ESAIM: M2AN 55 (2021) 2981–3016
https://doi.org/10.1051/m2an/2021075

ESAIM: Mathematical Modelling and Numerical Analysis
www.esaim-m2an.org

AN ANALYSIS OF THE UNIFIED FORMULATION FOR THE EQUILIBRIUM PROBLEM OF COMPOSITIONAL MULTIPHASE MIXTURES

Ibtihel Ben Gharbia[1], Mounir Haddou[2], Quang Huy Tran[1,*] and Duc Thach Son Vu[1]

Abstract. In this paper, we conduct a thorough mathematical analysis of the unified formulation advocated by Lauser *et al.* [*Adv. Water Res.* **34** (2011) 957–966] for compositional multiphase flows in porous media. The interest of this formulation lies in its potential to automatically handle the appearance and disappearance of phases. However, its practical implementation turned out to be not always robust for realistic fugacity laws associated with cubic equations of state, as shown by Ben Gharbia and Flauraud [*Oil Gas Sci. Technol.* **74** (2019) 43]. By focusing on the subproblem of phase equilibrium, we derive sufficient conditions for the existence of the corresponding system of equations. We trace back the difficulty of cubic laws to a deficiency of the Gibbs functions that comes into play due to the "unifying" feature of the new formulation. We propose a partial remedy for this problem by extending the domain of definition of these functions in a natural way. Besides, we highlight the crucial but seemingly unknown fact that the unified formulation encapsulates all the properties known to physicists on phase equilibrium, such as the tangent plane criterion and the minimization of the Gibbs energy of the mixture.

Mathematics Subject Classification. 76T99, 80M25, 90C33.

Received December 13, 2020. Accepted November 15, 2021.

1. Introduction

1.1. Motivation and objectives

In the numerical simulation of multicomponent (a.k.a. compositional) multiphase fluid flows, a delicate issue often arises in the handling of the appearance and disappearance of phases for various species, due to the laws of thermodynamic equilibrium. The traditional dynamic approach, known as *variable-switching* in reservoir simulations [10], considers only the unknowns and equations of the present phases. Albeit natural, it is awkward and even costly to implement, insofar as switching can occur all the time. Lauser *et al.* [17] proposed an alternative approach, called *unified formulation*, in which a fixed set of unknowns and equations is maintained during the calculations. This major theoretical advance is achieved by means of complementarity conditions, which allow distinct functioning regimes to be expressed in the same mathematical way, as is already the case

Keywords and phrases. Complementarity condition, unified formulation, phase equilibrium, Gibbs energy function, cubic EOS.

[1] IFP Energies nouvelles, 1 et 4 avenue de Bois Préau, 92852 Rueil-Malmaison Cedex, France.
[2] Univ Rennes, INSA, CNRS, IRMAR, UMR 6625, F-35000 Rennes, France.
*Corresponding author: quang-huy.tran@ifpen.fr

in a wide range of areas such as mechanics, electronics or geology [1,31]. Another key ingredient behind this "egalitarian" treatment of all regimes is the notion of extended partial fractions that must be assigned to species in all phases, including absent ones.

As a promise of more efficient simulations, the unified formulation has met with some success among numericists, as testified by the subsequent works by Ben Gharbia [5], Ben Gharbia and Jaffr [7], Masson *et al.* [20,21] and Beaude *et al.* [4]. These are all based on simple fugacity coefficients, such as given by Henry's law. Another series of works at IFPEN [6,8,19,30] is focused on realistic fugacity coefficients given by cubic equations of state, such as Peng–Robinson's law. Although the latter investigations have demonstrated a clear superiority of the unified formulation over the variable-switching one regarding computational time in some cases, the outcome remains unclear in other cases with single-phase transition: the nonlinear solver for the (unified) algebraic system of equations may not converge at all, unlike its competitor.

There are two possible explanations for this observed lack of robustness from the unified formulation. To sketch them out in a precise manner, we need the following formal setup. After discretization in time and space of the continuous flow model using the unified formulation, the system of equations to be solved at each time-step takes the abstract form

$$\Lambda(X) = 0, \tag{1.1a}$$

$$\min(G(X),\, H(X)) = 0, \tag{1.1b}$$

where $X \in \mathcal{D} \subset \mathbb{R}^\ell$ is the unknown vector and $\Lambda : \mathcal{D} \to \mathbb{R}^{\ell-m}$, $G : \mathcal{D} \to \mathbb{R}^m$ et $H : \mathcal{D} \to \mathbb{R}^m$ are continuously differentiable functions on the open domain $\mathcal{D}$. The componentwise action of the minimum function in (1.1b) is merely a convenient way of expressing the complementarity $0 \leqslant G(X) \perp H(X) \geqslant 0$. For conciseness, let us put

$$F(X) = \begin{bmatrix} \Lambda(X) \\ \min(G(X),\, H(X)) \end{bmatrix} \in \mathbb{R}^\ell, \tag{1.2}$$

so that (1.1) becomes $F(X) = 0$. We can then envision two scenarios that could cause the unified formulation to perform poorly:

(1) System (1.1) is ill-posed for some data and thermodynamic laws. In other words, it may not have a unique solution or may not have a solution at all. An even worse situation is when some components of Λ – and therefore of F – are not well-defined over the whole domain of interest $\mathcal{D}$, so that (1.1) no longer makes sense. As will be seen later, this occurs for cubic equations of state frequently used in realistic simulations.
(2) The numerical algorithm used to solve system (1.1) is not well suited to the semismooth nature of F. Indeed, the complementarity equations (1.1b) are not differentiable, which prevents the standard Newton method to be applied. A common remedy is the so-called *Newton-min* method [2,15]. However, Newton-min may suffer from periodic oscillations for large time-steps, as evidenced by Ben Gharbia and Flauraud [6].

The first issue originates from physical modeling. It is the subject of this article, whose primary objective is to clarify the conditions on thermodynamic laws under which system (1.1) is well-behaved and to propose some improvements of the model so as to guarantee the existence of a solution. The second issue pertains to numerical methods. It requires a new method to be designed in order to replace Newton-min and was addressed in a previous paper [35].

1.2. Main results and outline of the paper

Valuable insights into the difficulty can be gained if, instead of the fully discretized flow model, we focus on an elementary *phase equilibrium* problem that lies at the core of the thermodynamic part. This is why we start by stating the phase equilibrium problem for multicomponent mixtures in Section 2, comparing the variable-switching formulation to the unified formulation (Sect. 2.2) after recalling some preliminary notions in Section 2.1.

Section 3 is devoted to the analysis of the unified formulation. We first revisit two thermodynamic properties in light of the new framework, namely, the principle of Gibbs energy minimization (Sect. 3.1) and the tangent plane criterion (Sect. 3.2). Although these properties are well-known in thermodynamics by virtue of various physical arguments, the point we would like to make here is that they are all mathematical consequences of the unified formulation. In (3.3), we introduce an important phasewise subproblem called *local inversion* of extended fugacities in Section 3.3.1. Sufficient conditions are worked out to ensure the existence and uniqueness of a solution to this extended fugacities inversion subproblem. In essence, we require strict convexity of the Gibbs functions in each phase, as well as invertibility and surjectivity of their gradient maps. These assumptions are also shown in Section 3.4 to guarantee the existence of a solution to the full phase equilibrium problem.

Given a fugacity or activity law from physics textbooks, there is no reason for the corresponding molar Gibbs function to fulfill the hypotheses of strict convexity and invertibility/surjectivity of the gradient map. In Section 4, we further investigate the question of strict convexity for some simple fugacity and activity models, namely, Henry's laws (Sect. 4.1), Margules' law (Sect. 4.2), and Van Laar's law (Sect. 4.3). For each of these, we manage to determine the subregion in the space of parameters for which strict convexity holds.

A prominent category of fugacity laws widely used in realistic simulations of two-phase mixtures stems from *cubic equations of state* (EOS). As recalled in Section 5.1, the definition of the corresponding thermodynamic quantities involves solving a cubic equation which does not always have three real roots. After a careful study of the critical values (Sect. 5.2) and the frontier between the 1-root and 3-root regions (Sect. 5.3) for Peng–Robinson's law, one of the most advanced cubic EOS-based models, we explain the trouble with these laws regarding the domains of definition for different functions involved in (1.1). In a nutshell, since there are not always three real roots, the Gibbs functions and fugacity coefficients are not always well-defined simultaneously for both phases over the whole domain of generalized partial fractions. While this pathology is not detrimental to the variable-switching formulation, where only present phases are considered, it causes tremendous harm to the unified formulation, for which information relative to both phases must be permanently available. The uncovering of this difficulty in Section 6.1 prompts us to design an extension procedure for various thermodynamic functions, in an attempt to maintain a good behavior for the unified formulation, that is, to hope for the existence of a solution to (1.1). The basic idea, elaborated on in Section 6.2, is to extend the Gibbs functions by replacing the missing real root by the common real part of the two conjugate complex roots. This construction is supported by further calculations.

2. Phase equilibrium for multicomponent mixtures

2.1. Preliminary notions

In this paper, we are concerned with the advantages and drawbacks of the unified formulation for the phase equilibrium problem at fixed pressure and temperature. To state this problem, one needs some prerequisites on the thermodynamics of multiphase multicomponent mixtures.

2.1.1. Species, phases and fractions

A multicomponent mixture is a physical system consisting of several chemically distinct components or *species*, *e.g.*, hydrogen (H_2), water (H_2O), carbon dioxide (CO_2), methane (CH_4)... It can be thought of in a more abstract way by introducing the set of species

$$\mathcal{K} = \{\mathrm{I}, \mathrm{II}, \ldots, K\}, \qquad K \geqslant 2, \tag{2.1}$$

whose elements are labeled by Roman numerals. The total number of components $K = |\mathcal{K}|$ usually ranges from tens to hundreds. Each component $i \in \mathcal{K}$ may be present under one or many *phases*. Intuitively, a phase is more or less a state of matter, *e.g.*, gas (G), liquid (L), oil (O), solid (S)... This notion may be subtler, though, at high pressure [12]. Again, to lay down an abstract framework, we consider the set of all virtually possible phases

$$\mathscr{P} = \{1, 2, \ldots, P\}, \qquad P \geqslant 2, \tag{2.2}$$

whose elements are labeled by Arabic numerals. The choice of $\mathscr{P}$ within a model is the (difficult) task of physicists: P should be large enough to take into account the appearance of new phases in models with time evolution, but not too large for computations to remain feasible. In reservoir simulations, the maximum number of possible phases $P = |\mathscr{P}|$ is commonly about 3.

The relative importance of each phase $\alpha \in \mathscr{P}$ within the mixture is measured by the *phasic* fraction $Y_\alpha \in [0,1]$, such that

$$\sum_{\alpha\in\mathscr{P}} Y_\alpha = 1. \tag{2.3}$$

A phase for which $Y_\alpha = 0$ is said to be *absent*. Otherwise, it is *present*. The subset of present phases, namely,

$$\Gamma = \{\alpha \in \mathscr{P} \mid \ Y_\alpha > 0\} \subset \mathscr{P} \tag{2.4}$$

is referred to as the *context*. For a present phase $\alpha \in \Gamma$, it is possible to compare the relative contribution of each component $i \in \mathcal{K}$ within it by defining the partial fractions $x_\alpha^i \in [0,1]$, such that

$$\sum_{i\in\mathcal{K}} x_\alpha^i = 1. \tag{2.5}$$

The vector $\boldsymbol{x}_\alpha = \left(x_\alpha^{\mathrm{I}}, \ldots, x_\alpha^{K-1}\right) \in \overline{\Omega} \subset \mathbb{R}^{K-1}$ is called *partial composition* of phase α. It consists of only the first $K-1$ partial fractions, since the quantities x_α^{I}, x_α^{II}, …, x_α^K are not independent, in view of (2.5). Whenever a x_α^K turns up in any formula, it should be interpreted as $x_\alpha^K = 1 - x_\alpha^{\mathrm{I}} - \ldots - x_\alpha^{K-1}$. The domain of $\boldsymbol{x}_\alpha$ is the closure of

$$\Omega = \left\{\boldsymbol{x} = \left(x^{\mathrm{I}}, \ldots, x^{K-1}\right) \in \mathbb{R}^{K-1} \mid x^{\mathrm{I}} > 0, \ldots, x^{K-1} > 0, 1 - x^{\mathrm{I}} - \ldots - x^{K-1} > 0\right\}, \tag{2.6a}$$

namely,

$$\overline{\Omega} = \left\{\boldsymbol{x} = \left(x^{\mathrm{I}}, \ldots, x^{K-1}\right) \in \mathbb{R}^{K-1} \mid x^{\mathrm{I}} \geqslant 0, \ldots, x^{K-1} \geqslant 0, 1 - x^{\mathrm{I}} - \ldots - x^{K-1} \geqslant 0\right\}. \tag{2.6b}$$

Although this choice somehow breaks the symmetry, it is commonly resorted to in practice.

Finally, there is a third notion of fraction, called *global fractions* and denoted by $c^i \in [0,1]$, which quantifies the overall relative importance of each component $i \in \mathcal{K}$ inside the mixture. Of course, we have

$$\sum_{i\in\mathcal{K}} c^i = 1. \tag{2.7}$$

The vector $\boldsymbol{c} = \left(c^{\mathrm{I}}, \ldots, c^{K-1}\right) \in \overline{\Omega} \subset \mathbb{R}^{K-1}$ is called *global composition* of components. Again, because of the dependence (2.7), only the first $K-1$ values in $\boldsymbol{c}$. Whenever a c^K appears in the text, it should be understood as $c^K = 1 - c^{\mathrm{I}} - \ldots - c^{K-1}$. The material balance of component i implies that

$$c^i = \sum_{\alpha\in\Gamma} Y_\alpha x_\alpha^i. \tag{2.8}$$

Given the context Γ, the phasic fractions $\{Y_\alpha\}_{\alpha\in\Gamma}$ and the partial fractions $\{x_\alpha^i\}_{(i,\alpha)\in\mathcal{K}\times\mathscr{P}}$, it is straightforward to calculate the global composition $\{c^i\}_{i\in\mathcal{K}}$ by (2.8). The phase equilibrium problem takes exactly the opposite direction: given the global composition $\{c^i\}_{i\in\mathcal{K}}$ satisfying (2.7), is it possible to find the context Γ, the phasic fractions $\{Y_\alpha\}_{\alpha\in\Gamma}$ and the partial fractions $\{x_\alpha^i\}_{(i,\alpha)\in\mathcal{K}\times\mathscr{P}}$ satisfying (2.3), (2.5) and (2.8) beside positivity? Obviously, we do not have enough equations yet. The missing ones will be supplied at the end of Section 2.1.3.

Remark 2.1. We have deliberately not specified whether the three kinds of fractions Y_α, x_α^i and c^i are molar, volumic or specific fractions. In fact, this does not matter. The mathematical structure of the problem remains the same and the theoretical development is similar in all cases.

2.1.2. Gibbs energy and chemical potential

The behavior of each phase $\alpha \in \mathscr{P}$ is governed by a single fundamental function $g_\alpha : \overline{\Omega} \to \mathbb{R}$ known as the (intensive) *Gibbs free energy* of the phase. We require g_α to be as smooth as necessary in Ω and continuously extendable to $\partial\Omega$. However, ∇g_α may blow up on $\partial\Omega$. From g_α, we define K functions $\mu_\alpha^j : \Omega \to \mathbb{R}$, $j \in \mathcal{K}$, called *chemical potentials* by

$$\mu_\alpha^j(\boldsymbol{x}) = g_\alpha(\boldsymbol{x}) + \left\langle \nabla g_\alpha(\boldsymbol{x}),\, \boldsymbol{\delta}^j - \boldsymbol{x} \right\rangle \tag{2.9}$$

for $\boldsymbol{x} \in \Omega$, where the vector $\boldsymbol{\delta}^j = (\delta_{j,1}, \delta_{j,2}, \ldots, \delta_{j,K-1}) \in \mathbb{R}^{K-1}$ is made up of Kronecker's symbols. The following statement gives two helpful identities between g_α and μ_α^i. The first one (2.10a) relates the Gibbs energy to the potentials. The second one (2.10b) provides the gradient of the Gibbs energy from the potentials.

Lemma 2.2 (Connection between Gibbs energy and chemical potentials)**.** *For all $\boldsymbol{x} \in \Omega$:*

$$g_\alpha(\boldsymbol{x}) = \sum_{j=\mathrm{I}}^{K} x^j\, \mu_\alpha^j(\boldsymbol{x}); \tag{2.10a}$$

$$\frac{\partial g_\alpha}{\partial x^j}(\boldsymbol{x}) = \mu_\alpha^j(\boldsymbol{x}) - \mu_\alpha^K(\boldsymbol{x}), \qquad \forall j \in \mathcal{K}\backslash\{K\}. \tag{2.10b}$$

Proof. Multiplying (2.9) by x^j, summing over $j \in \mathcal{K}$ and noticing that $\sum_{j\in\mathcal{K}} x^j \boldsymbol{\delta}^j = \boldsymbol{x}$, we end up with (2.10a). To prove (2.10b), we subtract the last potential

$$\mu_\alpha^K(\boldsymbol{x}) = g_\alpha(\boldsymbol{x}) + \left\langle \nabla g_\alpha(\boldsymbol{x}),\, \boldsymbol{\delta}^K - \boldsymbol{x} \right\rangle$$

from each μ_α^j, $j \in \mathcal{K}\backslash\{K\}$, given by (2.9). This cancels out $g_\alpha(\boldsymbol{x})$ and the desired identity follows from $\boldsymbol{\delta}^K = (0, 0, \ldots, 0)$. □

Remark 2.3. In the above, we used the generic variable $\boldsymbol{x}$ to alleviate notations. Of course, g_α is to be evaluated at $\boldsymbol{x}_\alpha$, the composition of phase α. As a matter of fact, the Gibbs function also depends on the pressure P_α and the temperature T_α of the phase [33]. But since we work at fixed pressure and temperature, we purposely omit to write them down in order to concentrate on the dependency with respect to the fractions.

2.1.3. Fugacity, fugacity coefficient and quilibrium conditions

For a solid phase, μ_α^i is a constant. For fluid phases such as gas, liquid and oil, the chemical potentials take the form

$$\mu_\alpha^i(\boldsymbol{x}) = \ln\big(x^i \Phi_\alpha^i(\boldsymbol{x})\big), \tag{2.11a}$$

in which Φ_α^i is called the *fugacity coefficient* of component i in phase α. Note, however, that it depends on the whole composition vector. As for the quantity

$$f_\alpha^i(\boldsymbol{x}) = x^i \Phi_\alpha^i(\boldsymbol{x}), \tag{2.11b}$$

it is known as the *fugacity* of component i in phase α. Substituting the form (2.11a) into (2.10a), we obtain

$$g_\alpha(\boldsymbol{x}) = \sum_{i=\mathrm{I}}^{K} x^i \ln x^i + \sum_{i=\mathrm{I}}^{K} x^i \ln \Phi_\alpha^i(\boldsymbol{x}) \tag{2.12}$$

The first sum $\sum_{j=\mathrm{I}}^{K} x^j \ln x^j$ is the *ideal* part. The second sum, denoted by

$$\Psi_\alpha(\boldsymbol{x}) = \sum_{i=\mathrm{I}}^{K} x^i \ln \Phi_\alpha^i(\boldsymbol{x}), \tag{2.13}$$

is the *excess* part. In this perspective, a fluid phase α is assimilated to a "perturbation" of the ideal gas. As will be done in Section 6, we shall act only on the excess part to modify the Gibbs function.

