to the highest *NPV* solution from only one starting point, and yields a high-performance solvent but with an *NPV* which is 7.5% lower, from six of the starting points. These two top ranking molecules differ in chemical structure by one oxygen atom, which highlights the strong interplay between solvent choices and process performance. Of the five solutions generated, four of these have the same number of CH_2 groups. With initial guess 1, the algorithm converges in three iterations to one of the lowest ranking solutions, and non-convexity is detected at the third iteration by the global convexity test[62] that is implemented in the proposed algorithm. On average, primal evaluations are attempted for only 7.8% of the solvent design space for case study C2.

Finally, the outcome of the ten runs for case study C3 is presented in the last row of Table 6. The algorithm converges to the solution with the highest *NPV* from each of the starting points. The best solvent found is $CH_3O(CH_2O)_7CH_3$. The average numbers of iterations and primal evaluations and their standard deviations attain their smallest value in this case study as may be seen in Table 7. The results indicate that on an average 4.3% of solvents in the design space are tested by the algorithm to arrive at the solution. The primal problem is evaluated for a mere 2.8% of the solvent design space. It can be seen in Table 6 that the value of the optimal pressure is at its upper bound P^U_{abs}, indicating that a more profitable process may be possible at higher pressures, although designing such a process would require taking into account the cost of compressing the feed.

Table 5. Effectiveness of Tests 1, 2, and 3 Over All 109 Solvents, for the Specifications of Each of the Case Studies

	Test 1	Test 2	Test 3
Case study C1			
Number of molecules tested	109	88	88
Number of molecules eliminated by test	21	0	0
Arithmetic mean of updated bound on pressure (MPa)	N/A	–	–
Number of molecules for which the test is active	21	0	0
Case study C2			
Number of molecules tested	109	88	88
Number of molecules eliminated by test	21	0	0
Arithmetic mean of updated bound on pressure (MPa)	N/A	11.45	–
Number of molecules for which the test is active	21	18	0
Case study C3			
Number of molecules tested	109	88	88
Number of molecules eliminated by test	21	0	1
Arithmetic mean of updated bound on pressure (MPa)	N/A	–	0.32
Number of molecules for which the test is active	21	0	10

Table 6. Locally Optimal Solution for Each Case Study from Ten Different Starting Points

Initial guess ID	$n^*_{CH_3}$	$n^*_{CH_2}$	n^*_{eO}	n^*_{cO}	L^*_0/ kmol s^{-1}	P^*_{abs}/MPa	*NPV*/ 10^9 USD
Case C1							
1–10	2	5	2	4	0.846	3.832	1.724
Case C2							
4	2	8	2	1	0.339	9.695	0.040
2,5,6,7,9,10	2	8	2	0	0.304	10.035	0.037
8	2	8	0	3	0.248	10.482	0.035
1	2	6	0	4	0.233	11.177	0.015
3	2	8	1	0	0.337	9.669	0.014
Case C3							
1–10	2	7	2	6	1.457	9.8	0.329

The * superscript denotes locally optimal solutions

It is instructive to compare the solutions obtained in the three case studies. The top ranking solvent found by the algorithm for each case study, the corresponding optimal process degrees of freedom, and objective function value are reported in Table 6 in the row immediately below the name of each case study. As can be seen, significantly different optimal solvents and process degrees of freedom have been found in the three variants of the problem statement. The results confirm that a strong interaction exists between the process objective and the choice of solvent, process variables, process specifications, and constraints.

Overall, we find the proposed algorithm exhibits robust performance in solving the CAMPD problem over 30 distinct optimization runs. The computational time taken to execute the tests is typically negligible compared to the CPU time required to solve the primal problem. As can be seen in Table 7 the algorithm explores 8.1% of the space of solvents on average in identifying a locally optimal solution. As discussed in the description of the algorithm, Tests 2 and 3 are only applied to molecules that pass Test 1. It is evident from Table 7 that different tests are active in each of the case studies. Whether a test is active, that is, useful in reducing the domain, cannot be predicted *a priori* and the tests therefore complement each other in increasing the robustness of the algorithm. The *a priori* detection of infeasibility arising from the choice of solvent molecule, which occurs chiefly due to Test 1, and in two cases

Table 7. Performance of the Algorithm and Tests for Case Studies C1, C2, and C3, Based on Aggregate Values over the 10 Starting Points for Each Case Study

	C1	C2	C3
Arithmetic mean of the number of major iterations	11.8	9.9	4.7
Smallest number of major iterations	4	3	3
Largest number of major iterations	17	15	7
Standard deviation of the number of major iterations	4.1	4.6	1.1
Test 0 active	No	No	No
Percentage of iterations with Test 1 active	14.4	14.1	29.8
Percentage of iterations with Test 2 active	0.0	11.1	0.0
Percentage of iterations with Test 3 active	0.0	0.0	6.4
Arithmetic mean of the number of attempted primal evaluations	10.1	8.5	3.1

The percentages of iterations over which a given test is active are calculated based on the total number of major iterations for the 10 runs for the relevant case study.

Table 8. Outcome of Test 0 for Three Different Feed Conditions (CO_2 Mole Fraction, $y_{F_{CO_2}}$, Feed Temperature, T_F, Feed Pressure, P_F)

Feed conditions			Test 0 outcome	
$y_{F_{CO_2}}$	T_F/K	P_F/MPa	T_{cr}/K	P_{abs}^{L0}/MPa
0.30	290	12.000	237	3.637
0.50	301	10.000	260	4.312
0.80	301	8.000	288	6.532

The cricondentherm, T_c, and updated pressure bound, P_{abs}^{L0}, after Test 0 are reported.

Table 9. Detailed Outcome of the Proposed Algorithm for the Third Iteration of the Solution of Case Study C2, Using Initial Guess ID 4 as a Starting Point

Problem	Status	Updated bound
Test 0	No update	–
Test 1	Passed	–
Test 2	Feasible	$P_{abs}^{U(2)}=12.06490$ MPa
Test 3	Feasible	$P_{abs}^{L(2)}=0.100095$ MPa
Primal	Process model: feasible Design constraints: infeasible	–

The solvent candidate is $n_{CH_3}^{(3)}=2$, $n_{CH_2}^{(3)}=2$, $n_{eO}^{(3)}=2$, $n_{cO}^{(3)}=1$.

due to Test 3, makes it possible to avoid expensive process evaluations and optimizations at infeasible points in 17.8% of all the major iterations carried out over the 30 distinct runs. The elimination of infeasible solvents is especially desirable during the first iterations of the MINLP solver: in these the optimizer has little information about the domain and may generate a number of poor solvent choices, leading to an increased risk of numerical failure and increased computational cost. In 16 out of 30 runs, although the initial guess solvent passed Test 1, the solvent generated for the second iteration failed Test 1. The impact of each test is investigated in more detail in the next section.

Analysis of Tests Within the CAMPD Algorithm. The statistics on active test instances shown in Table 7 depend on the specific sequence of solvent candidates generated by the algorithm, which in turn depends on the problem specifications and starting point. As a result, in some cases the tests are found to be active in fewer instances than in the studies of the overall solvent space presented in the section on the application of tests to the entire search space: this is true for Test 3, which is most active in case study C3, in 6.4% of iterations, but which can in principle be active for 9.2% of the overall search space. Test 2 is active for 11.1% of iterations for case study C2 but can be active for 16.5% of the solvent design space. However, the reverse can also be true. For example, Test 1 is active for 29.8% of iterations in case study C3 but can eliminate up to 19.3% of the overall search space (21 out of a 109 solvents).

Test 0 is the only test which is never active; this is a result of the feed specifications set for each of the case studies (cf., Table 3). The potential of Test 0 to reduce the process domain for different specifications is demonstrated in Table 8, where three new sets of feed conditions are used. The cricondentherm is shown in the table, together with the lower bound on absorber pressure obtained after the application of Test 0; this bound is increased significantly compared to the initial value of 0.1 MPa.

Tests 1, 2, and 3 not only lead to a reduction in the size of the process-solvent domain but also aid in enhancing the convergence of the CAMPD algorithm. In conjunction with the initialization procedure used to find a starting point for the primal problem (Problem (P4I)), the tests thus increase the robustness of the CAMPD algorithm. Test 1 prevents premature termination of the algorithm by eliminating molecules that do not satisfy process constraint $\boldsymbol{g}_1 \leq 0$ and that may lead to failure to solve the nonlinear model equations for some solvents that fail Test 1. For example, consider molecule $\boldsymbol{n}=[2,8,2,7]^T$, which is generated at the second major iteration for 19 out of the 30 runs. This molecule is predicted to have a normal melting of $T_{mp}(\boldsymbol{n})=302$ K, and therefore, to be a solid at T_s, under the conditions relevant to the three case studies. If the tests are not applied, and instead the molecule is used directly to the fix the integer variables in the primal problem for case study C1, the NLP solver fails to converge and the solution of the primal problem is thus inconclusive. It is possible to investigate the effectiveness of the different constraints in problem (P1). In all the major iterations in which Test 1 is active, we found that the melting point constraint (P1) is violated in 57.8% of the runs, whereas the flash-point constraint is violated in 42.2% of the runs. The other two constraints are never active.

To illustrate the application of the tests in more detail, the fourth major iteration from case study C2, with initial guess ID 4 used as a starting point, is shown in Table 9. The candidate solvent passes Test 1. Both Tests 2 and 3 are feasible and result in an updated upper bound on pressure. Indeed, absorber pressures between the updated pressure upper bound of 12.065 MPa and the initial pressure upper bound of 12.9 MPa are found to lead to failed process model evaluations. With the updated bounds, while the process model is evaluated successfully at the initial point, no solution that meets the design constraint is found so the primal problem is deemed infeasible. As a result, linearizations of the primal problem functions are constructed and incorporated in the master problem as described in the section on the proposed CAMPD algorithm.

In order to assess the impact of the non-convexity of Problems (P2a) and (P3a), we also verified the bounds obtained by the tests by constructing phase diagrams for a few solvent-CO_2-CH_4 systems, for the example shown in Table 9 as well as in other cases. We used HELD,[67] an algorithm that can reliably solve constant temperature and pressure flash problems, to determine stable equilibrium phases, and we constructed the relevant phase diagrams (including the diagrams in Figures 3 and 4). We found that even though we had used local solvers to arrive at the bounds on pressure, these bounds are consistent with the fluid phase-behavior of the mixtures in all cases tested.

While infeasible solvents may be eliminated thanks to the tests, the evaluation of the primal problem for other solvents can often fail. The use of an initialization strategy (here, in the form of Problem (P4I)) for the primal problem is an essential component of the proposed approach. Consider case study C3, with the initial guess of solvent structure set as follows: $\boldsymbol{n}^{(0)}=[2,5,2,4]^T$ (Initial guess ID 2, Table 4). Without taking any steps to identify a starting point for the process model, the standard MINLP solver terminates prematurely during its first iteration, in which $\boldsymbol{n}$ is relaxed to a continuous variable. This is due to a failure to evaluate the process model, which is avoided when a starting point is generated with the initialization problem. In some cases, even when the initialization problem is infeasible because there is no feasible path between the initialization solvent $\boldsymbol{n}_0$ and corresponding initial values, $\boldsymbol{u}_0$ and $\mathbf{x}_0^d$, and the desired solvent, $\boldsymbol{n}^{(k)}$ and $\boldsymbol{u}_0$, the nonlinear

optimization solver can succeed in identifying a solution to the primal problem.