Owing to the regularity assumptions made on g_α and μ_α^i, the functions $\Psi_\alpha : \overline{\Omega} \to \mathbb{R}$ and $\ln \Phi_\alpha^i : \Omega \to \mathbb{R}$ are also as smooth as necessary, with Ψ_α extendable by continuity to $\overline{\Omega}$ but not the $\ln \Phi_\alpha^i$'s. The very useful relations below between Ψ_α and $\ln \Phi_\alpha^i$ are similar to those between g_α and μ_α^i.

Lemma 2.4 (Connection between excess energy and fugacity coefficients)**.** *For all* $\boldsymbol{x} \in \Omega$*:*

$$\ln \Phi_\alpha^j(\boldsymbol{x}) = \Psi_\alpha(\boldsymbol{x}) + \big\langle \nabla \Psi_\alpha(\boldsymbol{x}),\, \boldsymbol{\delta}^j - \boldsymbol{x} \big\rangle, \qquad \forall j \in \mathcal{K}; \tag{2.14a}$$

$$\frac{\partial \Psi_\alpha}{\partial x_\alpha^j}(\boldsymbol{x}) = \ln \Phi_\alpha^j(\boldsymbol{x}) - \ln \Phi_\alpha^K(\boldsymbol{x}), \qquad \forall j \in \mathcal{K}\backslash\{K\}; \tag{2.14b}$$

Proof. The proof is straightforward. For each identity from Lemma 2.2, we just have to separate the ideal part from the excess part. The ideal part vanishes trivially. □

In a multicomponent mixture without any chemical reaction (also called *non-reactive*), the presence of two phases $(\alpha, \beta) \in \Gamma \times \Gamma$ implies that some equilibrium conditions must be achieved. According to thermodynamics, these conditions are the equalities across the two phases of pressure, temperature, and the chemical potentials corresponding to each component $i \in \mathcal{K}$. In other words, the missing conditions for the phase equilibrium problem at fixed pressure and temperature are

$$\mu_\alpha^i(\boldsymbol{x}_\alpha) = \mu_\beta^i(\boldsymbol{x}_\beta), \qquad \text{for all } \ (i,\, \alpha,\, \beta) \in \mathcal{K} \times \Gamma \times \Gamma, \tag{2.15a}$$

or equivalently,

$$x_\alpha^i \Phi_\alpha^i(\boldsymbol{x}_\alpha) = x_\beta^i \Phi_\beta^i(\boldsymbol{x}_\beta), \qquad \text{for all } \ (i,\, \alpha,\, \beta) \in \mathcal{K} \times \Gamma \times \Gamma. \tag{2.15b}$$

The fugacity coefficients Φ_α^i are given empirically or inferred from an equation of state.

Remark 2.5. Our definitions (2.11) are not exactly those of textbooks, where

$$\widehat{\mu}_\alpha^i(\boldsymbol{x}_\alpha, \mathsf{P}, \mathsf{T}) = \widehat{\mu}_\bullet^i(\mathsf{P}, \mathsf{T}) + \mathsf{R}\mathsf{T} \ln\big(x_\alpha^i \Phi_\alpha^i(\boldsymbol{x}_\alpha, \mathsf{P}, \mathsf{T})\big), \tag{2.16a}$$

$$\widehat{f}_\alpha^i(\boldsymbol{x}_\alpha, \mathsf{P}, \mathsf{T}) = x_\alpha^i \Phi_\alpha^i(\boldsymbol{x}_\alpha, \mathsf{P}, \mathsf{T})\mathsf{P}, \tag{2.16b}$$

with R the universal gas constant and $\mu_\bullet^i(\mathsf{P}, \mathsf{T})$ a reference ideal value. Since P and T are equal across the phases, it is readily checked that the equality of "classical" chemical potentials $\widehat{\mu}_\alpha^i(\boldsymbol{x}_\alpha, \mathsf{P}, \mathsf{T}) = \widehat{\mu}_\beta^i(\boldsymbol{x}_\beta, \mathsf{P}, \mathsf{T})$ is indeed equivalent to (2.15a). Opting for (2.11) instead of (2.16) amounts to working with the Gibbs energy function g_α instead of

$$\widehat{g}_\alpha(\boldsymbol{x}_\alpha, \mathsf{P}, \mathsf{T}) = \sum_{i \in \mathcal{K}} \widehat{\mu}_\bullet^i(\mathsf{P}, \mathsf{T}) x_\alpha^i + \mathsf{R}\mathsf{T} g_\alpha(\boldsymbol{x}_\alpha),$$

which differs from g_α by an additive affine function and a multiplicative constant.

A given family of positive real-valued functions $\{\Phi_\alpha^i\}_{(i,\alpha) \in \mathcal{K} \times \mathscr{P}}$ is said to be *admissible* if, for each $\alpha \in \mathscr{P}$, there exists a Gibbs energy function g_α of which they are the fugacity coefficients.

2.2. Two mathematical formulations

Equipped with the preliminary notions of Section 1, we are now in a position to rigorously state the phase equilibrium problem in two different ways: the "traditional" one and the "modern" one.

2.2.1. Variable-switching formulation

The first formulation has the advantage of being "natural," insofar as it uses the variables that have been introduced so far. It also bears the name of *natural variable* formulation.

GIVEN

$$\mathcal{K},\ \mathscr{P},\ \{\Phi^i_\alpha\}_{(i,\alpha)\in\mathcal{K}\times\mathscr{P}} \text{ admissible},\ \boldsymbol{c}\in\overline{\Omega},$$

FIND

$$\Gamma\subset\mathscr{P},\ \{Y_\alpha\}_{\alpha\in\Gamma}>0,\ \{x^i_\alpha\}_{(i,\alpha)\in\mathcal{K}\times\Gamma}\geqslant 0$$

so as to satisfy

$$\sum_{\beta\in\Gamma} Y_\beta x^i_\beta - c^i = 0, \qquad \forall i\in\mathcal{K}; \tag{2.17a}$$

$$x^i_\alpha\Phi^i_\alpha(\boldsymbol{x}_\alpha) - x^i_\omega\Phi^i_\omega(\boldsymbol{x}_\omega) = 0, \qquad \forall(i,\alpha)\in\mathcal{K}\times\Gamma\backslash\{\omega\}, \tag{2.17b}$$

$$\sum_{j\in\mathcal{K}} x^j_\alpha - 1 = 0, \qquad \forall\alpha\in\Gamma, \tag{2.17c}$$

where ω is a fixed phase of Γ.

Obviously, equation (2.17b) is none other than (2.15b), but expressed in such a way to avoid redundancy. The material balances (2.17a), (2.17c) respectively match (2.8), (2.5). Note that (2.3) is not explicitly prescribed because it can be deduced from the existing equations by summing (2.17a) over $i\in\mathcal{K}$, switching order, and invoking (2.17c). For a given context Γ, system (2.17) contains $(K+1)|\Gamma|$ equations and unknowns. It must of course be assumed that the physical properties of the species involved are such that the $K(|\Gamma|-1)$ fugacity equalities (2.17b) are independent.

The price to be paid for naturality is that the context Γ is itself an unknown. To circumvent this difficulty, we start by making an "educated guess" for Γ. At every fixed Γ, we attempt to solve the algebraic equations (2.17): this is what physicists call a (P,T)-*flash*. After exiting the flash, we check the positivity of Y_α and the non-negativity of x^i_α, for $\alpha\in\Gamma$. Should one of these fractions have the wrong sign, we must change Γ by adding or deleting phases and go for another flash! The number of unknowns and equations for a flash strongly depends on the assumption currently made about the context Γ. There is a vast literature on numerical methods [22–24, 36] for the flash problem (2.17) at fixed Γ. In addition to the classical and generic Newton-Raphson method [3, 33], many special purpose algorithms have been dedicated to the flash problem. These are iterative methods based on various kinds of substitution [13], the most famous of them being the Rachford-Rice substitution [32].

2.2.2. Unified formulation

To avoid the annoyance of dynamically handling the context, Lauser *et al.* [17] put forward another formulation for the phase equilibrium problem.

GIVEN

$$\mathcal{K},\ \mathscr{P},\ \{\Phi^i_\alpha\}_{(i,\alpha)\in\mathcal{K}\times\mathscr{P}} \text{ admissible},\ \boldsymbol{c}\in\overline{\Omega},$$

FIND

$$\{Y_\alpha\}_{\alpha\in\mathscr{P}}\geqslant 0,\ \{\xi^i_\alpha\}_{(i,\alpha)\in\mathcal{K}\times\mathscr{P}}\geqslant 0$$

so as to satisfy

$$\sum_{\beta\in\mathscr{P}} Y_\beta\xi^i_\beta - c^i = 0, \qquad \forall i\in\mathcal{K}; \tag{2.18a}$$

$$\xi^i_\alpha\Phi^i_\alpha(\boldsymbol{x}_\alpha) - \xi^i_\omega\Phi^i_\omega(\boldsymbol{x}_\omega) = 0, \qquad \forall(i,\alpha)\in\mathcal{K}\times\mathscr{P}\backslash\{\omega\}, \tag{2.18b}$$

$$\min\left(Y_\alpha,\, 1-\sum_{j\in\mathcal{K}}\xi_\alpha^j\right)=0,\qquad \forall\alpha\in\mathscr{P}, \tag{2.18c}$$

where ω is a fixed phase of $\mathscr{P}$ and $\boldsymbol{x}_\alpha=(x_\alpha^{\mathrm{I}},\ldots,x_\alpha^{K-1})\in\mathbb{R}^{K-1}$ is *defined* as

$$x_\alpha^i=\frac{\xi_\alpha^i}{\sum_{j\in\mathcal{K}}\xi_\alpha^j}\cdot \tag{2.18d}$$

In this second formulation, the partial fractions x_α^i have been replaced by a new notion, that of *extended* fractions ξ_α^i. The latter are defined over $(i,\alpha)\in\mathcal{K}\times\mathscr{P}$ instead of being restricted to $(i,\alpha)\in\mathcal{K}\times\Gamma$. Although the connection between extended fractions and partial fractions is given by the renormalization (2.18d), the x_α^i's here are merely auxiliary variables that can be eliminated by inserting (2.18d) into (2.18b). The thermodynamic equilibrium (2.18b) is now the equality of *extended fugacity* across phases for each component.

The complementarity conditions (2.18c) actually mean that, for each $\alpha\in\mathscr{P}$,

$$Y_\alpha\geqslant 0,\qquad 1-\sum_{j\in\mathcal{K}}\xi_\alpha^j\geqslant 0,\qquad Y_\alpha\left(1-\sum_{j\in\mathcal{K}}\xi_\alpha^j\right)=0. \tag{2.19}$$

As a consequence, for each phase $\alpha\in\mathscr{P}$, there are three possible regimes:

- $Y_\alpha>0$ (phase α is present). This implies $\sum_{j\in\mathcal{K}}\xi_\alpha^j=1$ and by (2.18d), $\xi_\alpha^i=x_\alpha^i$. Hence, the extended fractions of a present phase coincide with the usual partial fractions.
- $1-\sum_{j\in\mathcal{K}}\xi_\alpha^j>0$. This entails $Y_\alpha=0$ (phase α is absent) and $\xi_\alpha^i\neq x_\alpha^i$. The extended fractions of an absent phase differ from the usual partial fractions (see exception below).
- $Y_\alpha=0$ and $1-\sum_{j\in\mathcal{K}}\xi_\alpha^j=0$. This corresponds to a *transition* point, where phase α starts appearing or disappearing.

It is legitimate to wonder about the origin of the sign condition $1-\sum_{j\in\mathcal{K}}\xi_\beta^j\geqslant 0$. After all, it brings a new piece of information that was not included in the variable-switching formulation (2.17). As will be proven in Section 3.2, this condition ensures a stability property known as the *tangent plane criterion* by physicists. It can also be related to the minimization of the Gibbs energy of the mixture, as will be done in Section 3.1.

The ability of formulation (2.18) to deal with all possible configurations (arising from the presence or the absence of each phase) in a unified way is very attractive not only for convenience but also for computational efficiency. The context Γ no longer appears in the statement of the problem, but can be determined *a posteriori* by collecting those phases α for which $Y_\alpha>0$. As before, note that the phase balance (2.3) is not explicitly imposed because it can be recovered from the existing equations by summing (2.18a) over $i\in\mathcal{K}$, permuting order and taking advantage of (2.18c). System (2.18) has $(K+1)P$ equations and unknowns. It can be cast under the abstract form (1.1) with

$$\ell=(K+1)P,\quad m=P,\quad \left(\{Y_\alpha\}_{\alpha\in\mathscr{P}},\,\{\xi_\alpha^i\}_{(i,\alpha)\in\mathcal{K}\times\mathscr{P}}\right)=:X\in\mathcal{D}:=\mathbb{R}_+^P\times\left(\mathbb{R}_+^K\right)^P.$$

The existence of a solution to (2.18) can be guaranteed under some sufficient conditions on the Gibbs functions g_α, as elucidated in Section 3.4.

3. Properties of the unified formulation

The unified formulation enjoys many remarkable properties that seem to be unknown so far, at least to our knowledge. In particular, by postulating $1-\sum_{i\in\mathcal{K}}\xi_\alpha^i\geqslant 0$ from the beginning, it achieves a deep connection with some classical results in thermodynamics.

3.1. Connection with Gibbs energy minimization

We would like to better understand where this sign information comes from. In the literature, the condition $1 - \sum_{i\in\mathcal{K}} \xi_\alpha^i \geqslant 0$ is customarily derived from a phase stability analysis [22] (see also [6] for a more recent presentation). However, this classical analysis suffers from a few limitations. First, it is restricted to two phases. Second, it is local: the Gibbs energy difference under study must be linearized *via* a first-order Taylor expansion, before minimizing. Third, the notion of extended fractions appears only at the end, in a very *ad hoc* way.

We propose a more direct connection between the unified formulation (2.18) and some Gibbs energy minimization problem expressed in terms of the extended fractions ξ_α^i, without any linearization. In this problem, the quantities $1 - \sum_{i\in\mathcal{K}} \xi_\alpha^i$ will appear to be the Lagrange multipliers associated with the constraints $Y_\alpha \geqslant 0$. Conversely, while not every critical point of the minimization problem $(\mathcal{P})$ is a solution of the unified formulation, some "natural" choice of critical points satisfies the unified formulation.

3.1.1. Towards a novel interpretation

In order to state the minimization problem, we need to introduce a new Gibbs function. For each phase $\alpha \in \mathscr{P}$, let $\mathfrak{g}_\alpha : \mathbb{R}_+^K \to \mathbb{R}$ be the *extended* molar Gibbs energy defined as

$$\mathfrak{g}_\alpha\big(\xi^{\mathrm{I}}, \dots, \xi^K\big) = \sum_{i\in\mathcal{K}} \xi^i \ln\big(\xi^i \Phi_\alpha^i(\boldsymbol{x})\big), \tag{3.1}$$

using the renormalization (2.18d) to compute $\boldsymbol{x} \in \overline{\Omega}$ from $\boldsymbol{\xi} = \big(\xi^{\mathrm{I}}, \dots, \xi^K\big) \in \mathbb{R}_+^K \backslash \{0\}$. For normalized fractions, $\mathfrak{g}_\alpha(x^{\mathrm{I}}, \dots, x^K) = g_\alpha(\boldsymbol{x})$. Thus, g_α lifts the intensive Gibbs function g_α to the domain of extended fractions, but it does not coincide with the usual extensive Gibbs function [33]. The following Lemma summarizes its most useful properties.

Lemma 3.1. *For $\boldsymbol{\xi} \in \mathbb{R}_+^K \backslash \{0\}$ and $j \in \mathcal{K}$, we have*

$$\frac{\partial \mathfrak{g}_\alpha}{\partial \xi^j}(\boldsymbol{\xi}) = \ln\big(\xi^j \Phi_\alpha^j(\boldsymbol{x})\big) + 1, \tag{3.2a}$$

$$\mathfrak{g}_\alpha(\boldsymbol{\xi}) = \sum_{i\in\mathcal{K}} \xi^i \frac{\partial \mathfrak{g}_\alpha}{\partial \xi^i}(\boldsymbol{\xi}) - \sigma, \qquad \text{with } \sigma = \sum_{i\in\mathcal{K}} \xi^i. \tag{3.2b}$$

$$1 = \sum_{i\in\mathcal{K}} \xi^i \frac{\partial \ln\big(\xi^i \Phi_\alpha^i\big)}{\partial \xi^j}(\boldsymbol{\xi}). \tag{3.2c}$$

Proof. The readers are referred to Lemma 2.3 from [34]. The calculations involve the extensive Gibbs energy that we have not introduced here for conciseness, but are not difficult. □

We can now consider the following minimization problem $(\mathcal{P})$.

GIVEN

$$\mathcal{K},\ \mathscr{P},\ \{\Phi_\alpha^i\}_{(i,\alpha)\in\mathcal{K}\times\mathscr{P}} \text{ admissible, } \boldsymbol{c} \in \overline{\Omega},$$

FIND

$$\min_{\substack{\{Y_\alpha\}_{\alpha\in\mathscr{P}} \\ \{\boldsymbol{\xi}_\alpha\}_{\alpha\in\mathscr{P}}}} \sum_{\alpha\in\mathscr{P}} Y_\alpha \mathfrak{g}_\alpha(\boldsymbol{\xi}_\alpha) \tag{3.3a}$$

subject to

$$\sum_{\alpha\in\mathscr{P}} Y_\alpha - 1 = 0, \tag{3.3b}$$

$$\sum_{\alpha\in\mathscr{P}} Y_\alpha \xi_\alpha^i - c^i = 0, \qquad \forall i \in \mathcal{K}, \tag{3.3c}$$

$$-Y_\alpha \leqslant 0, \qquad \forall \alpha \in \mathscr{P}. \tag{3.3d}$$

The objective function in (3.3a) represents the extended Gibbs energy for the mixture. The equality constraints (3.3b), (3.3c) are exactly the material balances (2.3), (2.18a). This time, there is no redundancy since we have not imposed the complementarity conditions (2.18c).

Let u, $\{v^i\}_{i\in\mathcal{K}}$ and $\{w_\alpha\}_{\alpha\in\mathscr{P}}$ be the Lagrange multipliers associated respectively with the constraints (3.3b), (3.3c) and (3.3d). The Lagrangian of the minimization problem (3.3) reads

$$\mathscr{L}\big(\{Y_\alpha\},\{\boldsymbol{\xi}_\alpha\},u,\{v^i\},\{w_\alpha\}\big) = \sum_{\alpha\in\mathscr{P}} Y_\alpha \mathfrak{g}_\alpha(\boldsymbol{\xi}_\alpha) + u\left(\sum_{\alpha\in\mathscr{P}} Y_\alpha - 1\right) + \sum_{i\in\mathcal{K}} v^i\left(\sum_{\alpha\in\mathscr{P}} Y_\alpha \xi_\alpha^i - c^i\right) - \sum_{\alpha\in\mathscr{P}} w_\alpha Y_\alpha.$$

The saddle-points of $\mathscr{L}$ are given by the Karush-Kuhn-Tucker (KKT) conditions [26]

$$\mathfrak{g}_\beta(\boldsymbol{\xi}_\beta) + u + \sum_{i\in\mathcal{K}} v^i \xi_\beta^i - w_\beta = 0, \qquad \forall \beta \in \mathscr{P} \tag{3.4a}$$

$$Y_\beta\left[\frac{\partial \mathfrak{g}_\beta}{\partial \xi^j}(\boldsymbol{\xi}_\beta) + v^j\right] = 0, \qquad \forall (j,\beta) \in \mathcal{K}\times\mathscr{P}, \tag{3.4b}$$

$$\sum_{\alpha\in\mathscr{P}} Y_\alpha - 1 = 0, \tag{3.4c}$$

$$\sum_{\alpha\in\mathscr{P}} Y_\alpha \xi_\alpha^i - c^i = 0, \qquad \forall i \in \mathcal{K}, \tag{3.4d}$$

$$\min(Y_\beta,\, w_\beta) = 0, \qquad \forall \beta \in \mathscr{P}. \tag{3.4e}$$

The last equation (3.4e) expresses the complementarity between each inequality constraint (3.3d) and its Lagrange multiplier at optimality. It can be rephrased as

$$Y_\beta \geqslant 0, \qquad w_\beta \geqslant 0, \qquad Y_\beta\, w_\beta = 0.$$

A set of values $\{(Y_\alpha,\, \boldsymbol{\xi}_\alpha)\}_{\alpha\in\mathscr{P}}$ is said to be a *critical point* for problem (3.3) if there exists a set of values $(u,\, \{v^i\}_{i\in\mathcal{K}},\, \{w_\alpha\}_{\alpha\in\mathscr{P}})$ such that the KKT optimality system (3.4) is satisfied.

3.1.2. From one formulation to the other

We first show that it is easy to go from the unified formulation to the minimization problem.

Theorem 3.2. *Every solution* $\big\{\big(\bar{Y}_\alpha,\, \bar{\boldsymbol{\xi}}_\alpha\big)\big\}_{\alpha\in\mathscr{P}}$ *of the unified formulation* (2.18) *is a critical point of the minimization problem* (3.3)*, with*

$$\bar{u} = 1, \qquad \bar{v}^j = -\big[\ln\big(\bar{\varphi}^j\big) + 1\big], \qquad \bar{w}_\beta = 1 - \bar{\sigma}_\beta, \tag{3.5}$$

where $\bar{\varphi}^j$ *is the common value of the extended fugacity* $\bar{\xi}_\alpha^j \Phi_\alpha^j(\bar{\boldsymbol{x}}_\alpha)$ *across all phases* $\alpha \in \mathscr{P}$*.*

Proof. Let $\big\{\big(\bar{Y}_\alpha,\, \bar{\boldsymbol{\xi}}_\alpha\big)\big\}_{\alpha\in\mathscr{P}}$ be a solution of (2.18). The material balances (3.4c), (3.4d) are naturally met, as observed in Section 2.2.2. The equality of extended fugacities (2.18b) makes it possible to define $\bar{v}^j = -\big[\ln\big(\varphi^j\big) + 1\big]$ in the way described in the theorem. This choice of $\bar{v}^j$ trivially fulfills (3.4b) because of (3.2a). The choice of $\bar{w}_\beta$ implies (3.4e) because of (2.18c). It remains to check (3.4a). To this end, we use Lemma 3.1 to write

$$\mathfrak{g}_\beta\big(\bar{\boldsymbol{\xi}}_\beta\big) + \bar{u} + \sum_{i\in\mathcal{K}} \bar{v}^i \bar{\xi}_\beta^i - \bar{w}_\beta = \sum_{i\in\mathcal{K}} \bar{\xi}_\beta^i \frac{\partial \mathfrak{g}_\beta}{\partial \bar{\xi}_\beta^i}\big(\bar{\boldsymbol{\xi}}_\beta\big) - \bar{\sigma}_\beta + 1 - \sum_{i\in\mathcal{K}} \bar{\xi}_\beta^i \frac{\partial \mathfrak{g}_\beta}{\partial \bar{\xi}_\beta^i}\big(\bar{\boldsymbol{\xi}}_\beta\big) - (1 - \bar{\sigma}_\beta) = 0.$$

This completes the proof. □

The reverse direction is more delicate. The main difficulty lies in the indetermination of the extended fractions for an absent phase.

Theorem 3.3. *Let* $\left\{\left(\widetilde{Y}_\alpha, \widetilde{\xi}_\alpha\right)\right\}_{\alpha\in\mathscr{P}}$ *be a critical point of the minimization problem* (3.3).

(1) *If two phases* $(\alpha,\beta)\in\mathscr{P}\times\mathscr{P}$ *are both present, i.e.,* $\widetilde{Y}_\alpha>0$ *and* $\widetilde{Y}_\beta>0$, *then*

$$\widetilde{\sigma}_\alpha=\widetilde{\sigma}_\beta=1,\qquad \widetilde{\xi}^i_\alpha\Phi^i_\alpha(\widetilde{\boldsymbol{x}}_\alpha)=\widetilde{\xi}^i_\beta\Phi^i_\beta(\widetilde{\boldsymbol{x}}_\beta)\quad\text{for all}\ \ i\in\mathcal{K}. \tag{3.6}$$

This implies that the complementarity condition (2.18c) *holds for both phases and that the extended fugacity equalities* (2.18b) *hold between the two phases considered.*

(2) *If phase* α *is present and phase* β *is absent, i.e.,* $\widetilde{Y}_\alpha>0$ *and* $\widetilde{Y}_\beta=0$, *then*

$$\widetilde{\sigma}_\alpha=1,\qquad \sum_{i\in\mathcal{K}}\widetilde{\xi}^i_\beta\Big[\ln\Big(\widetilde{\xi}^i_\beta\Phi^i_\beta(\widetilde{\boldsymbol{x}}_\beta)\Big)-\ln\Big(\widetilde{\xi}^i_\alpha\Phi^i_\alpha(\widetilde{\boldsymbol{x}}_\alpha)\Big)\Big]+1-\widetilde{\sigma}_\beta\geqslant 0. \tag{3.7}$$

This implies that, in general, the complementarity condition (2.18c) *does not hold for phase* β *and the extended fugacity equalities* (2.18b) *do not hold between* α *and* β. *But* (2.18c) *is automatically met for phase* β *as soon as* (2.18b) *holds between* α *and* β.

Proof. Let $\left\{\left(\widetilde{Y}_\alpha,\widetilde{\xi}_\alpha\right)\right\}_{\alpha\in\mathscr{P}}$, $\left(\widetilde{u},\{\widetilde{v}^i\}_{i\in\mathcal{K}},\{\widetilde{w}_\alpha\}_{\alpha\in\mathscr{P}}\right)$ be a solution of the KKT system (3.4). First, assume that $\widetilde{Y}_\alpha>0$ and $\widetilde{Y}_\beta>0$. Dividing (3.4b) by $\widetilde{Y}$, we obtain $\partial_{\xi^j}\mathfrak{g}_\alpha\left(\widetilde{\boldsymbol{\xi}}_\alpha\right)+\widetilde{v}^j=0$ and $\partial_{\xi^j}\mathfrak{g}_\beta\left(\widetilde{\boldsymbol{\xi}}_\beta\right)+\widetilde{v}^j=0$. From this, we deduce that $\partial_{\xi^j}\mathfrak{g}_\alpha\left(\widetilde{\boldsymbol{\xi}}_\alpha\right)=\partial_{\xi^j}\mathfrak{g}_\beta\left(\widetilde{\boldsymbol{\xi}}_\beta\right)=-\widetilde{v}^j$. According to (3.2a) (Lem. 3.1), this is equivalent to the equality of extended fugacities (2.18b), rewritten in the second part of (3.6). On the other hand, $\widetilde{Y}_\alpha>0$ implies $\widetilde{w}_\alpha=0$ by (3.4e). Equation (3.4a) then becomes

$$\mathfrak{g}_\alpha\left(\widetilde{\boldsymbol{\xi}}_\alpha\right)+\widetilde{u}-\sum_{i\in\mathcal{K}}\widetilde{\xi}^i_\alpha\frac{\partial\mathfrak{g}_\alpha}{\partial\xi^i}\left(\widetilde{\boldsymbol{\xi}}_\alpha\right)=0.$$

Combining this with (3.2b) (Lem. 3.1), we infer that $\widetilde{\sigma}_\alpha=\widetilde{u}$. Repeating the same reasoning for β, we also get $\widetilde{\sigma}_\beta=\widetilde{u}$. Hence, $\widetilde{\sigma}_\alpha=\widetilde{\sigma}_\alpha$. This means that $\widetilde{\sigma}$ takes on the same value $\widetilde{u}$ in all present phases. Let $\widetilde{\Gamma}$ be set of $\pi\in\mathscr{P}$ such that $\widetilde{Y}_\pi>0$. Note that $\widetilde{\Gamma}\neq\varnothing$ because of (3.4c). Summing (3.4d) over $i\in\mathcal{K}$ and permuting the order of summation yields

$$0=\sum_{i\in\mathcal{K}}\sum_{\pi\in\mathscr{P}}\widetilde{Y}_\pi\widetilde{\xi}^i_\pi-\sum_{i\in\mathcal{K}}c^i=\sum_{\pi\in\mathscr{P}}\widetilde{Y}_\pi\widetilde{\sigma}_\pi-1=\widetilde{u}\sum_{\pi\in\widetilde{\Gamma}}\widetilde{Y}_\pi-1=\widetilde{u}-1.$$

Therefore, $\widetilde{u}=1$, which proves the first part of (3.6).

Assume now that $\widetilde{Y}_\alpha>0$ and $\widetilde{Y}_\beta=0$. It is no longer possible to divide (3.4b) by $\widetilde{Y}_\beta$ to retrieve information on the extended fugacities. Likewise, we now simply have $w_\beta\geqslant 0$ from (3.4e). Equation (3.4a) for phase β leads to

$$\mathfrak{g}_\beta\left(\widetilde{\boldsymbol{\xi}}_\beta\right)+\widetilde{u}+\sum_{i\in\mathcal{K}}\widetilde{v}^i\widetilde{\xi}^i_\beta=\widetilde{w}_\beta\geqslant 0.$$

Because phase α is present, $\widetilde{\sigma}_\alpha=\widetilde{u}=1$ and $\widetilde{v}^i=-\partial_{\xi^i}\mathfrak{g}_\alpha\left(\widetilde{\boldsymbol{\xi}}_\alpha\right)$. Invoking (3.2b) (Lem. 3.1) for phase β, we can transform the above equality into

$$\sum_{i\in\mathcal{K}}\widetilde{\xi}^i_\beta\left[\frac{\partial\mathfrak{g}_\beta}{\partial\xi^i}\left(\widetilde{\boldsymbol{\xi}}_\beta\right)-\frac{\partial\mathfrak{g}_\alpha}{\partial\xi^i}\left(\widetilde{\boldsymbol{\xi}}_\alpha\right)\right]-\widetilde{\sigma}_\beta+1\geqslant 0.$$

This is none other than the second part of (3.7). □

To fully grasp the meaning of Theorem 3.3, it is capital to observe that when a critical point of (3.3) has a vanishing phase $\beta \in \mathscr{P}$ for which $\widetilde{Y}_\beta = 0$, the corresponding extended fractions $\widetilde{\boldsymbol{\xi}}_\beta$ cannot be uniquely determined. Indeed, $\widetilde{\boldsymbol{\xi}}_\beta$ plainly does not contribute to neither the objective function (3.3a) nor the constraint (3.3c) at fixed $\widetilde{Y}_\beta = 0$. To put it another way, changing $\widetilde{\boldsymbol{\xi}}_\beta$ to any other vector $\mathbb{R}^K_+$ will provide another acceptable critical point. Thus, as soon as there is a critical point of (3.3) for which $\widetilde{Y}_\beta = 0$, there are in fact an infinity of such critical points. Among this infinity of critical points, only those for which

$$\widetilde{\xi}^i_\beta \Phi^i_\beta(\widetilde{\boldsymbol{x}}_\beta) = \widetilde{\xi}^i_\alpha \Phi^i_\alpha(\widetilde{\boldsymbol{x}}_\alpha) \quad \text{for all} \quad i \in \mathcal{K}, \tag{3.8}$$

where α is present phase $\left(\widetilde{Y}_\alpha > 0\right)$, will be also solutions of the unified formulation (2.18). Combining this with Theorem 3.2, we can interpret the unified formulation as a set of equations that is slightly "stronger" than that of the KKT system for the critical points. It is stronger in the sense that it helps selecting some special critical points – and hopefully just one – among the infinity of possible critical points that appear when one of the phases disappears.

3.1.3. A continuity principle

We even have heuristic arguments to claim the critical points selected by the unified formulation are "natural" ones. By this, we mean that the additional conditions (3.8) to be prescribed on the extended fractions of an absent phase β can be construed as the limit of a continuous process during which β was present before vanishing. To build up this process, let us reformulate the minimization problem $(\mathcal{P})$ or (3.3) as the *bilevel* or *hierarchical* problem

$$\min_{Y_\beta} \quad \min_{\substack{\{Y_\alpha\}_{\alpha\in\mathscr{P}\setminus\{\beta\}} \\ \{\boldsymbol{\xi}_\alpha\}_{\alpha\in\mathscr{P}}}} \quad \sum_{\alpha\in\mathscr{P}\setminus\{\beta\}} Y_\alpha \mathfrak{g}_\alpha(\boldsymbol{\xi}_\alpha) + Y_\beta \mathfrak{g}_\beta(\boldsymbol{\xi}_\beta) \tag{3.9a}$$

subject to

$$\sum_{\alpha\in\mathscr{P}\setminus\{\beta\}} Y_\alpha + Y_\beta - 1 = 0, \tag{3.9b}$$

$$\sum_{\alpha\in\mathscr{P}\setminus\{\beta\}} Y_\alpha \xi^i_\alpha + Y_\beta \xi^i_\beta - c^i = 0, \qquad \forall i \in \mathcal{K}, \tag{3.9c}$$

$$-Y_\alpha \leqslant 0, \qquad \forall \alpha \in \mathscr{P}\setminus\{\beta\}. \tag{3.9d}$$

The constraints (3.9b)–(3.9d) are imposed on the inner minimization problem $(\mathcal{P}_{Y_\beta})$

$$\min_{\substack{\{Y_\alpha\}_{\alpha\in\mathscr{P}\setminus\{\beta\}} \\ \{\boldsymbol{\xi}_\alpha\}_{\alpha\in\mathscr{P}}}} \quad \sum_{\alpha\in\mathscr{P}\setminus\{\beta\}} Y_\alpha \mathfrak{g}_\alpha(\boldsymbol{\xi}_\alpha) + Y_\beta \mathfrak{g}_\beta(\boldsymbol{\xi}_\beta) \tag{3.10}$$

for a fixed $Y_\beta \geqslant 0$. Assume that for each small enough $Y_\beta > 0$ there is a unique critical point, denoted by $\{\widetilde{Y}_\alpha(Y_\beta)\}_{\alpha\in\mathscr{P}\setminus\{\beta\}}$, $\{\widetilde{\boldsymbol{\xi}}_\alpha(Y_\beta)\}_{\alpha\in\mathscr{P}}$. From the KKT conditions for (3.10) subject to (3.9b)–(3.9d), it follows that (see [34], Sect. 2.3.2.4 for details)

$$\widetilde{\xi}^\beta_i(Y_\beta)\Phi(\widetilde{\boldsymbol{x}}_\beta(Y_\beta)) = \widetilde{\xi}^\alpha_i(Y_\beta)\Phi(\widetilde{\boldsymbol{x}}_\alpha(Y_\beta)) \qquad \text{for all} \quad i \in \mathcal{K}.$$

Now, we let $Y_\beta \downarrow 0$. If all of the quantities involved in the above equality have finite limits, we clearly end up with (3.8).

3.2. Tangent plane criterion

Another set of properties can be established by looking at the geometric significance of the extended fugacity equalities (2.18b). Recall that $\overline{\Omega}$ defined in (2.6b), is the domain of the partial fractions $\boldsymbol{x}$ renormalized from $\boldsymbol{\xi}$ by (2.18d). The generic element of $\overline{\Omega} \times \mathbb{R}$ is denoted by $(\boldsymbol{x}, y)$. Let

$$\mathcal{G}_\alpha = \left\{(\boldsymbol{x},\, y) \in \overline{\Omega} \times \mathbb{R} \mid y = g_\alpha(\boldsymbol{x})\right\} \tag{3.11}$$

be the graph of the Gibbs energy function $g_\alpha : \overline{\Omega} \to \mathbb{R}$. For an interior point $\boldsymbol{x}_\alpha \in \Omega$, we designate by $T_{\boldsymbol{x}_\alpha}\mathcal{G}_\alpha$ the tangent hyperplane to $\mathcal{G}_\alpha$ at $\boldsymbol{x}_\alpha$. This tangent hyperplane, which exists thanks to the regularity assumptions on g_α, is the graph of the affine function $T_{\boldsymbol{x}_\alpha} g_\alpha : \mathbb{R}^{K-1} \to \mathbb{R}$ defined as

$$T_{\boldsymbol{x}_\alpha} g_\alpha(\boldsymbol{x}) = g_\alpha(\boldsymbol{x}_\alpha) + \langle \nabla g_\alpha(\boldsymbol{x}_\alpha),\, \boldsymbol{x} - \boldsymbol{x}_\alpha \rangle. \tag{3.12}$$

Let us assume that a solution $\big(\{\bar{Y}_\alpha\}_{\alpha\in\mathscr{P}},\, (\bar{\boldsymbol{\xi}}_\alpha)_{\alpha\in\mathscr{P}}\big)$ exists to the unified formulation (2.18), in which the renormalized fractions $\bar{\boldsymbol{x}}_\alpha \in \Omega$ are computed from $\bar{\boldsymbol{\xi}}_\alpha$ by (2.18d). We are going to learn as much as we can about it.