Conclusions

A modified outer-approximation algorithm is proposed to solve CAMPD problems for separation systems, enabling the simultaneous optimization of solvent and process variables. The approach is developed to overcome the numerical challenges that arise due to the strong nonlinear interactions between process and solvent, with the aim to enhance robustness and increase the likelihood of identifying high-performance solutions. Four tests are embedded within an outer approximation algorithm to reduce the domain of solvent and process variables before attempting to find an optimal set of process variables for a specified solvent by solving the primal problem. In Test 0, the effect of adjustments to the feed conditions on the feasible region of the process is quantified. Specifically, a lower bound is obtained on the pressure that can be achieved through the isenthalpic expansion of a stream with a positive Joule-Thomson coefficient. In Test 1, molecules which violate nonlinear constraints on the pure-component properties of the solvent are eliminated, such as the feasibility of solvent storage and handling. In Test 2, the feasibility of achieving two phases at equilibrium at a specific stage of the separation unit is evaluated and an upper bound on feasible pressures is obtained. For each solvent that passes this test, the application of Test 3 provides an assessment of the feasibility of achieving the required purification of the feed and a lower bound on the feasible pressures. For each solvent that passes all tests, an initialization strategy is deployed prior to solution of the primal problem. This, combined with the updated bounds, reduces the likelihood of numerical failure during the solution of the primal. Finally, information from the tests and the solution of the primal problem is embedded in the master problem formulation to tighten the formulation and global convexity cuts are used to avoid including linearizations of the feasible region that are not valid underestimators of the nonconvex feasible region in the master problem. The specific formulation of some of tests is developed with a focus on the design of absorption-desorption systems and three case studies on the design of a process for the separation of methane and carbon dioxide, given different specifications, are chosen to investigate the performance of the proposed approach. Different optimal solvents and process conditions are identified for each case study, confirming the strong interactions between solvent and process design.

A systematic investigation of the performance of the proposed algorithm is undertaken by solving each case study from 10 different starting points, using a fully automated implementation. Convergence is successfully achieved in all 30 runs. The results show that the tests offer several benefits in terms of increased robustness and computational efficiency. The realization that a candidate solvent molecule is infeasible early on in a major iteration eliminates the need to solve the primal problem at that iteration. Over the 30 runs, only 8.1% of the solvent design space is probed. Thanks to the removal of some solvents by the tests, evaluations of the primal problem are attempted for only 6.6% of the solvent design space, highlighting the advantages of the proposed optimization-based method over enumeration. Given that the solvent design space considered is relatively small and focused on classes of molecules that are known to offer good separation performance for CO_2 and CH_4,[47] the approach can be expected to be even more effective for larger molecular design spaces.

For feasible molecules, the tests help to remove infeasible regions from the search domain and this can enhance convergence in a number of ways: some combinations of pressures and temperatures which favour the incidence of discontinuities such as the appearance and disappearance of phases in the absorber can be eliminated by using Test 0 and Test 2; optimizing over a reduced domain, thanks to Test 0, Test 2, or Test 3, may lead to a smaller number of iterations of the nonlinear optimization solver, decreasing computational cost; within the reduced domain, there is a reduced likelihood of encountering points where the nature of phase behavior (the number and types of phases) deviates from expected behavior; finally, initial guesses that are feasible from a thermodynamic perspective can be identified more readily. Further, the tests, which carry only a small computational cost, do not require the introduction of any further complexity into the process model itself, and thus existing implementations of process models can be used directly in the proposed algorithm.

Due to its increased robustness, the proposed methodology makes it possible to tackle highly nonlinear CAMPD problems without resorting to problem decomposition. This moves the focus of CAMPD away from making simplifying assumptions that make the problem tractable. In developing the process and thermophysical property models, emphasis can be placed instead on choosing the most appropriate model in terms of accuracy. The increased robustness also makes it possible to adopt a strategy in which the nonconvex MINLP is solved from different starting points, increasing the likelihood of identifying the global solution, a useful capability given that the use of deterministic global optimization techniques is not yet feasible for problems of this size and complexity, characterized by numerous highly nonlinear constraints. To the best of our knowledge, deterministic global optimization techniques have not been currently applied to problems that include equilibrium stage-based models of separation units with phase equilibrium described by rigorous thermodynamic models.[68] The impact of nonconvexities is clearly seen in one of the case studies in which the algorithm converges to very different solutions from different starting points.

There is scope to apply the proposed modified outer-approximation algorithm or tests to increase solution robustness to a wide range of design problems. Given that the tests are derived from a thermodynamic analysis of pure component and mixture behavior, they can readily be applied to solve property-based CAMD problems, where the process model is not embedded within the optimization formulation. They can also be used to facilitate process optimization even when the solvent molecule is fixed, by deploying Tests 0, 2, and 3 and the initialization strategy prior to solving the full process optimization problem. A similar test-based strategy may be applied to the design of other types of solvent-based separation systems, such as liquid-liquid extraction systems, through appropriate modifications of the proposed tests. For example, in the design of a liquid-liquid extraction system, the use of a test similar to Test 2 might eliminate solvents that are fully miscible with the feed. In an extractive distillation system, the use of a test based on Test 3 may allow for the screening of process conditions and entrainers with which the required distillate composition may be obtained. Finally, the set of tests can be expanded by incorporating additional implicit and explicit process constraints, just as in our current work.

Acknowledgments

The authors gratefully acknowledge financial support from the Engineering and Physical Sciences Research Council (EPSRC) grants EP/E016340, EP/J014958/1 and EP/J003840/1. SG is grateful for an Imperial College PhD Scholarship.