Theorem 3.4. *For any pair $(\alpha, \beta) \in \mathscr{P} \times \mathscr{P}$ of phases, present or absent:*

(1) *The potentials in phase β are shifted from their counterparts in phase α by a same constant, i.e.,*

$$\mu_\beta^j(\bar{\boldsymbol{x}}_\beta) = \mu_\alpha^j(\bar{\boldsymbol{x}}_\alpha) + [\ln \bar{\sigma}_\alpha - \ln \bar{\sigma}_\beta] \tag{3.13a}$$

for all $j \in \mathcal{K}$, where

$$\bar{\sigma}_\alpha = \sum_{i\in\mathcal{K}} \bar{\xi}_\alpha^i. \tag{3.13b}$$

(2) *The two tangent hyperplanes $T_{\bar{\boldsymbol{x}}_\alpha}\mathcal{G}_\alpha$ and $T_{\bar{\boldsymbol{x}}_\beta}\mathcal{G}_\beta$ are parallel. Put another way,*

$$\nabla g_\alpha(\bar{\boldsymbol{x}}_\alpha) = \nabla g_\beta(\bar{\boldsymbol{x}}_\beta). \tag{3.13c}$$

Proof. For each phase $\alpha \in \mathscr{P}$, let us define $\bar{\sigma}_\alpha$ as in (3.13b), so that for $j \in \mathcal{K}$ we have $\bar{\xi}_\alpha^j = \bar{\sigma}_\alpha \bar{x}_\alpha^j$ in view of (2.18d). The extended fugacity equalities (2.18b) then become

$$\bar{\sigma}_\alpha \bar{x}_\alpha^j \Phi_\alpha^j(\bar{\boldsymbol{x}}_\alpha) = \bar{\sigma}_\beta \bar{x}_\beta^j \Phi_\beta^j(\bar{\boldsymbol{x}}_\beta). \tag{3.14}$$

Taking the logarithm of both sides and recalling (2.11a), we obtain

$$\ln \bar{\sigma}_\alpha + \mu_\alpha^j(\bar{\boldsymbol{x}}_\alpha) = \ln \bar{\sigma}_\beta + \mu_\beta^j(\bar{\boldsymbol{x}}_\beta). \tag{3.15}$$

From this, we deduce the translation property (3.13). Subtracting the last equality $\ln \bar{\sigma}_\alpha + \mu_\alpha^K(\bar{\boldsymbol{x}}_\alpha) = \ln \bar{\sigma}_\beta + \mu_\beta^K(\bar{\boldsymbol{x}}_\beta)$ from (3.15) and recalling (2.10b) (Lem. 2.2), we have

$$\frac{\partial g_\alpha}{\partial x^j}(\bar{\boldsymbol{x}}_\alpha) = \frac{\partial g_\beta}{\partial x^j}(\bar{\boldsymbol{x}}_\beta)$$

for all $j \in \{\mathrm{I}, \mathrm{II}, \dots, K-1\}$. This completes the proof for (3.13c). □

The first part of Theorem 3.4 reveals that, in general, there is no equality of chemical potentials if these were computed using the renormalized partial fractions. The second part of Theorem 3.4 is more interesting. Let us investigate this further by making an additional assumption on one of the phases. We recall the definition (3.12) of the linearized expansion $T_{\boldsymbol{x}_\alpha} g_\alpha(\boldsymbol{x})$.

Theorem 3.5 (Tangent plane criterion)**.** *Assume that a phase $\alpha \in \mathscr{P}$ is present, i.e., $\bar{Y}_\alpha > 0$. Then, for any other phase $\beta \in \mathscr{P}$, absent or present,*

$$T_{\bar{\boldsymbol{x}}_\beta} g_\beta(\boldsymbol{x}) \geqslant T_{\bar{\boldsymbol{x}}_\alpha} g_\alpha(\boldsymbol{x}), \qquad \text{for all } \boldsymbol{x} \in \mathbb{R}^{K-1}. \tag{3.16}$$

Thus, the tangent hyperplane $T_{\bar{\boldsymbol{x}}_\beta}\mathcal{G}_\beta$ lies above or coincides with the tangent hyperplane $T_{\bar{\boldsymbol{x}}_\alpha}\mathcal{G}_\alpha$.

Proof. From equality (3.13a), we have

$$\mu_\beta^K(\bar{\boldsymbol{x}}_\beta) = \mu_\alpha^K(\bar{\boldsymbol{x}}_\alpha) + C_{\alpha\beta}, \qquad C_{\alpha\beta} = \ln\bar{\sigma}_\alpha - \ln\bar{\sigma}_\beta.$$

Since $\bar{Y}_\alpha > 0$, the complementarity condition (2.18c) entails $\bar{\sigma}_\alpha = \sum_{j\in\mathcal{K}} \bar{\xi}_\alpha^j = 1$, hence $\ln\bar{\sigma}_\alpha = 0$. For any other $\beta \in \mathscr{P}$, we have $\bar{\sigma}_\beta = \sum_{j\in\mathcal{K}} \bar{\xi}_\beta^j \leqslant 1$, also by virtue of (2.18c). Therefore, $\ln\bar{\sigma}_\beta \leqslant 0$ and $C_{\alpha\beta} \geqslant 0$. Hence, $\mu_\beta^K(\bar{\boldsymbol{x}}_\beta) \geqslant \mu_\alpha^K(\bar{\boldsymbol{x}}_\alpha)$. Using (2.9) from Lemma 2.2, we can rewrite the previous inequality as

$$g_\beta(\bar{\boldsymbol{x}}_\beta) - \langle \nabla_{\boldsymbol{x}} g_\beta(\bar{\boldsymbol{x}}_\beta),\, \bar{\boldsymbol{x}}_\beta\rangle \geqslant g_\alpha(\bar{\boldsymbol{x}}_\beta) - \langle \nabla_{\boldsymbol{x}} g_\alpha(\bar{\boldsymbol{x}}_\alpha),\, \bar{\boldsymbol{x}}_\alpha\rangle. \tag{3.17}$$

On the other hand, taking the dot product of the equality of gradients (3.13c) with any $\boldsymbol{x} \in \Omega$, we have

$$\langle \nabla_{\boldsymbol{x}} g_\beta(\bar{\boldsymbol{x}}_\beta),\, \boldsymbol{x}\rangle = \langle \nabla_{\boldsymbol{x}} g_\alpha(\bar{\boldsymbol{x}}_\alpha),\, \boldsymbol{x}\rangle. \tag{3.18}$$

Adding together (3.17) and (3.18), we end up with

$$g_\beta(\bar{\boldsymbol{x}}_\beta) + \langle \nabla_{\boldsymbol{x}} g_\beta(\bar{\boldsymbol{x}}_\beta),\, \boldsymbol{x} - \bar{\boldsymbol{x}}_\beta\rangle \geqslant g_\alpha(\bar{\boldsymbol{x}}_\beta) + \langle \nabla_{\boldsymbol{x}} g_\alpha(\bar{\boldsymbol{x}}_\alpha),\, \boldsymbol{x} - \bar{\boldsymbol{x}}_\alpha\rangle$$

which is the desired result (3.16). □

This result, notoriously known as the *tangent plane criterion*, is usually derived by physicists from a local analysis of phase stability [22] (see also Sect. 3.1). Theorem 3.5 testifies to the fact that this stability property is already encoded in the unified formulation *via* the sign of $1 - \sum_{j\in\mathcal{K}} \bar{\xi}_\beta^j$. If phase β is "strictly" absent, namely, if $1 - \sum_{j\in\mathcal{K}} \bar{\xi}_\beta^j > 0$ and $\bar{Y}_\beta = 0$, then the tangent hyperplane $T_{\bar{\boldsymbol{x}}_\beta}\mathcal{G}_\beta$ will lie strictly above $T_{\bar{\boldsymbol{x}}_\alpha}\mathcal{G}_\alpha$.

We now push one step further by looking at the case of several present phases. Let $\bar{\Gamma}$ be the set of all $\alpha \in \mathscr{P}$ such that $\bar{Y}_\alpha > 0$.

Corollary 3.6 (Common tangent hyperplane)**.** *At a solution of the unified formulation satisfying $\bar{\boldsymbol{x}}_\alpha \in \Omega$ for all $\alpha \in \mathscr{P}$, the tangent hyperplanes $\{T_{\bar{\boldsymbol{x}}_\alpha}\mathcal{G}_\alpha\}_{\alpha\in\bar{\Gamma}}$, are all the same. Moreover,*

$$\boldsymbol{c} = \big(c^{\mathrm{I}}, \ldots, c^{K-1}\big) \in \operatorname{int}(\operatorname{conv}(\{\bar{\boldsymbol{x}}_\alpha\}_{\alpha\in\bar{\Gamma}})), \tag{3.19}$$

i.e., the global composition point belongs to the open convex hull spanned by the points $\{\bar{\boldsymbol{x}}_\alpha\}_{\alpha\in\bar{\Gamma}}$.

Proof. Let $(\alpha,\beta) \in \bar{\Gamma}\times\bar{\Gamma}$. Applying Theorem 3.5 twice and switching their roles, we have $T_{\bar{\boldsymbol{x}}_\beta} g_\beta(\boldsymbol{x}) \geqslant T_{\bar{\boldsymbol{x}}_\alpha} g_\alpha(\boldsymbol{x})$ and $T_{\bar{\boldsymbol{x}}_\alpha} g_\alpha(\boldsymbol{x}) \geqslant T_{\bar{\boldsymbol{x}}_\beta} g_\beta(\boldsymbol{x})$, whence $T_{\bar{\boldsymbol{x}}_\alpha} g_\alpha(\boldsymbol{x}) = T_{\bar{\boldsymbol{x}}_\beta} g_\beta(\boldsymbol{x})$ for all $\boldsymbol{x} \in \Omega$. Thus, $T_{\bar{\boldsymbol{x}}_\alpha}\mathcal{G}_\alpha = T_{\bar{\boldsymbol{x}}_\beta}\mathcal{G}_\beta$. The material balance (2.18a) reads

$$c^i = \sum_{\beta\in\mathscr{P}} \bar{Y}_\beta \bar{\xi}_\beta^i = \sum_{\alpha\in\bar{\Gamma}} \bar{Y}_\alpha \bar{x}_\alpha^i,$$

where the last equality comes from retaining only those summands in $\bar{\Gamma}$. Extracting the first $K-1$ components from the above equation yields

$$\boldsymbol{c} = \sum_{\alpha\in\bar{\Gamma}} \bar{Y}_\alpha \bar{\boldsymbol{x}}_\alpha. \tag{3.20}$$

Since $\bar{Y}_\alpha > 0$ and $\sum_{\alpha\in\Gamma} \bar{Y}_\alpha = 1$, the point $\boldsymbol{c}$ belongs to the interior of $\operatorname{conv}(\{\bar{\boldsymbol{x}}_\alpha\}_{\alpha\in\bar{\Gamma}})$. □

From this *common tangent plane* property, a purely geometric procedure can be devised in order to build a solution of the phase equilibrium problem (2.18). The construction involves the lower convex envelope of the function $\boldsymbol{x} \mapsto \min_{\alpha\in\mathscr{P}} g_\alpha(\boldsymbol{x})$. More details will be given in Section 3.4.

3.3. Existence and uniqueness for a phasewise subproblem

The key step towards ensuring the existence of a solution to (2.18) is to study a phasewise subproblem that arises in two different forms.

3.3.1. Extended fugacities local inversion problem

To solve (2.18) in practice, Lauser *et al.* [17] advocated using the common values $\{\varphi^i\}_{i\in\mathcal{K}}$ of extended fugacity across phases as main unknowns. This gives rise to a two-level algorithm. In the inner level, we solve P nonlinear systems of size $K \times K$

$$\xi_\alpha^i \Phi_\alpha^i(\boldsymbol{x}_\alpha) = \varphi^i, \qquad \forall i \in \mathcal{K}, \tag{3.21}$$

one for each $\alpha \in \mathscr{P}$. These *local fugacity inversion* problems express the extended fractions as implicit functions $\xi_\alpha^i(\boldsymbol{\varphi})$ of the extended fugacity vector $\boldsymbol{\varphi} = \left(\varphi^{\mathrm{I}}, \dots \varphi^K\right) \in \mathbb{R}_+^K$. In the outer level, we solve one nonlinear system consisting of the $K + P$ remaining equations

$$\textstyle\sum_{\beta\in\mathscr{P}} Y_\beta \xi_\beta^i(\boldsymbol{\varphi}) - c^i = 0, \qquad \forall i \in \mathcal{K}, \tag{3.22a}$$

$$\min\Big(Y_\alpha,\, 1 - \textstyle\sum_{j\in\mathcal{K}} \xi_\alpha^j(\boldsymbol{\varphi})\Big) = 0, \qquad \forall \alpha \in \mathscr{P}, \tag{3.22b}$$

in the $K + P$ unknowns $\left(\{Y_\alpha\}_{\alpha\in\mathscr{P}}, \{\varphi^i\}_{i\in\mathcal{K}}\right)$. This approach, the interest of which is to involve only "small" systems, was adopted by many authors [6, 8, 20, 21].

Taking the logarithm of both sides of (3.21), writing $\xi_\alpha^i = \sigma_\alpha x_\alpha^i$ and proceeding as in the proof of Theorem 3.4, we can transform the inner system (3.21) into

$$\nabla g_\alpha(\boldsymbol{x}_\alpha) = \left\{\ln \varphi^i - \ln \varphi^K\right\}_{1\leqslant i\leqslant K-1}, \tag{3.23a}$$

$$\ln \sigma_\alpha + \mu_\alpha^K(\boldsymbol{x}_\alpha) = \ln \varphi^K. \tag{3.23b}$$

Thus, our ability to solve (3.23) for all reasonable inputs $\boldsymbol{\varphi} \in \mathbb{R}_+^K$ relies on the existence of an unambiguous reciprocal function $[\nabla g_\alpha]^{-1}$.

3.3.2. Extended fractions for a single-phase solution

The second situation occurs when the solution of (2.18) is single-phase, say, in phase β. Put another way, $\bar{Y}_\beta = 1$ and $\bar{Y}_\alpha = 0$ for all $\alpha \in \mathscr{P}\backslash\{\beta\}$. By (2.18a), rewritten as (3.20), we have $\bar{\boldsymbol{x}}_\beta = \boldsymbol{c}$. Assume $\boldsymbol{c} \in \Omega$. After Theorem 3.4, the extended fractions in a vanishing phase $\alpha \in \mathscr{P}\backslash\{\beta\}$ satisfy

$$\nabla g_\alpha(\bar{\boldsymbol{x}}_\alpha) = \nabla g_\beta(\boldsymbol{c}), \tag{3.24a}$$

$$\ln \bar{\sigma}_\alpha + \mu_\alpha^K(\bar{\boldsymbol{x}}_\alpha) = \mu_\beta^K(\boldsymbol{c}), \tag{3.24b}$$

If the function ∇g_α were invertible, we could write $\bar{\boldsymbol{x}}_\alpha = [\nabla g_\alpha]^{-1}(\nabla g_\beta(\boldsymbol{c}))$. Then, it could deduced from (3.24b) that $\bar{\sigma}_\alpha = \exp[\mu_\beta^K(\boldsymbol{c}) - \mu_\alpha^K(\bar{\boldsymbol{x}}_\alpha)]$ and $\bar{\xi}_\alpha^i = \bar{\sigma}_\alpha \bar{x}_\alpha^i$. Hence, phase α would be entirely known. The ability to assign well-determined values to the extended fractions in an absent phase is an important feature of the unified formulation. System (3.24) has the same structure as (3.23).

3.3.3. Fundamental assumptions

The superiority of the unified formulation over the variable-switching formulation hinges upon the invertibility of (3.23) and (3.24), which cannot be taken for granted. To this end, additional assumptions need to be made. Below is the most natural set of assumptions.

Hypotheses 3.7. The gradient map $\nabla_{\boldsymbol{x}} g_\alpha : \Omega \to \mathbb{R}^{K-1}$ is surjective. Moreover, the Gibbs energy $g_\alpha : \Omega \to \mathbb{R}$ is *strictly* convex, that is, it satisfies one of the two conditions below, which are equivalent for a twice differentiable function:

(a) For all $(\boldsymbol{x}, \boldsymbol{y}) \in \Omega \times \Omega$ with $\boldsymbol{x} \neq \boldsymbol{y}$,

$$\left\langle \nabla g_\alpha(\boldsymbol{x}) - \nabla g_\alpha(\boldsymbol{y}),\, \boldsymbol{x} - \boldsymbol{y} \right\rangle > 0. \tag{3.25}$$

(b) For all $\boldsymbol{x} \in \Omega$, the Hessian matrix $\nabla^2 g_\alpha(\boldsymbol{x})$ is positive definite .

We refer the reader to [9] for the notion of strict convexity and for the equivalence between the two conditions (a) and (b) for twice differentiable functions.

Theorem 3.8 (Existence and uniqueness of the phasewise subproblem)**.** *Under Hypotheses 3.7, the extended fugacities local inversion problem* (3.23) *has a unique solution.*

Proof. Surjectivity provides existence of a solution $\boldsymbol{x} \in \Omega$ to $\nabla g_\alpha(\boldsymbol{x}) = \boldsymbol{u}$ for all $\boldsymbol{u} \in \mathbb{R}^{K-1}$. Strict convexity enforces uniqueness of such a solution. □

Hypotheses 3.7 is neither unrealistic nor unreachable, as shown by the following example.

Proposition 3.9. *The Gibbs energy function of an ideal gas*

$$g_\alpha(\boldsymbol{x}) = \sum_{i=\mathrm{I}}^{K} x^i \ln x^i, \tag{3.26}$$

where $x^K = 1 - x^\mathrm{I} - \ldots - x^{K-1}$*, satisfies Hypotheses 3.7.*

Proof. The gradient $\nabla g_\alpha : \Omega \to \mathbb{R}^{K-1}$ is given by

$$\nabla g_\alpha(\boldsymbol{x}) = \big(\ln x^\mathrm{I} - \ln x^K, \ldots, \ln x^{K-1} - \ln x^K\big). \tag{3.27}$$

This map is continuous over Ω. For any given $\boldsymbol{u} = \big(u^\mathrm{I}, \ldots, u^{K-1}\big) \in \mathbb{R}^{K-1}$, the nonlinear system $\nabla g_\alpha(\boldsymbol{x}) = \boldsymbol{u}$ can be easily inverted and the only solution is

$$x^j = \frac{\exp\big(u^j\big)}{1 + \sum_{i=\mathrm{I}}^{K-1} \exp(u^i)}, \qquad j \in \{\mathrm{I}, \mathrm{II}, \ldots, K-1\}.$$

This defines a unique continuous inverse map $[\nabla g_\alpha]^{-1} : \mathbb{R}^{K-1} \to \Omega$.

From the expression (3.27) of the gradient, the Hessian matrix can be found to be

$$\nabla^2 g(\boldsymbol{x}) = \frac{1}{x^K}\mathbf{E} + \mathrm{Diag}\left(\frac{1}{x^\mathrm{I}}, \ldots, \frac{1}{x^{K-1}}\right),$$

where $\mathbf{E}$ is the matrix whose all entries are equal to 1. It follows that, for a generic $\boldsymbol{v} \in \mathbb{R}^{K-1}$,

$$\big\langle \nabla^2 g(\boldsymbol{x})\boldsymbol{v},\, \boldsymbol{v}\big\rangle = \frac{1}{x^K}\big|v^\mathrm{I} + \ldots + v^{K-1}\big|^2 + \sum_{i=1}^{K-1} \frac{|v^i|^2}{x^i}.$$

When $\boldsymbol{x} \in \Omega$, it is obvious that $\big\langle \nabla^2 g(\boldsymbol{x})\boldsymbol{v},\, \boldsymbol{v}\big\rangle > 0$ for all $\boldsymbol{v} \neq \boldsymbol{0}$. □

3.4. Existence for the phase equilibrium problem in the unified formulation

Thanks to Hypotheses 3.7, a solution of (2.18) can also be worked out explicitly. Its construction is inspired by Gibbs' geometric one [12] for the two-phase binary ($K = 2$) case. This settles the issue of existence under some minor technicalities.

Hypotheses 3.7 are taken for granted throughout this section. Additionally, we recall that the functions $\{g_\alpha\}_{\alpha\in\mathscr{P}}$ are smooth (say, twice differentiable), take finite values on the boundary $\partial\Omega$ but their gradients blow up there, *i.e.*, $\lim_{\boldsymbol{x}\to\partial\Omega}\|\nabla g_\alpha(\boldsymbol{x})\| = +\infty$. The latter is due to the presence of logarithms in the ideal parts of the Gibbs functions. The function

$$g = \min_{\alpha\in\mathscr{P}} g_\alpha \tag{3.28}$$

is continuous on $\bar{\Omega}$ but may not be differentiable. Let $\breve{g}$ be the lower convex envelope of g on $\bar{\Omega}$. By design, $\breve{g}$ is a convex function. Like g, $\breve{g}$ is continuous. Here, we have a stronger property.

Lemma 3.10. *The lower convex envelope $\breve{g}$ is differentiable at all interior points $\boldsymbol{c} \in \Omega$.*

Proof. The lower convex envelope at $\boldsymbol{c}$ can be characterized as

$$\breve{g}(\boldsymbol{c}) = \sup_{h \leqslant g,\ h \text{ affine}} h(\boldsymbol{c}). \tag{3.29}$$

It is a convex function, which allows us to consider its subdifferential $\partial\breve{g}(\boldsymbol{c})$ at $\boldsymbol{c} \in \Omega$. It is known that $\partial\breve{g}(\boldsymbol{c})$ is a nonempty and convex set [9]. Let us distinguish two cases.

Case 1: $\breve{g}(\boldsymbol{c}) = g(\boldsymbol{c})$. Let $\boldsymbol{p} \in \partial\breve{g}(\boldsymbol{c})$. By definition of a subgradient, $\breve{g}(\boldsymbol{x}) \geqslant \breve{g}(\boldsymbol{c}) + \langle \boldsymbol{p}, \boldsymbol{x} - \boldsymbol{c}\rangle$ for all $\boldsymbol{x} \in \bar{\Omega}$. Let $\alpha \in \mathscr{P}$ such that $g(\boldsymbol{c}) = g_\alpha(\boldsymbol{c})$. Combining the previous inequality with $g_\alpha(\boldsymbol{x}) \geqslant g(\boldsymbol{x}) \geqslant \breve{g}(\boldsymbol{x})$ and $\breve{g}(\boldsymbol{c}) = g_\alpha(\boldsymbol{c})$, we obtain $g_\alpha(\boldsymbol{x}) \geqslant g_\alpha(\boldsymbol{c}) + \langle \boldsymbol{p}, \boldsymbol{x} - \boldsymbol{c}\rangle$ for all $\boldsymbol{x} \in \bar{\Omega}$. This means that $\boldsymbol{p} \in \partial g_\alpha(\boldsymbol{c}) = \{\nabla g_\alpha(\boldsymbol{c})\}$, which results in $\boldsymbol{p} = \nabla g_\alpha(\boldsymbol{c})$. Since the subdifferential $\partial\breve{g}(\boldsymbol{c})$ is reduced to a singleton, $\breve{g}$ is differentiable at $\boldsymbol{c}$.

Case 2: $\breve{g}(\boldsymbol{c}) < g(\boldsymbol{c})$. Let $\boldsymbol{p} \in \partial\breve{g}(\boldsymbol{c})$. Since $g(\boldsymbol{x}) \geqslant \breve{g}(\boldsymbol{x}) \geqslant \breve{g}(\boldsymbol{c}) + \langle \boldsymbol{p}, \boldsymbol{x} - \boldsymbol{c}\rangle$ for all $\boldsymbol{x} \in \bar{\Omega}$, the affine map $h(\boldsymbol{x}) = \breve{g}(\boldsymbol{c}) + \langle \boldsymbol{p}, \boldsymbol{x} - \boldsymbol{c}\rangle$ is a legitimate "competitor" in the supremum (3.29). If the graph of h does not intersect that of g, namely, if $h(\boldsymbol{x}) < g(\boldsymbol{x})$ for all $\boldsymbol{x} \in \bar{\Omega}$, we can find $\epsilon > 0$ such that $h(\boldsymbol{x}) + \epsilon < g(\boldsymbol{x})$ for all $\boldsymbol{x} \in \bar{\Omega}$, thanks to continuity of the functions and compactness of the domain. But then $h + \epsilon$ would be a better "candidate" in (3.29), as it would raise by ϵ the value of $\breve{g}(\boldsymbol{c})$. To avoid this contradiction, there exists $\bar{\boldsymbol{x}}_\alpha \in \bar{\Omega}$ such that $h(\bar{\boldsymbol{x}}_\alpha) = g_\alpha(\bar{\boldsymbol{x}}_\alpha) = g(\bar{\boldsymbol{x}}_\alpha)$.

Let us investigate $\breve{g}(\bar{\boldsymbol{x}}_\alpha)$. On the one hand, $h(\bar{\boldsymbol{x}}_\alpha) \leqslant \breve{g}(\bar{\boldsymbol{x}}_\alpha) \leqslant g(\bar{\boldsymbol{x}}_\alpha)$. On the other hand, $h(\bar{\boldsymbol{x}}_\alpha) = g(\bar{\boldsymbol{x}}_\alpha)$ as said above. Therefore, $\breve{g}(\bar{\boldsymbol{x}}_\alpha) = g(\bar{\boldsymbol{x}}_\alpha)$. By the same argument as in Case 1, we conclude that $\breve{g}$ is differentiable at $\bar{\boldsymbol{x}}_\alpha$ and $\nabla\breve{g}(\bar{\boldsymbol{x}}_\alpha) = \nabla g_\alpha(\bar{\boldsymbol{x}}_\alpha)$. From the inequality $g_\alpha(\boldsymbol{x}) \geqslant g(\boldsymbol{x}) \geqslant \breve{g}(\boldsymbol{c}) + \langle \boldsymbol{p}, \boldsymbol{x} - \boldsymbol{c}\rangle$ and the equality $g_\alpha(\bar{\boldsymbol{x}}_\alpha) = \breve{g}(\boldsymbol{c}) + \langle \boldsymbol{p}, \bar{\boldsymbol{x}}_\alpha - \boldsymbol{c}\rangle$, we infer that $\bar{\boldsymbol{x}}_\alpha$ achieves the minimum of the function $\boldsymbol{x} \mapsto g_\alpha(\boldsymbol{x}) - \breve{g}(\boldsymbol{c}) - \langle \boldsymbol{p}, \boldsymbol{x} - \boldsymbol{c}\rangle$ over $\bar{\Omega}$. Since the latter function is strictly convex with unbounded derivatives on the boundary, the minimum cannot take place on $\partial\Omega$. Thus, $\bar{\boldsymbol{x}}_\alpha \in \Omega$ and minimality then entails $\boldsymbol{p} = \nabla g_\alpha(\bar{\boldsymbol{x}}_\alpha)$. Hence, $\bar{\boldsymbol{x}}_\alpha$ is a tangent point.

At this stage, we have proved that to each $\boldsymbol{p} \in \partial\breve{g}(\boldsymbol{c})$ there corresponds a phase $\alpha \in \mathscr{P}$ and a point $\bar{\boldsymbol{x}}_\alpha \in \Omega$ such that $\boldsymbol{p} = \nabla g_\alpha(\bar{\boldsymbol{x}}_\alpha)$ and $\breve{g}(\boldsymbol{c}) = g_\alpha(\bar{\boldsymbol{x}}_\alpha) + \langle \boldsymbol{p}, \boldsymbol{c} - \bar{\boldsymbol{x}}_\alpha\rangle$ (the last condition simply expresses that $\boldsymbol{c}$ belongs to the tangent hyperplane $T_{\bar{\boldsymbol{x}}_\alpha} g_\alpha$. Assume that $\partial\breve{g}(\boldsymbol{c})$ contains two distinct elements $\boldsymbol{p} \neq \boldsymbol{q}$. By convexity, $(1-u)\boldsymbol{p} + u\boldsymbol{q} \in \partial\breve{g}(\boldsymbol{c})$ for all $u \in [0,1]$. To each $u \in [0,1]$ there correspond a phase $\alpha(u) \in \mathscr{P}$ and point $\bar{\boldsymbol{x}}_{\alpha(u)}(u) \in \Omega$ such that $(1-u)\boldsymbol{p} + u\boldsymbol{q} = \nabla g_{\alpha(u)}(\bar{\boldsymbol{x}}_{\alpha(u)}(u))$. If necessary and up to a reparametrization, we can always take another $\boldsymbol{q}$ in this segment that is sufficiently close to $\boldsymbol{p}$ so that $\alpha(u) \equiv \alpha$ for all u. Let

$$y_{\boldsymbol{c}}(u) = T_{\bar{\boldsymbol{x}}_{\alpha(u)}} g_\alpha(\boldsymbol{c}) = g_\alpha(\bar{\boldsymbol{x}}_\alpha(u)) + \langle (1-u)\boldsymbol{p} + u\boldsymbol{q},\, \boldsymbol{c} - \bar{\boldsymbol{x}}_\alpha(u)\rangle \tag{3.30}$$

be the value at $\boldsymbol{c}$ of the tangent map at $\bar{\boldsymbol{x}}_\alpha(u)$. Since $y_{\boldsymbol{c}}(u) = \breve{g}(\boldsymbol{c})$ for all $u \in [0,1]$, the derivative of $y_{\boldsymbol{c}}$ with respect to u must identically vanish. The calculation of this derivative leads to

$$\langle \boldsymbol{q} - \boldsymbol{p},\, \boldsymbol{c} - \bar{\boldsymbol{x}}_\alpha(u)\rangle = 0. \tag{3.31}$$

Taking the difference of (3.31) between $u = 0$ and $u = 1$ leads to

$$\langle \nabla g_\alpha(\bar{\boldsymbol{x}}_\alpha(0)) - \nabla g_\alpha(\bar{\boldsymbol{x}}_\alpha(1)),\, \bar{\boldsymbol{x}}_\alpha(0) - \bar{\boldsymbol{x}}_\alpha(1)\rangle = 0,$$

which violates the strict convexity condition (3.25). Therefore, $\partial\breve{g}(\boldsymbol{c})$ is a singleton. □

Thus, for $\boldsymbol{c} \in \Omega$, it makes sense to speak about the gradient $\nabla\breve{g}(\boldsymbol{c})$ and the tangent hyperplane, defined as the graph of the linearized expansion $T_{\boldsymbol{c}}\breve{g}(\boldsymbol{x}) = \breve{g}(\boldsymbol{c}) + \langle \nabla\breve{g}(\boldsymbol{c}),\, \boldsymbol{x} - \boldsymbol{c}\rangle$. We introduce

$$\bar{\Gamma}(\boldsymbol{c}) = \{\alpha \in \mathscr{P} \mid \exists\, \bar{\boldsymbol{x}}_\alpha \in \Omega,\ \ g_\alpha(\bar{\boldsymbol{x}}_\alpha) = T_{\boldsymbol{c}}\breve{g}(\bar{\boldsymbol{x}}_\alpha),\ \ \nabla g_\alpha(\bar{\boldsymbol{x}}_\alpha) = \nabla\breve{g}(\boldsymbol{c})\} \tag{3.32}$$

as the set of thoses phases whose Gibbs function g_α is tangent to the hyperplane $T_{\boldsymbol{c}}\breve{g}$.

Lemma 3.11. *For $\boldsymbol{c} \in \Omega$, the following properties hold true:*

(1) $\bar{\Gamma}(\boldsymbol{c}) \neq \varnothing$.
(2) *For each phase $\alpha \in \bar{\Gamma}(\boldsymbol{c})$, the contact point $\bar{\boldsymbol{x}}_\alpha$ is unique.*
(3) *If $P \leqslant K$, then $\boldsymbol{c} \in \operatorname{conv}\{\bar{\boldsymbol{x}}_\alpha\}_{\alpha\in\bar{\Gamma}(\boldsymbol{c})}$.*

Proof. Existence of at least a contact point. The argument is similar to the proof of Lemma 3.10, with $\boldsymbol{p} = \nabla\breve{g}(\boldsymbol{c})$.

Uniqueness of the contact point in each phase. If the hyperplane $T_{\boldsymbol{c}}\breve{g}$ were tangent to g_α at two distinct points $\bar{\boldsymbol{x}}_\alpha \neq \widetilde{\boldsymbol{x}}_\alpha$, then $\nabla\breve{g}(\boldsymbol{c}) = \nabla g_\alpha(\bar{\boldsymbol{x}}_\alpha) = \nabla g_\alpha(\widetilde{\boldsymbol{x}}_\alpha)$, and we would have $\langle \nabla g_\alpha(\bar{\boldsymbol{x}}_\alpha) - \nabla g_\alpha(\widetilde{\boldsymbol{x}}_\alpha),\, \bar{\boldsymbol{x}}_\alpha - \widetilde{\boldsymbol{x}}_\alpha\rangle = 0$, which violates the strict convexity condition (3.25).

Convex hull of the contact points. In the characterization (3.29) of $\breve{g}$, the supremum is also a maximum reached at $h = T_{\boldsymbol{c}}\breve{g}$ due to differentiability. The idea is to express this optimality by rotating the common tangent hyperplane of the contact points $\{\bar{\boldsymbol{x}}_\alpha\}_{\alpha\in\bar{\Gamma}(\boldsymbol{c})}$ using the gradient vector $\boldsymbol{p}$ as parameter. For $\boldsymbol{p} \in \mathbb{R}^{K-1}$, we define $\widetilde{\boldsymbol{x}}_\alpha(\boldsymbol{p}) = [\nabla g_\alpha]^{-1}(\boldsymbol{p})$ for $\alpha \in \bar{\Gamma}(\boldsymbol{c})$. Note that $\widetilde{\boldsymbol{x}}_\alpha(\nabla\breve{g}(\boldsymbol{c})) = \bar{\boldsymbol{x}}_\alpha$. The tangent map of g_α at $\widetilde{\boldsymbol{x}}_\alpha(\boldsymbol{p})$ reads

$$T_{\widetilde{\boldsymbol{x}}_\alpha(\boldsymbol{p})} g_\alpha(\boldsymbol{x}) = g_\alpha(\widetilde{\boldsymbol{x}}_\alpha(\boldsymbol{p})) + \langle \boldsymbol{p},\, \boldsymbol{x} - \widetilde{\boldsymbol{x}}_\alpha(\boldsymbol{p})\rangle = \langle \boldsymbol{p},\, \boldsymbol{x}\rangle - g_\alpha^*(\boldsymbol{p}), \tag{3.33}$$

where g_α^* stands for the Legendre conjugate of g_α. Let $\bar{\Gamma}(\boldsymbol{c}) = \{\alpha, \beta, \ldots, \pi, \omega\}$ and consider the maximization problem

$$\max_{\boldsymbol{p}} T_{\widetilde{\boldsymbol{x}}_\alpha(\boldsymbol{p})} g_\alpha(\boldsymbol{c}) \tag{3.34a}$$

subject to the $|\bar{\Gamma}(\boldsymbol{c})| - 1$ equality constraints

$$g_\alpha^*(\boldsymbol{p}) = g_\beta^*(\boldsymbol{p}), \quad g_\alpha^*(\boldsymbol{p}) = g_\gamma^*(\boldsymbol{p}), \quad \ldots, \quad g_\alpha^*(\boldsymbol{p}) = g_\pi^*(\boldsymbol{p}), \quad g_\alpha^*(\boldsymbol{p}) = g_\omega^*(\boldsymbol{p}). \tag{3.34b}$$

The constraints (3.34b) are aimed at making the $|\bar{\Gamma}(\boldsymbol{c})|$ functions (3.33) coincide with each other, so as to preserve common tangency. Since $|\bar{\Gamma}(\boldsymbol{c})| \leqslant P \leqslant K$, the number of constraints does not exceed the space dimension and problem (3.34) keeps a chance of being feasible. The objective function (3.34a) is the value taken by this common tangent hyperplane at $\boldsymbol{c}$.

If $\boldsymbol{p}$ stays in a small enough neighborhood of $\nabla\breve{g}(\boldsymbol{c})$, then $T_{\widetilde{\boldsymbol{x}}_\alpha(\boldsymbol{p})}$ remains below g (thanks to the compactness of $\bar{\Omega}$) and $h = T_{\widetilde{\boldsymbol{x}}_\alpha(\boldsymbol{p})}$ can be considered as a valid "candidate" in (3.29). Therefore, it is expected that $\boldsymbol{p} = \nabla\breve{g}(\boldsymbol{c})$ achieves a local optimum of (3.34). The first-order optimality conditions for (3.34) imply

$$\boldsymbol{c} - \nabla g_\alpha^*(\boldsymbol{p}) + \lambda_\beta\big(\nabla g_\alpha^*(\boldsymbol{p}) - \nabla g_\beta^*(\boldsymbol{p})\big) + \ldots + \lambda_\omega(\nabla g_\alpha^*(\boldsymbol{p}) - \nabla g_\omega^*(\boldsymbol{p})) = 0, \tag{3.35}$$

where $\lambda_\beta, \ldots, \lambda_\omega$ are the Lagrange multiplies associated with the constraints (3.34b). Plugging $\boldsymbol{p} = \nabla\breve{g}(\boldsymbol{c})$ into (3.35) and using $\nabla g_\alpha^*(\boldsymbol{p}) = \widetilde{\boldsymbol{x}}_\alpha(\boldsymbol{p})$, we end up with

$$\boldsymbol{c} = (1 - \lambda_\beta - \ldots \lambda_\omega)\bar{\boldsymbol{x}}_\alpha + \lambda_\beta\bar{\boldsymbol{x}}_\beta + \ldots + \lambda_\pi\bar{\boldsymbol{x}}_\pi + \lambda_\omega\bar{\boldsymbol{x}}_\omega. \tag{3.36}$$

Since the coefficients in the right-hand side sum to 1, at least one of them must be positive. Up to a permuation of $\bar{\Gamma}(\boldsymbol{c})$, we can assume that $1 - \lambda_\beta - \ldots - \lambda_\omega > 0$. Let us prove that the other coefficients are nonnegative. Suppose that $\lambda_\omega < 0$. The idea is now to rotate the common tangent hyperplane but to leave out the tangency constraint for ω, so that the new affine function becomes strictly lower that g_ω, remains tangent to the other Gibbs functions in $\bar{\Gamma}(\boldsymbol{c})$ and achieves a higher value at $\boldsymbol{c}$, which is contradiction.