Literature Cited

1. Adjiman CS, Galindo A, Jackson G. Molecules matter: the expanding envelope of process design. In: Eden MR, Siirola JD, Towler GP, editors. *Proceedings of the 8th International Conference on Foundations of Computer-Aided Process Design*, vol. 34. *Computer Aided Chemical Engineering*. Elsevier B.V., Amsterdam, 2014:55–64. Available at: http://www.sciencedirect.com/science/article/pii/B9780444634337500079
2. Eden MR, Jørgensen SB, Gani R, El-Halwagi MM. A novel framework for simultaneous separation process and product design. *Chem Eng Process*. 2004;43:595–608.
3. Hostrup M, Harper PM, Gani R. Design of environmentally benign processes: integration of solvent design and separation process synthesis. *Comput Chem Eng*. 1999;23:1395–1414.
4. Pistikopoulos EN, Stefanis SK. Optimal solvent design for environmental impact minimization. *Comput Chem Eng*. 1998;22(6):717–733.
5. Pereira FE, Keskes E, Galindo A, Jackson G, Adjiman CS. Integrated solvent and process design using a SAFT-VR thermodynamic description: high-pressure separation of carbon dioxide and methane. *Comput Chem Eng*. 2011;35:474–491.
6. Odele O, Macchietto S. Computer aided molecular design: a novel method for optimal solvent selection. *Fluid Phase Equilib*. 1993;82(0):47–54.
7. Pretel EJ, Lpez PA, Bottini SB, Brignole EA. Computer-aided molecular design of solvents for separation processes. *AIChE J*. 1994;40(8):1349–1360.
8. Buxton A, Livingston AG, Pistikopoulos EN. Optimal design of solvent blends for environmental impact minimization. *AIChE J*. 1999;45(4):817–843.
9. Struebing H, Ganase Z, Karamertzanis P, Siougkrou E, Haycock P, Piccione P, Armstrong A, Galindo A, Adjiman C. Computer-aided molecular design of solvents for accelerated reaction kinetics. *Nat Chem*. 2013;5:952–957.
10. Apostolakou A, Adjiman CS. Chapter 4: Optimization methods in CAMD-II. In: Achenie LE, Gani R, Venkatasubramanian V, editors. *Computer Aided Molecular Design: Theory and Practice*, vol. 12. *Computer Aided Chemical Engineering*. Elsevier, 2002:63–93. Available at: http://www.sciencedirect.com/science/article/pii/S1570794603800068
11. Churi N, Achenie LEK. On the use of a mixed integer nonlinear programming model for refrigerant design. *Int Trans Oper Res*. 1997;4:45–54.
12. Papadopoulos AI, Stijepovic M, Linke P, Seferlis P, Voutetakis S. Toward optimum working fluid mixtures for organic rankine cycles using molecular design and sensitivity analysis. *Ind Eng Chem Res*. 2013;52(34):12116–12133.
13. Jonuzaj S, Akula PT, Kleniati PM, Adjiman CS. The formulation of optimal mixtures with generalized disjunctive programming: a solvent design case study. *AIChE J*. In press 2016. doi:10.1002/aic.15122
14. Jonuzaj S, Adjiman CS. Designing optimal mixtures using generalized disjunctive programming: hull relaxations. *Chem Eng Sci*. In press 2016.
15. Fredenslund A, Jones RL, Prausnitz JM. Group-contribution estimation of activity coefficients in nonideal liquid mixtures. *AIChE J*. 1975;21(6):1086–1099.
16. Papaioannou V, Lafitte T, Avendaño C, Adjiman CS, Jackson G, Müller EA, Galindo A. Group contribution methodology based on the statistical associating fluid theory for heteronuclear molecules formed from Mie segments. *J Chem Phys*. 2014;140(5):054107.
17. Majer C, Marquardt W, Gilles E. Reinitialization of DAE's after discontinuities. *Comput Chem Eng*. 1995;19(Supplement 1):507–512.
18. Yeomans H, Grossmann IE. Disjunctive programming models for the optimal design of distillation columns and separation sequences. *Ind Eng Chem Res*. 2000;39(6):1637–1648.
19. Kamath RS, Biegler LT, Grossmann IE. An equation-oriented approach for handling thermodynamics based on cubic equation of state in process optimization. *Comput Chem Eng*. 2010;34(12):2085–2096.
20. Skiborowski M, Harwardt A, Marquardt W. Efficient optimization-based design for the separation of heterogeneous azeotropic mixtures. *Comput Chem Eng*. 2015;72:34–51.
21. Eljack FT, Solvason CC, Chemmangattuvalappil N, Eden MR. A property based approach for simultaneous process and molecular design. *Chin J Chem Eng*. 2008;16:424–434.
22. Bommareddy S, Chemmangattuvalappil NG, Solvason CC, Eden MR. An algebraic approach for simultaneous solution of process and molecular design problems. *Braz J Chem Eng*. 2010;27:441–450.
23. Chemmangattuvalappil NG, Eden MR. A novel methodology for property-based molecular design using multiple topological indices. *Ind Eng Chem Res*. 2013;52(22):7090–7103.
24. Bardow A, Steur K, Gross J. Continuous-molecular targeting for integrated solvent and process design. *Ind Eng Chem Res*. 2010;49(6):2834–2840.
25. Bardow A, Steur K, Gross J. A continuous targeting approach for integrated solvent and process design based on molecular thermodynamic models. In: Nascimento C, Alves R, Biscaia E, editors. *10th International Symposium on Process Systems Engineering: Part A*, vol. 27. *Computer Aided Chemical Engineering*. Elsevier B.V., Amsterdam, 2009:813–818.
26. Gross J, Sadowski G. Perturbed-chain SAFT: An equation of state based on a perturbation theory for chain molecules. *Ind Eng Chem Res*. 2001;40(4):1244–1260.
27. Lampe M, Gross J, Bardow A. Simultaneous process and working fluid optimisation for organic rankine cycles (ORC) using PC-SAFT. *Comput Aided Chem Eng*. 2012;30:572–576.
28. Lampe M, Stavrou M, Schilling J, Sauer E, Gross J, Bardow A. Computer-aided molecular design in the continuous-molecular targeting framework using group-contribution PC-SAFT. *Comput Chem Eng*. 2015;81:278–287.
29. Karunanithi AT, Achenie LEK, Gani R. A new decomposition-based computer-aided molecular/mixture design methodology for the design of optimal solvents and solvent mixtures. *Ind Eng Chem Res*. 2005;44(13):4785–4797.
30. Roughton BC, Christian B, White J, Camarda KV, Gani R. Simultaneous design of ionic liquid entrainers and energy efficient azeotropic separation processes. *Comput Chem Eng*. 2012;42:248–262.
31. Kossack S, Kraemer K, Gani R, Marquardt W. A systematic synthesis framework for extractive distillation processes. *Chem Eng Res Design*. 2008;86(7):781–792.
32. Papadopoulos AI, Linke P. Multiobjective molecular design for integrated process-solvent systems synthesis. *AIChE J*. 2006;52(3):1057–1070.
33. Papadokonstantakis S, Badr S, Hungerbühler K, Papadopoulos AI, Damartzis T, Seferlis P, Forte E, Chremos A, Galindo A, Jackson G, Adjiman CS. Chapter 11 - Toward sustainable solvent-based post-combustion CO_2 capture: from molecules to conceptual flowsheet design. In: You F, editor. *Sustainability of Products, Processes and Supply Chains Theory and Applications*, vol. 36. *Computer Aided Chemical Engineering*. Elsevier B.V., Amsterdam, 2015:279–310. Available at: http://www.sciencedirect.com/science/article/pii/B9780444634726000112
34. Papadopoulos AI, Linke P. Efficient integration of optimal solvent and process design using molecular clustering. *Chem Eng Sci*. 2006;61:6316–6336.
35. Papadopoulos A, Linke P. A unified framework for integrated process and molecular design. *Chem Eng Res Design*. 2005;83:674–678.
36. Papadopoulos AI, Seferlis P, Linke P. A framework for the integration of solvent and process design with controllability assessment. *Chem Eng Sci*. In press 2016. Available at: http://www.sciencedirect.com/science/article/pii/S0009250916302135
37. Zhou T, Zhou Y, Sundmacher K. A hybrid stochastic-deterministic optimization approach for integrated solvent and process design. *Chem Eng Sci*. In press 2016. Available at: http://www.sciencedirect.com/science/article/pii/S0009250916301348
38. Jaksland CA, Gani R, Lien KM. Separation process design and synthesis based on thermodynamic insights. *Chem Eng Sci*. 1995;50(3):511–530.
39. Gil-Villegas A, Galindo A, Whitehead PJ, Mills SJ, Jackson G, Burgess AN. Statistical associating fluid theory for chain molecules with attractive potentials of variable range. *J Chem Phys*. 1997;106:4168–4186.
40. Galindo A, Davies LA, Gil-Villegas A, Jackson G. The thermodynamics of mixtures and the corresponding mixing rules in the SAFT-VR approach for potentials of variable range. *Mol Phys*. 1998;93:241–252.
41. Zhou T, McBride K, Zhang X, Qi Z, Sundmacher K. Integrated solvent and process design exemplified for a Diels-Alder reaction. *AIChE J*. 2015;61(1):147–158.
42. Cheng HC, Wang FS. Computer-aided biocompatible solvent design for an integrated extractive fermentation-separation process. *Chem Eng J*. 2010;162(2):809–820.
43. Exler O, Schittkowski K. A trust region sqp algorithm for mixed-integer nonlinear programming. *Optim Lett*. 2007;1(3):269–280.
44. Schilling J, Lampe M, Gross J, Bardow A. 1-stage CoMT-CAMD: an approach for integrated design of ORC process and working fluid