Let $\boldsymbol{p} = \nabla\breve{g}(\boldsymbol{c}) + \delta\boldsymbol{p}$, where $\delta\boldsymbol{p}$ is orthogonal to the subspace spanned by $\bar{\boldsymbol{x}}_\alpha - \bar{\boldsymbol{x}}_\beta, \ldots, \bar{\boldsymbol{x}}_\alpha - \bar{\boldsymbol{x}}_\pi$. Since $\bar{\boldsymbol{x}}_\alpha - \bar{\boldsymbol{x}}_\beta = \Big(\nabla g_\alpha^* - \nabla g_\beta^*\Big)(\nabla\breve{g}(\boldsymbol{c}))$ and similarly for other phases, the $|\bar{\Gamma}(\boldsymbol{c})| - 2$ constraints $g_\alpha^*(\boldsymbol{p}) = g_\beta^*(\boldsymbol{p}), \ldots, g_\alpha^*(\boldsymbol{p}) = g_\pi^*(\boldsymbol{p})$ remain satisfied at first-order expansion. Let $y_{\boldsymbol{c}}(\boldsymbol{p}) = T_{\widetilde{\boldsymbol{x}}_\alpha(\boldsymbol{p})} g_\alpha(\boldsymbol{c})$ and $y_{\bar{\boldsymbol{x}}_\omega}(\boldsymbol{p}) = T_{\widetilde{\boldsymbol{x}}_\alpha(\boldsymbol{p})} g_\alpha(\bar{\boldsymbol{x}}_\omega)$ be the values of the new hyperplane evaluated at $\boldsymbol{c}$ and $\bar{\boldsymbol{x}}_\omega$. It is straigthforward to show that $\nabla y_{\boldsymbol{c}}(\boldsymbol{p}) = \boldsymbol{c} - \widetilde{\boldsymbol{x}}_\alpha(\boldsymbol{p})$ and $\nabla y_{\bar{\boldsymbol{x}}_\omega}(\boldsymbol{p}) = \bar{\boldsymbol{x}}_\omega - \widetilde{\boldsymbol{x}}_\alpha(\boldsymbol{p})$, so that $\nabla y_{\boldsymbol{c}}(\nabla\breve{g}(\boldsymbol{c})) = \boldsymbol{c} - \bar{\boldsymbol{x}}_\alpha$ and $\nabla y_{\bar{\boldsymbol{x}}_\omega}(\breve{g}(\boldsymbol{c})) = \bar{\boldsymbol{x}}_\omega - \bar{\boldsymbol{x}}_\alpha$. In view of (3.36),

$$c - \bar{\boldsymbol{x}}_\alpha = \lambda_\beta(\bar{\boldsymbol{x}}_\beta - \bar{\boldsymbol{x}}_\alpha) + \ldots + \lambda_\pi(\bar{\boldsymbol{x}}_\pi - \bar{\boldsymbol{x}}_\alpha) + \lambda_\omega(\bar{\boldsymbol{x}}_\omega - \bar{\boldsymbol{x}}_\alpha).$$

Taking the dot product with $\delta\boldsymbol{p}$ yields $\langle\nabla y_{\boldsymbol{c}}(\nabla\check{g}(\boldsymbol{c})),\,\delta\boldsymbol{p}\rangle = \lambda_\omega\,\langle\nabla y_{\bar{\boldsymbol{x}}_\omega}(\check{g}(\boldsymbol{c})),\,\delta\boldsymbol{p}\rangle$. Hence, it is possible to choose $\delta\boldsymbol{p}$ so as to increase $y_{\boldsymbol{c}}$ and to decrease $y_{\bar{\boldsymbol{x}}_\omega}$, in other words such that $T_{\tilde{\boldsymbol{x}}_\alpha(\boldsymbol{p})}g_\alpha(\boldsymbol{c}) > \check{g}(\boldsymbol{c})$ and $T_{\tilde{\boldsymbol{x}}_\alpha(\boldsymbol{p})}g_\alpha(\bar{\boldsymbol{x}}_\omega) < g_\omega(\bar{\boldsymbol{x}}_\omega)$ at first-order expansion. The affine function $h = T_{\tilde{\boldsymbol{x}}_\alpha(\boldsymbol{p})}g_\alpha$ would then be a strictly better candidate than $T_{\boldsymbol{c}}\check{g}$ in (3.29). This is impossible. □

The last property means that $\boldsymbol{c}$ is a convex combination of the contact points, that is, there exist $\{\bar{Y}_\alpha\}_{\alpha\in\bar{\Gamma}(\boldsymbol{c})} \geqslant 0$ such that $\sum_{\alpha\in\bar{\Gamma}(\boldsymbol{c})}\bar{Y}_\alpha = 1$ and $\sum_{\alpha\in\bar{\Gamma}(\boldsymbol{c})}\bar{Y}_\alpha\bar{\boldsymbol{x}}_\alpha = \boldsymbol{c}$. We are now ready to describe a solution.

Theorem 3.12. *Assume $P \leqslant K$, $\boldsymbol{c}\in\Omega$. Let $\{\bar{\boldsymbol{x}}_\alpha\}_{\alpha\in\bar{\Gamma}(\boldsymbol{c})}$, $\{\bar{Y}_\alpha\}_{\alpha\in\bar{\Gamma}(\boldsymbol{c})}$ be defined as above and set*

$$\bar{\xi}^i_\alpha = \bar{x}^i_\alpha, \qquad \text{for } (\alpha,i)\in\bar{\Gamma}(\boldsymbol{c})\times\mathcal{K}. \tag{3.37a}$$

For $(\beta,i)\in\mathscr{P}\backslash\bar{\Gamma}(\boldsymbol{c})\times\mathcal{K}$, set

$$\bar{Y}_\beta = 0, \qquad \bar{\boldsymbol{x}}_\beta = [\nabla g_\beta]^{-1}(\nabla\check{g}(\boldsymbol{c})), \qquad \bar{\xi}^i_\beta = \exp[T_{\boldsymbol{c}}\check{g}(\bar{\boldsymbol{x}}_\beta) - g_\beta(\bar{\boldsymbol{x}}_\beta)]\bar{x}^i_\beta. \tag{3.37b}$$

This procedure supplies us with a solution of (2.18).

Proof. The material balance (2.18a) is an easy consequence of $\sum_{\alpha\in\bar{\Gamma}(\boldsymbol{c})}\bar{Y}_\alpha\bar{\boldsymbol{x}}_\alpha = \boldsymbol{c}$, $\sum_{\alpha\in\bar{\Gamma}(\boldsymbol{c})}\bar{Y}_\alpha = 1$ and $\bar{Y}_\beta = 0$ for $\beta\in\mathscr{P}\backslash\bar{\Gamma}(\boldsymbol{c})$.

The equality of extended fugacities (2.18b) across phases is equivalent to that of $\nabla g_\alpha(\bar{\boldsymbol{x}}_\alpha)$ and of $\ln\bar{\sigma}_\alpha + \mu^K(\bar{\boldsymbol{x}}_\alpha)$, as was pointed out in Sections 3.2 and 3.3. On the one hand, it follows from (3.32) and (3.37b) that $\nabla g_\alpha(\bar{\boldsymbol{x}}_\alpha) = \nabla\check{g}(\boldsymbol{c})$ for all $\alpha\in\mathscr{P}$. On the other hand, if $\alpha\in\bar{\Gamma}(\boldsymbol{c})$, the common tangency $\check{g}(\boldsymbol{c}) + \langle\nabla\check{g}(\boldsymbol{c}),\,\bar{\boldsymbol{x}}_\alpha - \boldsymbol{c}\rangle = g_\alpha(\bar{\boldsymbol{x}}_\alpha)$ implies that

$$\mu^K(\bar{\boldsymbol{x}}_\alpha) = g_\alpha(\bar{\boldsymbol{x}}_\alpha) - \langle\nabla g_\alpha(\bar{\boldsymbol{x}}_\alpha),\,\bar{\boldsymbol{x}}_\alpha\rangle = \check{g}(\boldsymbol{c}) - \langle\nabla\check{g}(\boldsymbol{c}),\,\boldsymbol{c}\rangle.$$

Since $\bar{\sigma}_\alpha = \sum_{i\in\mathcal{K}}\bar{\xi}^i_\alpha = 1$ after (3.37a), we have $\ln\bar{\sigma}_\alpha + \mu^K(\bar{\boldsymbol{x}}_\alpha) = \check{g}(\boldsymbol{c}) - \langle\nabla\check{g}(\boldsymbol{c}),\,\boldsymbol{c}\rangle$. If $\beta\in\mathscr{P}\backslash\bar{\Gamma}(\boldsymbol{c})$, by virtue of (3.37b),

$$\bar{\sigma}_\beta = \sum_{i\in\mathcal{K}}\bar{\xi}^i_\beta = \exp[T_{\boldsymbol{c}}\check{g}(\bar{\boldsymbol{x}}_\beta) - g_\beta(\bar{\boldsymbol{x}}_\beta)]. \tag{3.38}$$

Therefore, $\ln\bar{\sigma}_\beta + \mu^K(\bar{\boldsymbol{x}}_\beta) = T_{\boldsymbol{c}}\check{g}(\bar{\boldsymbol{x}}_\beta) - g_\beta(\bar{\boldsymbol{x}}_\beta) = \check{g}(\boldsymbol{c}) - \langle\nabla\check{g}(\boldsymbol{c}),\,\boldsymbol{c}\rangle$. As a result, $\ln\bar{\sigma}_\alpha + \mu^K(\bar{\boldsymbol{x}}_\alpha)$ takes the same value for all $\alpha\in\mathscr{P}$.

To prove the complementarity conditions (2.18c), we first notice from various definitions that $\bar{Y}_\alpha \geqslant 0$ and $\bar{Y}_\alpha\big(1 - \sum_{i\in\mathcal{K}}\bar{\xi}^i_\alpha\big) = 0$ for all $\alpha\in\mathscr{P}$. Moreover, $1 - \sum_{i\in\mathcal{K}}\bar{\xi}^i_\alpha = 0$ for $\alpha\in\bar{\Gamma}(\boldsymbol{c})$. Hence, it just remains to prove that $1 - \sum_{i\in\mathcal{K}}\bar{\xi}^i_\beta \geqslant 0$ for $\beta\in\mathscr{P}\backslash\bar{\Gamma}(\boldsymbol{c})$. Starting from (3.38) and invoking the convexity of $\check{g}$, we have

$$\bar{\sigma}_\beta \leqslant \exp[\check{g}(\bar{\boldsymbol{x}}_\beta) - g_\beta(\bar{\boldsymbol{x}}_\beta)] \leqslant \exp[g(\bar{\boldsymbol{x}}_\beta) - g_\beta(\bar{\boldsymbol{x}}_\beta)] \leqslant \exp(0) = 1,$$

which is the desired result. □

The assumption $P \leqslant K$ turns out to be true in practice: there are about two or three phases at most for tens to hundreds of components. For the two-phase binary case, namely, when $K = P = 2$, the previous solution can be proved to be unique ([34], Thm. 2.5).

4. Convexity analysis of simple laws

The goal of this section is to review some commonly used laws that satisfy Hypotheses 3.7 unconditionally or conditionally. Each law will be given by the excess Gibbs function Ψ_α, which is connnected to the Gibbs function g_α by

$$g_\alpha(\boldsymbol{x}) = \sum_{i=\mathrm{I}}^{K} x^i\ln x^i + \Psi_\alpha(\boldsymbol{x}). \tag{4.1}$$

The subscript α stands for the phase to which the physical law under consideration applies.

4.1. Henry's law

In Section 3.3 (Prop. 3.9), we saw that an ideal gas $\Psi_\alpha \equiv 0$ fulfills Hypotheses 3.7. Next in the level of complexity is Henry's law [14]

$$\Psi_\alpha(\boldsymbol{x}) = \sum_{i=1}^{K} x^i \ln k^i \tag{4.2}$$

where $\{k^i\}_{i\in\mathcal{K}}$ are positive constants, each of them embodying a property of the corresponding species. The fugacity coefficients calculated by (2.14a) are then

$$\ln \Phi_\alpha^j(\boldsymbol{x}) = \ln k^i, \qquad \text{for all } \ j \in \mathcal{K}. \tag{4.3}$$

This is why this law is also called the *constant coefficients* law.

Proposition 4.1. *For all* $\big(k^{\mathrm{I}},\dots,k^K\big) \in \big(\mathbb{R}_+^*\big)^K$, *the Gibbs energy function* g_α *associated with Henry's law fulfills Hypotheses 3.7.*

Proof. Since Ψ_α is affine with respect to $\boldsymbol{x} = \big(x^{\mathrm{I}},\dots,x^{K-1}\big)$, its second derivatives all vanish. Therefore, the Hessian matrix $\nabla^2 g_\alpha$ coincides with that of the Gibbs function of the ideal gas. But this matrix was shown to be positive definite in Proposition 3.9. We still have to check that the range of the gradient map

$$\nabla g_\alpha(\boldsymbol{x}) = \big(\ln\big(k^{\mathrm{I}}x^{\mathrm{I}}\big) - \ln\big(k^K x^K\big),\dots,\ln\big(k^{K-1}x^{K-1}\big) - \ln\big(k^K x^K\big)\big).$$

is equal to $\mathbb{R}^{K-1}$. For a given $\boldsymbol{u} = \big(u^{\mathrm{I}},\dots,u^{K-1}\big) \in \mathbb{R}^{K-1}$, the nonlinear system $\nabla g_\alpha(\boldsymbol{x}) = \boldsymbol{u}$ can be easily inverted and the only solution is

$$x^j = \frac{\exp\big(u^j\big)/k^{\mathrm{I}}}{1/k^K + \sum_{i=\mathrm{I}}^{K-1}\exp(u^i)/k^i}, \qquad j \in \{\mathrm{I},\mathrm{II},\dots,K-1\}.$$

This defines a unique continuous inverse map $[\nabla g_\alpha]^{-1} : \mathbb{R}^{K-1} \to \Omega$. □

4.2. Margules' law

We now consider two laws dedicated to liquid binary mixtures ($K = 2$), namely, Margules' and Van Laar's [29]. For liquids, physicists rather talk about *activity* coefficients instead of fugacity coefficients. This distinction is however anecdotical, since the mathematical structure of thermodynamic equilibria remains the same [27].

Since $K - 1 = 1$, we simply write x instead of x^{I} and $\boldsymbol{x}$. The excess function associated with Margules' law is

$$\Psi_\alpha(x) = x(1-x)[A_{12}(1-x) + A_{21}x], \tag{4.4}$$

where $(A_{12},\, A_{21}) \in (\mathbb{R}^*)^2$ are two nonzero parameters. By (2.14a), the fugacity coefficients are

$$\ln \Phi_\alpha^{\mathrm{I}}(x) = [A_{12} + 2(A_{21} - A_{12})x](1-x)^2, \tag{4.5a}$$

$$\ln \Phi_\alpha^{\mathrm{II}}(x) = [A_{21} + 2(A_{12} - A_{21})(1-x)]x^2. \tag{4.5b}$$

To meet Hypotheses 3.7, the pair (A_{12}, A_{21}) must be restricted to some region of $\mathbb{R}^2$.

Proposition 4.2. *Let* $S = A_{12} + A_{21}$ *and* $D = A_{12} - A_{21}$. *Then, the Gibbs energy function* g_α *associated with Margules' law fulfills Hypotheses 3.7 if and only if*

$$S < 4 \text{ and } \ |D| < \frac{1}{3}\Big[S^2 - 18S + 54 + 2(9-2S)^{3/2}\Big]^{1/2}. \tag{4.6}$$

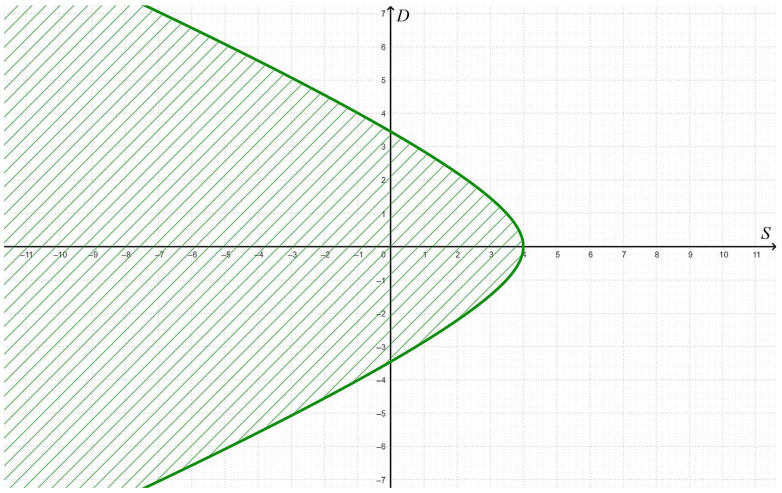


FIGURE 1. Region of strict convexity for the parameters of Margules' law in the (S, D)-plane.

The region indicated by (4.6) is colored in striped green in Figure 1. Its right-most point is located at $(S, D) = (4, 0)$, where it has a vertical tangent.

Proof. The first derivative of g_α is

$$g'_\alpha(x) = \ln x - \ln(1-x) + A_{12} + (2A_{21} - 4A_{12})x + 3(A_{12} - A_{21})x^2.$$

This is a continuous function over $(0, 1)$, with $\lim_{x\downarrow 0} g'_\alpha(x) = -\infty$ and $\lim_{x\uparrow 1} g'_\alpha(x) = +\infty$. Thus, g'_α has range in $\mathbb{R}$.

The second derivative of g_α, multiplied by $x(1-x)$ to remove singularities, is equal to

$$h(x) := x(1-x)g''_\alpha(x) = 1 + \big(x - x^2\big)[2A_{21} - 4A_{12} + 6(A_{12} - A_{21})x].$$

Let us change the variable to $y = x - \frac{1}{2} \in \big(-\frac{1}{2}, \frac{1}{2}\big)$ to work with the more symmetric function

$$H(y) := h\left(x - \frac{1}{2}\right) = 1 + \left(\frac{1}{4} - y^2\right)[6(A_{12} - A_{21})y - (A_{12} + A_{21})].$$

It remains to study H in order to determine the region of strict positivity $H(y) > 0$. The technical details can be found in [16] or Proposition 3.2 of [34]. □

4.3. Van Laar's law

Van Laar's law is also a model for activity coefficients of a liquid [29]. The excess Gibbs function associated with it is

$$\Psi_\alpha(x) = \frac{A_{12}A_{21}x(1-x)}{A_{12}x + A_{21}(1-x)}, \tag{4.7}$$

where $(A_{12},\, A_{21}) \in (\mathbb{R}^*)^2$ are two nonzero parameters. By (2.14a), the fugacity coefficients are

$$\ln \Phi^{\mathrm{I}}_\alpha(x) = A_{12}\left[\frac{A_{21}(1-x)}{A_{12}x + A_{21}(1-x)}\right]^2, \tag{4.8a}$$

$$\ln \Phi^{\mathrm{II}}_\alpha(x) = A_{21}\left[\frac{A_{12}x}{A_{12}x + A_{21}(1-x)}\right]^2. \tag{4.8b}$$

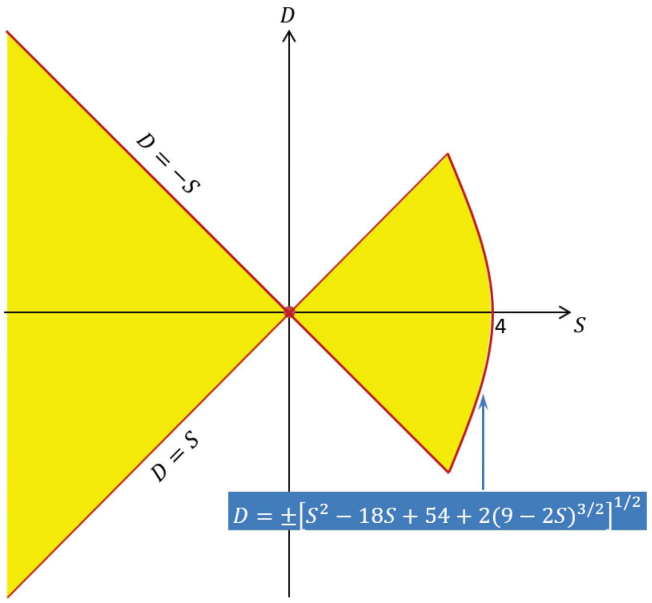


FIGURE 2. Region of strict convexity for the parameters of Van Laar's law.

To make sure that formulae (4.7) and (4.8) are well-defined over $x \in (0,1)$, the denominator $A_{12}x + A_{21}(1-x)$ must keep the same sign. This amounts to requiring that

$$A_{12}A_{21} > 0. \tag{4.9}$$

Besides, the pair (A_{12}, A_{21}) must be further restricted in order to comply with Hypotheses 3.7.

Proposition 4.3. *Let $S = A_{12} + A_{21}$ and $D = A_{12} - A_{21}$. Then, the Gibbs energy function g_α associated with Van Laar's law fulfills Hypotheses 3.7 if and only if*

$$(S, D) \in \mathcal{R}_- \cup \mathcal{R}_+, \tag{4.10a}$$

where

$$\mathcal{R}_- = \{S < 0 \text{ and } |D| < -S\}, \tag{4.10b}$$

$$\mathcal{R}_+ = \left\{0 < S < 4 \text{ and } |D| < \min\left(S; \left[S^2 - 18S + 54 + 2(9-2S)^{3/2}\right]^{1/2}\right)\right\}. \tag{4.10c}$$

The region indicated by (4.10) is colored in yellow in Figure 2. It lies inside the cone $D^2 < S^2$ that corresponds to condition (4.9). The origin $(0,0)$ must be excluded.

Proof. The first derivative of g_α is

$$g'_\alpha(x) = \ln x - \ln(1-x) + A_{12}A_{21}\frac{A_{21}(1-x)^2 - A_{12}x^2}{[A_{12}x + A_{21}(1-x)]^2}.$$

Under assumption (4.9), this is a continuous function over $(0,1)$, with $\lim_{x\downarrow 0} g'_\alpha(x) = -\infty$ and $\lim_{x\uparrow 1} g'_\alpha(x) = +\infty$. Thus, g'_α has range in $\mathbb{R}$.

The second derivative of g_α, multiplied by $x(1-x)$ to get rid of singularities, is equal to

$$h(x) := x(1-x)g''_\alpha(x) = 1 - 2A_{12}^2A_{21}^2\frac{x(1-x)}{(A_{12}x + A_{21}(1-x))^3}.$$

Let us change the variable to $y = x - \frac{1}{2} \in \left(-\frac{1}{2}, \frac{1}{2}\right)$ to work with the more symmetric function

$$H(y) := h\left(x - \frac{1}{2}\right) = 1 - 2A_{12}^2A_{21}^2\frac{\frac{1}{4} - y^2}{\left[\frac{1}{2}(A_{12}+A_{21}) + (A_{12}-A_{21})y\right]^3}.$$

It remains to study H in order to determine the region of strict positivity $H(y) > 0$. The technical details can be found in [16] or Proposition 3.3 in [34]. □

5. Cubic equation of states for two-phase mixtures

The fugacity laws investigated in Section 4 are simple and apply to a selected phase α, regardless of the remaining ones. We are now going to examine a prominent category of laws for a two-phase (gas and liquid) mixture, in which the fugacity coefficients for both phases are computed in a "simultaneous" way. Throughout the rest of this paper, it is therefore assumed that

$$\mathscr{P} = \{G, L\}, \qquad P = 2. \tag{5.1}$$

The new labels G (gas) and L (liquid) are aimed at being more meaningful. To fix ideas, the presentation is done for Peng–Robinson's law [28]. The philosophy is the same for other cubic laws.

5.1. Peng–Robinson's law

5.1.1. Mixing rules and cubic equation

Each component $i \in \mathcal{K}$ in a pure state is characterized by a pair of positive parameters a^i (cohesion term) and b^i (covolume). These are highly sophisticated functions of the pressure and the temperature, but at fixed (P, T) can be considered as constants. This gives rise at fixed (P, T) to a pair of positive dimensionless parameters

$$A^i = \frac{\mathsf{P}a^i}{(\mathsf{RT})^2}, \qquad B^i = \frac{\mathsf{P}b^i}{\mathsf{RT}}. \tag{5.2}$$

As before, we shall never write down explicitly the dependency of $\left(A^i, B^i\right)$ on (P, T).

A multicomponent mixture is supposed to behave approximately as a fictitious pure component endowed with an averaged value for the pair (A, B). The latter is computed from the $\left(A^i, B^i\right)$'s and the current partial fractions by means of a *mixing rule*. More specifically, let $\boldsymbol{x} \in \bar{\Omega}$ be the partial fractions of some phase. There can be found [27] a wide variety of mixing rules $\boldsymbol{x} \mapsto (A(\boldsymbol{x}), B(\boldsymbol{x}))$. We require mixing rules to be *smooth* and to satisfy the compatibility relation $\left(A(\boldsymbol{\delta}^i), B(\boldsymbol{\delta}^i)\right) = \left(A^i, B^i\right)$ for all $i \in \mathcal{K}$. We recall that $\boldsymbol{\delta}^i = (\delta_{i,1}, \delta_{i,2}, \ldots, \delta_{i,K-1})$ was introduced in Section 2.1.2 for $i \in \mathcal{K}$ and consists of elementary Kronecker symbols $\delta_{i,j}$.

The next step is to consider the cubic equation

$$Z^3(\boldsymbol{x}) + (B(\boldsymbol{x}) - 1)Z^2(\boldsymbol{x}) + \left[A(\boldsymbol{x}) - 2B(\boldsymbol{x}) - 3B^2(\boldsymbol{x})\right]Z(\boldsymbol{x}) + \left[B^2(\boldsymbol{x}) + B^3(\boldsymbol{x}) - A(\boldsymbol{x})B(\boldsymbol{x})\right] = 0 \tag{5.3}$$

in the variable $Z(\boldsymbol{x})$. This accounts for the name "cubic EOS." The dimensionless quantity $Z(\boldsymbol{x})$, called *compressibility factor*, can be intuitively understood by noting that for a pure component, the cubic equation

$$Z^3 + (B-1)Z^2 + \left(A - 2B - 3B^2\right)Z + \left(B^2 + B^3 - AB\right) = 0 \tag{5.4}$$

simply results from an algebraic transformation of the equation of state

$$\mathsf{P} = \frac{\mathsf{RT}}{\mathsf{V} - b} - \frac{a}{\mathsf{V}^2 + 2\mathsf{V}b - b^2}, \tag{5.5}$$

using

$$Z = \frac{\mathsf{PV}}{\mathsf{RT}}, \qquad A = \frac{\mathsf{P}a}{(\mathsf{RT})^2}, \qquad B = \frac{\mathsf{P}b}{\mathsf{RT}}. \tag{5.6}$$

Thus, for an ideal gas ($a = b = 0$), we have $Z = 1$.

In the most favorable situation, there are three real roots, all greater than $B(\boldsymbol{x})$. These are then named

$$B(\boldsymbol{x}) < Z_L(\boldsymbol{x}) < Z_I(\boldsymbol{x}) < Z_G(\boldsymbol{x}). \tag{5.7}$$

In other words, the smallest root is associated with the liquid phase L, while the largest one is associated with the gas phase G. This is grounded on the physical fact that, at the same pressure and temperature, the gas phase occupies a larger volume than the liquid phase, which by (5.6) implies that $Z_G(\boldsymbol{x}) > Z_L(\boldsymbol{x})$. As for the intermediate root $Z_I(\boldsymbol{x})$, it does not have any physical meaning.

5.1.2. Gibbs functions and fugacity coefficients

Let $\alpha \in \{G, L\}$ and assume that the real root $Z_\alpha(\boldsymbol{x}) > B(\boldsymbol{x})$ is well-defined. Then, the Peng–Robinson excess Gibbs energy is defined as

$$\Psi_\alpha(\boldsymbol{x}) = Z_\alpha(\boldsymbol{x}) - 1 - \ln\big[Z_\alpha(\boldsymbol{x}) - B(\boldsymbol{x})\big] - \frac{A(\boldsymbol{x})}{2\sqrt{2}B(\boldsymbol{x})} \ln\left[\frac{Z_\alpha(\boldsymbol{x}) + \big(1+\sqrt{2}\big)B(\boldsymbol{x})}{Z_\alpha(\boldsymbol{x}) - \big(\sqrt{2}-1\big)B(\boldsymbol{x})}\right]. \tag{5.8}$$

From this, the fugacity coefficients can be deduced with the help of (2.14a).

Theorem 5.1. *The Peng–Robinson fugacity coefficients are given by*

$$\begin{aligned}\ln \Phi_\alpha^i(\boldsymbol{x}) = {} & \frac{B(\boldsymbol{x}) + \big\langle \nabla B(\boldsymbol{x}),\, \boldsymbol{\delta}^i - \boldsymbol{x}\big\rangle}{B(\boldsymbol{x})}\big[Z_\alpha(\boldsymbol{x}) - 1\big] - \ln[Z_\alpha(\boldsymbol{x}) - B(\boldsymbol{x})] \\ & + \left[\frac{B(\boldsymbol{x}) + \big\langle \nabla B(\boldsymbol{x}),\, \boldsymbol{\delta}^i - \boldsymbol{x}\big\rangle}{B(\boldsymbol{x})} - \frac{2A(\boldsymbol{x}) + \big\langle \nabla A(\boldsymbol{x}),\, \boldsymbol{\delta}^i - \boldsymbol{x}\big\rangle}{A(\boldsymbol{x})}\right] \\ & \cdot \frac{A(\boldsymbol{x})}{2\sqrt{2}B(\boldsymbol{x})} \ln\left[\frac{Z_\alpha(\boldsymbol{x}) + \big(1+\sqrt{2}\big)B(\boldsymbol{x})}{Z_\alpha(\boldsymbol{x}) - \big(\sqrt{2}-1\big)B(\boldsymbol{x})}\right],\end{aligned} \tag{5.9}$$

for all $i \in \mathcal{K}$ and for any phase $\alpha \in \{G, L\}$ in which $Z_\alpha(\boldsymbol{x}) > B(\boldsymbol{x})$ is well-defined.

Proof. Taking the gradient of (5.8), we have

$$\begin{aligned}\nabla \Psi_\alpha = {} & \left\{1 - \frac{1}{Z_\alpha - B} - \frac{A}{2\sqrt{2}B\big[Z + \big(\sqrt{2}+1\big)B\big]} + \frac{A}{2\sqrt{2}B\big[Z - \big(\sqrt{2}+1\big)B\big]}\right\}\nabla Z_\alpha \\ & + \left\{\frac{1}{Z_\alpha - B} - \frac{A\big(\sqrt{2}+1\big)}{2\sqrt{2}\big[Z + \big(\sqrt{2}+1\big)B\big]} - \frac{A\big(\sqrt{2}-1\big)}{2\sqrt{2}\big[Z - \big(\sqrt{2}+1\big)B\big]}\right. \\ & \left. + \frac{A}{2\sqrt{2}B^2} \ln\left[\frac{Z_\alpha(\boldsymbol{x}) + \big(1+\sqrt{2}\big)B(\boldsymbol{x})}{Z_\alpha(\boldsymbol{x}) - \big(\sqrt{2}-1\big)B(\boldsymbol{x})}\right]\right\}\nabla B - \frac{1}{2\sqrt{2}B} \ln\left[\frac{Z_\alpha(\boldsymbol{x}) + \big(1+\sqrt{2}\big)B(\boldsymbol{x})}{Z_\alpha(\boldsymbol{x}) - \big(\sqrt{2}-1\big)B(\boldsymbol{x})}\right]\nabla A,\end{aligned}$$

in which we dropped the variable $\boldsymbol{x}$ for clarity. By virtue of the cubic equation (5.3),

$$1-\frac{1}{Z_\alpha-B}-\frac{A}{2\sqrt{2}B\big[Z+\big(\sqrt{2}+1\big)B\big]}+\frac{A}{2\sqrt{2}B\big[Z-\big(\sqrt{2}+1\big)B\big]}=0$$

and

$$\frac{1}{Z_\alpha-B}-\frac{A\big(\sqrt{2}+1\big)}{2\sqrt{2}\big[Z+\big(\sqrt{2}+1\big)B\big]}-\frac{A\big(\sqrt{2}-1\big)}{2\sqrt{2}\big[Z-\big(\sqrt{2}+1\big)B\big]}=\frac{Z_\alpha-1}{B}.$$

Thus,

$$\nabla\Psi_\alpha=\frac{Z_\alpha-1}{B}\nabla B+\frac{A}{2\sqrt{2}B}\ln\left[\frac{Z_\alpha(\boldsymbol{x})+\big(1+\sqrt{2}\big)B(\boldsymbol{x})}{Z_\alpha(\boldsymbol{x})-\big(\sqrt{2}-1\big)B(\boldsymbol{x})}\right]\left\{\frac{1}{B}\nabla B-\frac{1}{A}\nabla A\right\}.$$

Applying (2.14a) and using (5.8), we arrive at the desired result. □

5.1.3. Two crucial questions

Formulae (5.8) and (5.9) are well-known to thermodynamicists. A delicate but less often clarified issue is to know which phase $\alpha\in\{G,L\}$ they can be applied to, especially in the unfavorable situation when equation (5.3) has only one real root greater than $B(\boldsymbol{x})$. The simple-minded idea of taking $Z_G=Z_L$ equal to this real root is of common practice in industrial codes, but is ill-advised since it gives rise to discontinuities in the Gibbs functions, as will be explained in Remark 6.1.

In fact, in the one-root scenario, two subcases have to be envisaged. If we manage to assign a "natural" phase label $\alpha=G$ or L to the real root, then the corresponding excess Gibbs energy Ψ_α is defined by (5.8), leaving its counterpart in the other phase undefined. If we do not succeed in attributing a "logical" phase label to the real root, then Ψ_α is undefined in both phases. This process is intuitive enough to describe with words, but raises two serious questions:

(1) When does the cubic equation have three real roots greater than $B(\boldsymbol{x})$ and when does it have only one real root greater than $B(\boldsymbol{x})$?
(2) When and how can a "natural" phase label be assigned to the unique real root greater than $B(\boldsymbol{x})$?

The upcoming subsections answer to these questions by working on the generic form (5.4). Part of these issues was already addressed in [18] for Van der Waals' law. We are not aware of any similar work for Peng–Robinson's law. This is why we are taking this opportunity to undertake a rigorous study.

5.2. Assignment of phase labels to roots

Instead of the polynomial

$$\Upsilon_{A,B}(Z)=Z^3+(B-1)Z^2+\big(A-2B-3B^2\big)Z+\big(B^2+B^3-AB\big),\tag{5.10}$$

which is naturally associated with (5.4), it is more convenient to work with the rational function

$$\Pi_{A,B}(Z)=\frac{1}{Z-B}-\frac{A}{Z^2+2BZ-B^2}-1,\tag{5.11}$$

obtained from $\Upsilon_{A,B}$ through division by $-(Z-B)(Z^2+2BZ-B^2)$. Insofar as the roots of $Z^2+2BZ-B^2$, namely, $-B\big(\sqrt{2}+1\big)$ and $B\big(\sqrt{2}-1\big)$, are both lesser than B, $\Pi_{A,B}$ and $\Upsilon_{A,B}$ have the same roots over $(B,+\infty)$. Since

$$\lim_{Z\downarrow B}\Pi_{A,B}(Z)=+\infty,\qquad\lim_{Z\to+\infty}\Pi_{A,B}(Z)=-1,\tag{5.12}$$

there is at least one root larger than B.

A triplet $(Z_c, A_c, B_c) \in (B, +\infty) \times (\mathbb{R}_+^*)^2$ is said to be a *critical point* if

$$\Pi_{A_c,B_c}(Z_c) = 0, \qquad \Pi'_{A_c,B_c}(Z_c) = 0, \qquad \Pi''_{A_c,B_c}(Z_c) = 0. \tag{5.13}$$

Critical points, also called *triple points*, are physically important. Here, this notion will help us divide the space of parameters into various subregions with physically distinct behaviors.

Lemma 5.2. *For Peng–Robinson' law, there is a unique critical point given by*

$$Z_c = \frac{1}{32}\left[11 + \sqrt[3]{16\sqrt{2} - 13} - \sqrt[3]{16\sqrt{2} + 13}\right], \tag{5.14a}$$

$$A_c = \frac{1}{512}\left[-59 + 3\sqrt[3]{276\,831 - 192\,512\sqrt{2}} + 3\sqrt[3]{276\,231 + 192\,512\sqrt{2}}\right], \tag{5.14b}$$

$$B_c = \frac{1}{32}\left[-1 - 3\sqrt[3]{16\sqrt{2} - 13} + 3\sqrt[3]{16\sqrt{2} + 13}\right]. \tag{5.14c}$$

Approximately,

$$Z_c \approx 0.307401308, \qquad A_c \approx 0.457235529, \qquad B_c \approx 0.077796073. \tag{5.14d}$$

In physics textbooks [27,33], only decimal approximations (5.14d) of the critical point can be found, without any proof. The interest of Lemma 5.2 is to derive the exact values (5.14a)–(5.14c) of the critical point.

Proof. System (5.13) can be turned into a set of 3 polynomial equations in (Z_c, A_c, B_c). By eliminating A_c, we obtain the cubic equation $z_c^3 - 3z_c^2 - 3z_c - 3 = 0$ in $z_c = Z_c/B_c$, whose only real root is

$$z_c = 1 + \sqrt[3]{4 - 2\sqrt{2}} + \sqrt[3]{4 + 2\sqrt{2}} \approx 3.951373036. \tag{5.15}$$

From this A_c/B_c can be deduced exactly. Once this is done, we can get back to (Z_c, A_c, B_c). See Lemma 3.3 in [34] for more details. □

Theorem 5.3 (Supercritical and subcritical regimes)**.**

(1) *If $B/A > B_c/A_c \approx 0.170144420$, the function $\Pi_{A,B}$ is decreasing over $(B, +\infty)$ and has only one zero greater than B.*

(2) *If $B/A < B_c/A_c \approx 0.170144420$, the function $\Pi_{A,B}$ has two disctinct local extrema. In other words, there exist two distinct values $\zeta_L < \zeta_G$ in $(B, +\infty)$ such that*

$$\Pi'_{A,B}(\zeta_L) = \Pi'_{A,B}(\zeta_G) = 0.$$

Then, $\Pi_{A,B}$ is decreasing on (B, ζ_L), increasing on (ζ_L, ζ_G) and decreasing on $(\zeta_G, +\infty)$. It may have one or three distinct zeros over $(B, +\infty)$.