using PC-SAFT. *Chem Eng Sci*. In press 2016. Available at: http://www.sciencedirect.com/science/article/pii/S0009250916302196
45. Geoffrion A. Generalized benders decomposition. *J Optim Theory Appl*. 1972;10(4):237–260.
46. Giovanoglou A, Barlatier J, Adjiman CS, Pistikopoulos EN, Cordiner JL. Optimal solvent design for batch separation based on economic performance. *AIChE J*. 2003;49(12):3095–3109.
47. Burger J, Papaioannou V, Gopinath S, Jackson G, Galindo A, Adjiman CS. A hierarchical method to integrated solvent and process design of physical CO_2 absorption using the SAFT-γ Mie approach. *AIChE J*. 2015;61:3249–3269.
48. Gopinath S, Galindo A, Jackson G, Adjiman CS. A feasibility-based algorithm for computer aided molecular and process design of solvent-based separation systems. In: Kravanja Z, editor. *Proceedings of the 26th European Symposium on Computer Aided Process Engineering. Computer Aided Chemical Engineering*. Elsevier B.V., Amsterdam, 2016.
49. Duran MA, Grossmann IE. An outer-approximation algorithm for a class of mixed-integer nonlinear programs. *Math Program*. 1986; 36(3):307–339.
50. Fletcher R, Leyffer S. Solving mixed integer nonlinear programs by outer approximation. *Math Program*. 1994;66(1-3):327–349.
51. Chapman WG, Gubbins KE, Jackson G, Radosz M. New reference equation of state for associating liquids. *Ind Eng Chem Res*. 1990; 29:1709–1721.
52. Chapman WG, Gubbins KE, Jackson G, Radosz M. SAFT: equation-of-state solution model for associating fluids. *Fluid Phase Equilib*. 1989;52:31–38.
53. McCabe C, Galindo A. Chapter 8: SAFT associating fluids and fluid mixtures. In: Goodwin ARH, Sengers JV, Peters CJ editors. *Applied Thermodynamics of Fluids. The Royal Society of Chemistry. Cambridge*, 2010:215–279. Available at: http://dx.doi.org/10.1039/9781849730983-00215
54. Kontogeorgis GM, Folas GK. *Thermodynamic Models for Industrial Applications: From Classical and Advanced Mixing Rules to Association Theories*. New York: Wiley, 2010.
55. Hukkerikar AS, Sarup B, Kate AT, Abildskov J, Sin G, Gani R. Group-contribution$^+$ (GC^+) based estimation of properties of pure components: Improved property estimation and uncertainty analysis. *Fluid Phase Equilibria*. 2012;321:25–43.
56. Sastri S, Rao K. A new group contribution method for predicting viscosity of organic liquids. *Chem Eng J*. 1992;50:9–25.
57. Macleod DB. On a relation between surface tension and density. *Trans Faraday Soc*. 1923;19:38–41.
58. Process Systems Enterprise. gPROMS. 1997−2015. Available at: www.psenterprise.com/gproms
59. Harper PM, Gani R, Kolar P, Ishikawa T. Computer-aided molecular design with combined molecular modeling and group contribution. *Fluid Phase Equilib*. 1999;158160(0):337–347.
60. Hunter TG, Nash AW. The application of physico-chemical principles to the design of liquid-liquid contact equipment. Part II. Application of phase-rule graphical methods. *J Soc Chem Ind*. 1934.
61. Henley EJ, Seader JD, Roper DK. *Separation Process Principles, 3rd ed, International student version*. John Wiley & Sons, Asia, 2011.
62. Kravanja Z, Grossmann IE. New developments and capabilities in PROSYN-an automated topology and parameter process synthesizer. *Comput Chem Eng*. 1994;18(1112):1097–1114.
63. Samudra AP, Sahinidis NV. Optimization-based framework for computer-aided molecular design. *AIChE J*. 2013;59:3686–3701.
64. Gurobi Optimization, Inc. Gurobi optimizer reference manual, 2015. Available at: http://www.gurobi.com
65. Tan L, Shariff A, Lau K, Bustam M. Impact of high pressure on high concentration carbon dioxide capture from natural gas by monoethanolamine/n-methyl-2-pyrrolidone solvent in absorption packed column. *Int J Greenhouse Gas Control*. 2015;34:25–30.
66. Baojia H, Xusheng L, Xianming X. Origin and accumulation of CO_2 in natural gases of the Yinggehai-Qiongdongnan basins, offshore South China Sea. Paper presented to AAPG Hedberg Conference-"Natural Gas Geochemistry: Recent developments, applications, and technologies", 2011, Beijing, China. http://www.searchanddiscovery.com/pdfz/abstracts/pdf/2011/hedberg-beijing/abstracts/ndx_huang02.pdf.html.
67. Pereira FE, Jackson G, Galindo A, Adjiman CS. The HELD algorithm for multicomponent, multiphase equilibrium calculations with generic equations of state. *Comput Chem Eng*. 2012;36:99–118.
68. Skiborowski M, Rautenberg M, Marquardt W. A hybrid evolutionary-deterministic optimization approach for conceptual design. *Ind Eng Chem Res*. 2015;54(41):10054–10072.

Appendix A

Pressure can affect fluid-phase behavior in different ways depending on the mixture under consideration. In order to illustrate different situations, and given assumption 2, four types of ternary phase diagrams that may be observed are sketched in Figure 6, at fixed pressure and temperature, illustrating instances in which different pairs of compounds are partially miscible. In Figure 6a, only the solvent-product binary mixture exhibits vapor-liquid equilibrium (as in Figure 3); in Figure 6b, only the solvent-solute pair exhibits phase separation; in Figure 6c, the two binary pairs of solvent-solute and solvent-product exhibit vapor-liquid equilibrium; finally, in Figure 6d, the two binary pairs of product-solute and product-solvent are partially miscible.

For a mixture at conditions P, T_N that exhibits behavior of the types shown in Figures 3 and 6a, a further increase in pressure may result in a decrease in the size of the vapor-liquid region. Thus, for systems that exhibit full miscibility of the solvent and the solute the necessary condition identified in Test 2 may be used to find the maximum feasible value of pressure. Consider a mixture that exhibits partial miscibility of the solute-solvent pair and full miscibility of the product-solvent pair (cf., Figure 6b). At some pressure $P' > P$, let the vapor-liquid envelope at P', shown by the dashed curve in Figure 6b, result in a vapor leaving stage N that has a lower concentration of product than the feed. Hence, the operation of the stage at P' is infeasible with respect to achieving separation in the desired direction. As $\overline{y_F x_s}$ does not intersect the curve at conditions P', T_N, the necessary condition derived here is not satisfied at P', but is satisfied at some feasible pressure P, where $P' > P$. Thus, the application of the test to systems of the type shown in Figure 6b can also yield a maximum pressure beyond which separation is infeasible. Next, consider mixtures that exhibit partial miscibility of the solvent-solute and the solvent-product pairs (cf. Figure 6c) at all pressures and temperatures within the process domain. The line $\overline{y_F x_s}$ intersects the two-phase region at the conditions P, T_N. Thus, as these mixtures satisfy the necessary condition identified, all pressures up to the user-defined upper bound may be found to be feasible. Similarly, for systems that exhibit partial product-solute miscibility, as in Figure 6d, if line $\overline{\mathbf{y}_F \mathbf{x}_s}$ intersects the two-phase region at P, T_N, pressure P is feasible. For mixtures for which the size of the two-phase region increases with pressure, the necessary condition of Test 2 may yield a maximum feasible pressure at or above the user-defined upper bound on allowable pressures, offering no improvement over the user-defined bound. Although ternary mixtures may exhibit more than one type of phase diagram as the conditions of pressure and temperature are varied, the necessary condition for Test 2 is valid for each of type of ternary diagram and, therefore, across different conditions.

Appendix B

To reformulate Problem (P2), the objective function is multiplied by $\tanh\left(\beta(\| \mathbf{y}_N - \mathbf{x}_N \|_2 - \epsilon)\right)$. Here, β is a positive scaling factor to ensure that the tanh function yields values very close to unity when its argument is positive. When the third constraint in (P2), which ensures that two distinct phases are obtained, is violated, the following holds

$$\| \mathbf{y}_N - \mathbf{x}_N \|_2 - \epsilon < 0$$

and the objective function is thus multiplied by a negative number. This ensures that no increasing direction for the objective function is found when no two-phase solution is found. Thus, the solver converges to a solution in the neighborhood of $P_N^{U(k)}$. The formulation is given by

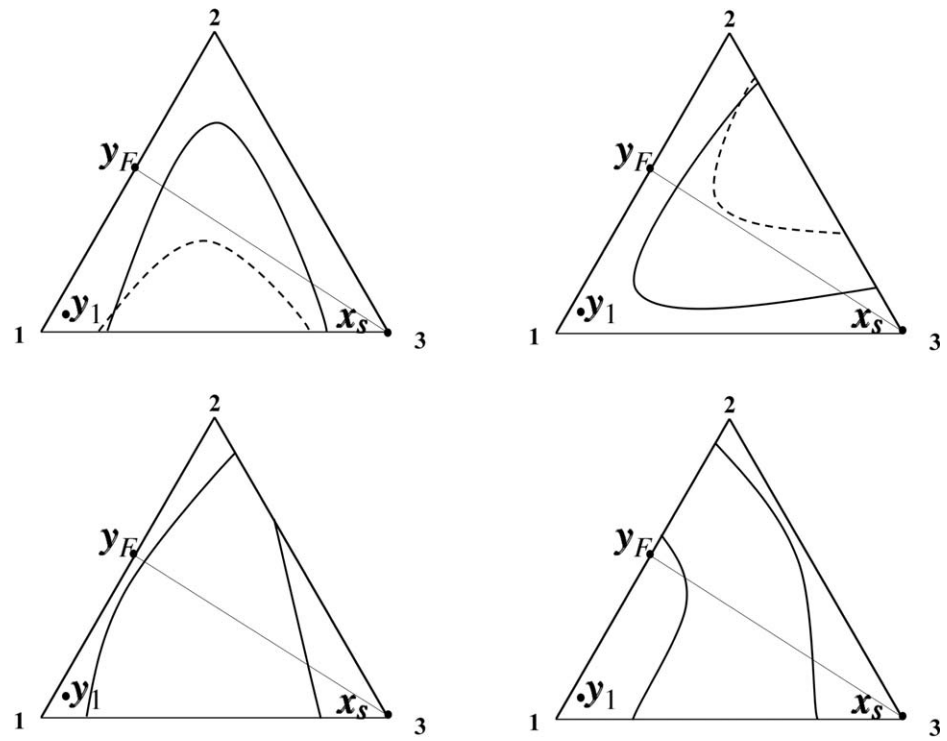

Figure 6. Four types of ternary phase diagrams for a product (1), solute (2), and solvent (3) at constant pressure and temperature.

(a) The solvent and product pair is partially miscible and other pairs are fully miscible; (b) The solvent and solute pair is partially miscible and other pairs are fully miscible; (c) The solvent and solute pair is partially miscible and the solvent and product pair is partially miscible; (d) The product and solute pair is partially miscible and the solvent and product pair is partially miscible. y_F denotes the feed composition, y_1 denotes the product composition and x_s the lean solvent composition. The thick curves denote the vapor-liquid envelope at pressure P. The dashed curves indicate the vapor-liquid envelope at a pressure P', such that $P'>P$.

$$
\begin{aligned}
P_N^{U(k)} = \max_{P_N, T_N, \mathbf{y}_N, \mathbf{x}_N} \quad & P_N \tanh\left(\beta(\| \mathbf{y}_N - \mathbf{x}_N \|_2 - \epsilon)\right) \\
\text{s.t.} \quad & \frac{y_{F,1} - x_{N_1}}{y_{F,2} - x_{N,2}} = \frac{y_{F,1} - x_{s,1}}{y_{F,2} - x_{s,2}} \\
& \mu_i^V(\mathbf{y}_N, T_N, P_N, \boldsymbol{n}^{(k)}) = \mu_i^L(\mathbf{x}_N, T_N, P_N, \boldsymbol{n}^{(k)}) \quad \forall\, i = 1, \ldots, NC \\
& \sum_{i=1}^{NC} x_{N,i} = 1 \\
& \sum_{i=1}^{NC} y_{N,i} = 1 \\
& \| \mathbf{y}_N - \mathbf{x}_N \|_2 \geq \epsilon \\
& P_{N+1}^L - \Delta P \leq P_N \leq P_{N+1}^U \\
& \max\left(T_{\text{mp}}(\boldsymbol{n}^{(k)}) + 10, T_N^L\right) \leq T_N \leq \min\left(T_F + 20, T_N^U\right) \\
& 0 \leq \mathbf{x}_N \leq 1 \\
& 0 \leq \mathbf{y}_N \leq 1
\end{aligned}
\tag{P2a}
$$

Appendix C

In order to solve for phase equilibrium more robustly for Test 3, the TPFlash routine in gSAFT, an isothermal-isobaric flash routine, is used. The routine yields compositions and thermodynamic properties of up to three phases in equilibrium, denoted by mole fractions **y**, **xa**, and **xb**. If a phase does not exist, the corresponding compositions are set equal to zero in the output of this flash routine. Two phases in equilibrium may either be ordered as **y** and **xa**, or as **xa** and **xb** in the output of the flash routine.

In the input to the flash routine we used an overall composition of z, chosen to be a point very close to the product-solvent boundary. An overall composition on that boundary, such that z_1 is smaller than y_p, is one most likely to result in an equilibrium composition of the vapor phase that is rich in the product. z_1 was set at 0.9 in this study, where the required purity y_p ranges from 0.97 to 0.99 assuming the liquid

boundary has compositions of the product lower than 0.9. Hence, if two-phases exist at y_p or greater, the overall composition z will split into two. Alternatively, z_1 could be set equal to y_p, thus requiring no assumption on the liquid boundary

$$
\begin{aligned}
P_1^{L(k)} = \min_{P_1,T_1} \quad & P_1 \\
s.t. \quad & z=[0.9 \;\; \epsilon \;\; 0.1-\epsilon]^T \\
& [\mathbf{y} \;\; \mathbf{xa} \;\; \mathbf{xb}]^T = \mathrm{TPFlash}(z, T_1, P_1, \boldsymbol{n}^{(k)}) \\
& \max(y_1, xa_1, xb_1) > y_p \\
& \| \mathbf{y}-\mathbf{xa} \|_2 > \epsilon \\
& \| \mathbf{xa}-\mathbf{xb} \|_2 > \epsilon \;\; 0.5 \le \Sigma_i^{\mathrm{NC}} (y_i + xb_i) \le 1.5 \\
& P_1^L \le P_1 \le P_1^U \\
& \max(T_{\mathrm{mp}}(\boldsymbol{n}^{(k)})+10, T_1^L) \le T_1 \le \min(T_F+20, T_1^U)
\end{aligned}
\tag{P3a}
$$

Appendix D

The following steps are taken to overcome numerical issues arising from the nonlinearity and nonconvexity of Problems (P2a), (P3a), and (P4):

1. The robust algorithm for constant pressure and temperature flash calculations implemented within gPROMS is used in Test 3 instead of including the necessary conditions for phase equilibrium as part of the model equations. Note that while the flash algorithm was used in (P3a), it was not used in (P2a) or (P4). In the gPROMS modeling platform, which follows a feasible path approach with respect to equality constraints, a variable is either a degree of freedom (or input) or its value is obtained by the solution of a system of equations. The flash algorithm requires temperature, pressure, and total composition as inputs. In Test 3, all of the inputs to the flash methods are degrees of freedom, hence it is straightforward to use the flash algorithm to solve for phase equilibrium. However, in an initial analysis, the use of the flash algorithm seemed to make the solution of problems (P4) and (P2a) less robust as in these the temperature and composition variables, respectively, are not degrees of freedom.
2. Starting points that are likely to be feasible for the two tests and Problem (P4) are generated at each iteration. For Test 2, the initial guess of pressure is set at P_N^L, a pressure which is most likely to be feasible. The initial guess of temperature is set as $\min(T_F+20, T_N^U)$. A flash problem is then solved for a mixture with total composition equal to the arithmetic mean of the feed and pure solvent compositions and at the initial guess of pressure and temperature. The equilibrium compositions, if they exist, are then used to initialize problem (P2a). For Test 3, the initial guess of T_1 is set at its lower bound as it is most likely to be feasible. The pressure on the other hand, is set at its upper bound. For Problem (P4), given a known solution $(\boldsymbol{u}_0, \boldsymbol{n}_0, \mathbf{x}_0^d)$ to the model equations, an initial guess for $\mathbf{x}^d$ at iteration k, where the solvent is given by $\boldsymbol{n}^{(k)}$, is obtained by solving the following problem

$$
\begin{aligned}
& \boldsymbol{h}(\boldsymbol{n}(t), \mathbf{x}^d(t), \boldsymbol{u}_0) = 0 \\
& \frac{d\boldsymbol{n}(t)}{dt} = 0.001(\boldsymbol{n}^{(k)} - \boldsymbol{n}_0) \\
& \boldsymbol{n}(0) = \boldsymbol{n}_0 \\
& \boldsymbol{n}(t=1000) = \boldsymbol{n}^{(k)}
\end{aligned}
\tag{P4I}
$$

so that $(\boldsymbol{u}_0, \boldsymbol{n}^{(k)}, \mathbf{x}^d(t=1000))$ can be used as an initial guess for (P4). Problem (P4I) is a differential-algebraic system of equations in which initial and final conditions on the solvent structure are specified. Such a problem can be solved provided that the physical property models allow the solvent structure to be set with a real-valued input for the number of groups of a given type. Thus, provided that one feasible point is known for the equalities in the primal problem, it can often be used to derive initial guesses for the solution of other primal problems. Problem (P4I) is implemented in gPROMS, as the other primal subproblems.

3. The infeasibility of Problems (P2a) and (P3a) is handled in different ways depending on the root cause: in gPROMS, a problem is reported to be infeasible if either no solution to the model equations (equality constraints) is found at the starting guess, that is the equality constraints cannot be initialized at the initial guess, or if no solution is found that satisfies both the equality and inequality constraints, that is the solver generates points where the equality constraints are satisfied but the inequality constraints are violated. If the first case occurs, in this algorithm, the test is treated as inconclusive in the CAMPD algorithm, as a failure to initialize model equations, whereas in the latter case the test is treated as infeasible.

APPENDIX E

The values assigned to the constants that appear in the formulations (P2a), (P3a), and (M) are shown in Table E1.

Table E1. Values Assigned to Constants in the Problems (P2a), (P3a), and (M)

Constant	Value
ϵ	10^{-3}
ϵ_a	-10^{-3}
ϵ_c	9
ϵ_n	10^{-3}
β	10^4
b	9
M^L	−7372
M^U	7372

Manuscript received Apr. 29, 2016, and revision received June 23, 2016.