Theorem 5.3 paves the way to a natural association of a root with a phase in the subcritical regime. Note that B/A plays the role of a temperature (up to a multiplicative constant).

Definition 5.4 (Phase label assignment)**.** The region $0 < B < (B_c/A_c)A$ is said to be *subcritical*. In the subcritical region, a root $Z > B$ of the cubic equation (5.4) is said to be *associated with the liquid phase* L if $Z < \zeta_L$; a root $Z > B$ of the cubic equation (5.4) is said to be *associated with the gas phase* G if $Z > \zeta_G$.

Let us elaborate on this definition before proving Theorem 5.3. If there is only one root $Z > B$, this root cannot belong to (ζ_L, ζ_G). Therefore, either $Z \in (B, \zeta_L)$ or $Z \in (\zeta_G, +\infty)$. This way of assigning a phase label to Z is most natural, since it extends by continuity the "topological" pattern observed in the case of three real roots.

The region $B > (B_c/A_c)A$ is said to be *supercritical*. The graph of $\Pi_{A,B}$ no longer has two discernable branches. In this configuration, there is no natural way to associate Z with a phase. Physically speaking, the critical threshold B_c/A_c corresponds to a critical temperature T_c. Above the critical temperature, the distinction between gas and liquid phases no longer holds [12].

Proof. To find the local extrema of $\Pi_{A,B}$ on $(B,+\infty)$, we search for the zeros on $(B,+\infty)$ of

$$\Pi'_{A,B}(Z) = -\frac{1}{(Z-B)^2} + \frac{A(2Z+2B)}{\left(Z^2+2BZ-B^2\right)^2},$$

or equivalently, of the polynomial

$$Q_{A,B}(Z) = -\left(Z^2+2BZ-B^2\right)^2 + 2A(Z+B)(Z-B)^2,$$

which is equal to $(Z-B)^2\left(Z^2+2BZ-B^2\right)^2\Pi'_{A,B}(Z)$. An even more convenient choice is to set $\mathfrak{T} = (Z-B)/B \in (0,+\infty)$ and to study

$$q_{A,B}(\mathfrak{T}) := \frac{1}{B^4}\,Q_{A,B}(B\mathfrak{T}+B) = -\left(\mathfrak{T}^2+4\mathfrak{T}+2\right)^2 + 2\,\frac{A}{B}\,(\mathfrak{T}+2)\mathfrak{T}^2. \tag{5.16}$$

The key idea is to insert A_c/B_c and to put the latter function under the form

$$q_{A,B}(\mathfrak{T}) = (\mathfrak{T}-\mathfrak{T}_c)\left(q_0+q_1\mathfrak{T}+q_2\mathfrak{T}^2\right) + 2\left(\frac{A}{B}-\frac{A_c}{B_c}\right)(\mathfrak{T}+2)\mathfrak{T}^2$$

where $\mathfrak{T}_c = z_c - 1$ and $q_0 < 0$, $q_1 < 0$, $q_2 < 0$. See Theorem 3.5 in [34] for the calculation of q_0, q_1, q_2. Thus, the graph of q_{A_c,B_c} is tangent to the $\mathfrak{T}$-axis at $\mathfrak{T} = \mathfrak{T}_c$ with $q_{A_c,B_c}(\mathfrak{T}) \leqslant 0$ for $\mathfrak{T} \geqslant 0$. If $A/B > A_c/B_c$, then $q_{A,B}(\mathfrak{T}_c) > 0$ and $q_{A,B}$ vanishes twice on $(0,+\infty)$. If $A/B < A_c/B_c$, then $q_{A,B}(\mathfrak{T}) < q_{A_c,B_c}(\mathfrak{T})$ for all $\mathfrak{T} > 0$ and $q_{A,B}$ does not vanish on $(0,+\infty)$. □

5.3. Three-root and one-root regions

We also have the following necessary (and perhaps sufficient) condition for the existence of three real roots greater than B. To the best of our knowledge, the frontier given by (5.17) has never been investigated before.

Theorem 5.5. *In the quarter-plane $(A,B) \in \left(\mathbb{R}_+^*\right)^2$, the region for which Peng–Robinson's cubic equation (5.4) has three real roots, all greater than B, is contained in the region*

$$\{0 < B < B_c, \quad A_G(B) < A < A_L(B)\}, \tag{5.17a}$$

where $A_G(B)$ and $A_L(B)$ are respectively the middle root and greatest roots of the cubic equation

$$\begin{aligned} -4A^3 - \left(8B^2-40B-1\right)A^2 + \left(16B^4-112B^3-88B^2-8B\right)A \\ + \left(32B^6+128B^5+160B^4+64B^3+8B^2\right) = 0. \end{aligned} \tag{5.17b}$$

The region (5.17) lies itself inside the subcritical domain $0 < B < (B_c/A_c)A$. Moreover,

- *for $\{0 < B < B_c,\ (A_c/B_c)B < A < A_G(B)\}$, the only real root is associated with the gas phase G, in the sense of Definition 5.4;*
- *for $\{0 < B < B_c,\ A_L(B) < A\}$ or $\{B_c < B,\ (A_c/B_c)B < A\}$, the only real root is associated with the liquid phase L, in the sense of Definition 5.4.*

The region characterized by (5.17) is colored in cyan in Figure 3. Inside it, Peng–Robinson's cubic equation (5.4) has three real roots. Nevertheless, we could not prove that all the roots are greater than B, despite abundant numerical evidences supporting the validity of this claim. The first branch $A_G(\cdot)$ starts at $(A,B) = (0,0)$ with slope $A'_G(B=0) = 4+2\sqrt{2}$. The second branch $A_L(\cdot)$ starts at $(A,B) = (1/4,0)$ with slope $A'_L(B=0) = 2$. Both branches end at $(A,B) = (A_c,B_c)$, with the common slope $A'_G(B=B_c) = A'_L(B=B_c) \approx 2.95686087$.

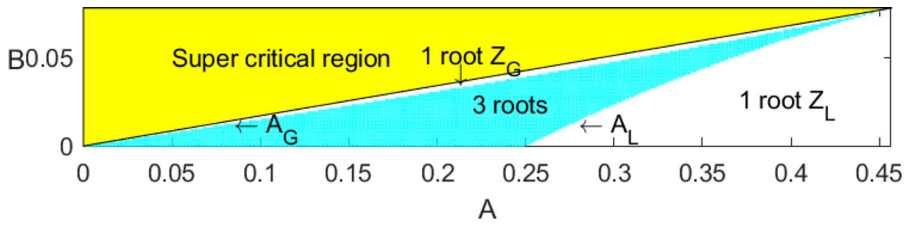


FIGURE 3. Number of roots for Peng–Robinson's law in the (A, B)-quarter plane.

Proof. The discriminant[3] of the cubic equation (5.4) is

$$\begin{aligned}\Delta(A,B) = &-4A^3 - \big(8B^2 - 40B - 1\big)A^2 + \big(16B^4 - 112B^3 - 88B^2 - 8B\big)A \\ &+ \big(32B^6 + 128B^5 + 160B^4 + 64B^3 + 8B^2\big).\end{aligned} \tag{5.18}$$

For (5.4) to have three real roots, $\Delta(A, B)$ must be positive. If the polynomial (5.18) has only one real root $A_0(B)$, since the leading coefficient -4 is negative, we must have $A < A_0(B)$ to ensure $\Delta(A, B) > 0$. If (5.18) has three real roots $A_0(B) < A_G(B) < A_L(B)$, we must have $A < A_0(B)$ or $A \in (A_G(B), A_L(B))$. The discriminant of (5.18) with respect to A is equal to

$$\Delta_A(B) = -32B^2\big(64B^3 + 6B^2 + 12B - 1\big) \cdot \big(4096B^6 + 768B^5 + 1572B^4 + 16B^3 + 132B^2 - 24B + 1\big).$$

It can be shown that $\Delta_A(B) > 0$ for $B \in (0, B_c)$, $\Delta_A(B_c) = 0$ and $\Delta_A(B) < 0$ for $B > B_c$. Therefore, if $B > B_c$, only $A_0(B)$ exists. If $B \in (0, B_c)$, there exist $A_0(B) < A_G(B) < A_L(B)$.

The rest of the proof goes as follows. We first show that $A_0(B) > 0$. Then, we rule out the region $\{0 < B < B_c, 0 < A < A_0(B)\}$ which in fact belongs to the supercritical region. Next, in the region $\{(A_c/B_c)B < A < A_0(B), B_* < B\}$, where $B_* \approx 2.435425$ is the ordinate at which the graphe of $A_0(\cdot)$ enters the subcritical region, we show that the three real real roots cannot be all larger than B. In conclusion, the only way for (5.4) to have three real roots, all greater than B, is that $B \in (0, B_c)$ and $A \in (A_G(B), A_L(B))$. This region is shown to be contained in the subcritical domain. The comprehensive discussion can be found in Theorem 3.6 from [34]. □

6. Domain extension for cubic EOS-based Gibbs functions

From Section 5.3, it appears that the cubic equation (5.3) may not always have three real roots greater than $B(\boldsymbol{x})$ for all $\boldsymbol{x} \in \bar{\Omega}$. As a consequence, the domain of definition for the functions Ψ_α, Φ^i_α for a given phase α may not always cover the whole simplex $\bar{\Omega}$. This turns out to be detrimental to the unified formulation (2.18).

6.1. Difficulty of the unified formulation with cubic EOS

In a nutshell, the Gibbs energy functions g_α may grossly violate Hypotheses 3.7. To give a visual picture of the nature of the obstruction, let us consider the simplistic case of a two-phase binary ($K = 2$) mixture, governed by Peng–Robinson's law combined with the mixing rule

$$A(x) = \Big[x\sqrt{A^{\mathrm{I}}} + (1-x)\sqrt{A^{\mathrm{II}}}\Big]^2, \qquad B(x) = xB^{\mathrm{I}} + (1-x)B^{\mathrm{II}}. \tag{6.1}$$

For an arbitrary choice of the two pairs $\big(A^{\mathrm{I}}, B^{\mathrm{I}}\big)$ and $\big(A^{\mathrm{II}}, B^{\mathrm{II}}\big)$ in the subcritical region $0 < B < (B_c/A_c)A$, the parametrized curve $\gamma : [0, 1] \ni x \mapsto (A(x), B(x)) \in \big(\mathbb{R}^*_+\big)^2$ is an arc of parabola.

[3]The discriminant of the equation $aX^3 + bX^2 + cX + d = 0$ is $\Delta = b^2c^2 - 4ac^3 - 4b^3d - 27a^2d^2 + 18abcd$.

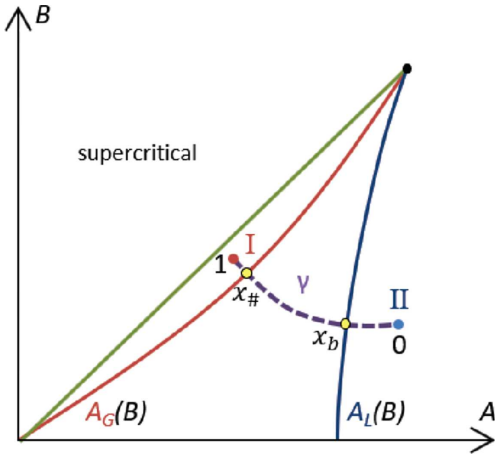


FIGURE 4. Curve γ defined by the mixing rule in the (A, B)-plane.

Assume that $\left(A^{\mathrm{I}}, B^{\mathrm{I}}\right)$ belongs to the one-root region labelled G, while $\left(A^{\mathrm{II}}, B^{\mathrm{II}}\right)$ belongs to the one-root region labelled L, as illustrated in Figure 4. At $x = 0$, the curve γ starts from $\left(A^{\mathrm{II}}, B^{\mathrm{II}}\right)$ in the L-root region. At some parameter value $x = x_\flat \in (0, 1)$, it enters the three-root region. At some further value $x = x_\sharp \in (x_\flat, 1)$, it exits the three-root region. At $x = 1$, it finally meets $\left(A^{\mathrm{I}}, B^{\mathrm{I}}\right)$ in the G-root region. It is not difficult to realize that:

- the quantities $Z_L(x)$, $\Psi_L(x)$, $g_L(x)$ are well-defined only for $x \in [0, x_\sharp]$; $g_L\left(x_\sharp^-\right)$ and $g'_L\left(x_\sharp^-\right)$ remain bounded, while $g''_L\left(x_\sharp^-\right)$ and $Z'_L\left(x_\sharp^-\right)$ blow up;
- the quantities $Z_G(x)$, $\Psi_G(x)$, $g_G(x)$ are well-defined only for $x \in [x_\flat, 1]$; $g_G\left(x_\flat^+\right)$ and $g'_G\left(x_\flat^+\right)$ remain bounded, while $g''_G\left(x_\flat^+\right)$ and $Z'_G\left(x_\flat^+\right)$ blow up.

Since $g'_G\left(x_\flat^+\right)$ and $g'_L\left(x_\sharp^-\right)$ are finite, the image sets $g'_G([x_\flat, 1))$ and $g'_L((0, x_\sharp])$ are not equal to $\mathbb{R}$. This prevents us from being always able to assign a well-defined extended fraction to the vanishing phase for the single-phase problem (3.24) of Section 3.3.2. Indeed, when c is sufficiently close to 0, $g'_L(c) \notin g'_G([x_\flat, 1))$ because $\lim_{x \downarrow 0} g'_L(x) = -\infty$, and it is impossible to find $\bar{x}_G \in [x_\flat, 1)$ such that $g'_G(\bar{x}_G) = g'_L(c)$. Likewise, when c is sufficiently close to 1, $g'_G(c) \notin g'_L((0, x_\sharp])$ because $\lim_{x \uparrow 1} g'_G(x) = +\infty$, and it is impossible to find $\bar{x}_L \in (0, x_\sharp]$ such that $g'_L(\bar{x}_L) = g'_G(c)$. Figure 5 depicts this situation.

It could be argued that the same flaw of cubic EOS laws should cause the same prejudice to the variable-switching formulation of Section 2.2.1. But this is not true. In the variable-switching formulation, if the context is correctly guessed, we do not need to compute anything from the absent phase and the above problem is irrelevant. If the context is incorrectly alleged, the flash does not converge or may even crash, but there is an opportunity for us to make up for it by changing the context. The natural variable formulation does not seek to explore the regions where information is missing. The unified formulation has to do so, by its very vocation to treat all phases on an equal footing.

Remark 6.1. From Figure 5, it can be seen that if we abruptly take $Z_G = Z_L$ in the one-root regions $x \in [0, x_\flat)$ and $x \in (x_\sharp, 1]$, as often done by practitioners, then we will have $g_G = g_L$ over these two intervals. As a consequence, g_G will exhibit a discontinuity at $x_\flat$ and g_L a discontinuity at $x_\sharp$. These unphysical discontinuities are not a favorable feature for numerical robustness.

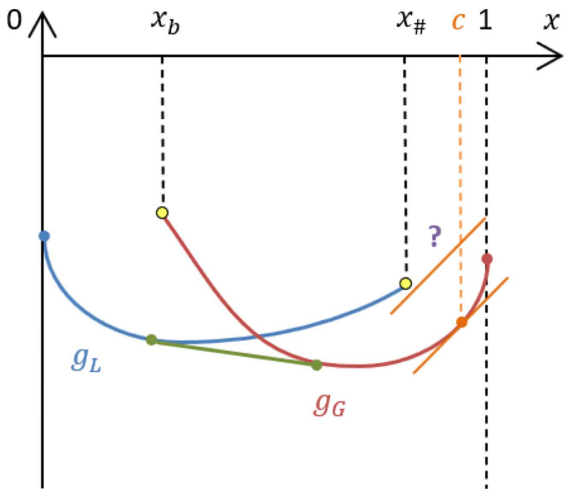


FIGURE 5. Typical situation where the fraction in the absent phase cannot be computed.

6.2. A natural domain extension method

To enhance the performance of the unified formulation, it is essential that the domains of definition for the excess functions Ψ_α can be extended to $\bar{\Omega}$. Note that here we just want to extend the domains of definition of various functions. We do not strive to fulfill Hypotheses 3.7, since these assumptions may already be violated for the original unextended Gibbs functions. Even without strict convexity, a smooth extension of the Gibbs functions helps iterative methods [35] to remain well-defined everywhere.

6.2.1. Construction in the one-root region

When the cubic equation does not have three real roots, our idea is to use the common real part of the two complex conjugate roots, as a "surrogate" of the missing real root. Assume that Z_α is the only real root greater than B of Peng–Robinson's cubic equation

$$Z^3 + (B-1)Z^2 + \big(A - 2B - 3B^2\big)Z + \big(B^2 + B^3 - AB\big) = 0,$$

where the label $\alpha \in \{G, L\}$ has been assigned in accordance with Definition 5.4. To alleviate notations, we do not explicitly indicate the dependency of A, B and Z on $\boldsymbol{x}$.

Let β be the other phase, that is, $\beta = L$ if $\alpha = G$ and $\beta = G$ if $\alpha = L$. Since the sum of the three (complex) roots is equal to $1 - B$, the two remaining conjugate roots share the common real part

$$W_\beta = \frac{1 - B - Z_\alpha}{2}. \tag{6.2}$$

The following properties of W_β speak in favor of its enrollment as a substitute for Z_β, which would have been subject the same constraints, had it existed.

Lemma 6.2. *Let (A, B) be a pair in the subcritical region $0 < B < (B_c/A_c)A$ and assume that Peng–Robinson's cubic equation has only one real root $Z_\alpha > B$ that corresponds to phase α.*

(1) *If $B < 0.206813$, then*

$$W_\beta > B. \tag{6.3a}$$

(2) *If $B < 0.137072$, then*

$$Z_\alpha < W_\beta \ \text{ if } \ \alpha = L, \qquad W_\beta < Z_\alpha \ \text{ if } \ \alpha = G. \tag{6.3b}$$

Proof. The proof is based on the rational function $\Pi_{A,B}$ introduced in (5.11) and its behavior described in Theorem 5.3. Full details are available in Lemma 3.5 of [34]. □

By restricting ourselves to $B < 0.137072$, which is reasonable since $B_c \approx 0.077796$, we can rely on Lemma 6.2 to stipulate that

$$\Psi_\beta = W_\beta - 1 - \ln\big[W_\beta - B\big] - \frac{A}{2\sqrt{2}B}\ln\left[\frac{W_\beta + \big(\sqrt{2}+1\big)B}{W_\beta - \big(\sqrt{2}-1\big)B}\right] \tag{6.4}$$

for the missing phase β. By (2.14a), we can derive the corresponding fugacity coefficients.

Theorem 6.3. *With extension (6.4), the Peng–Robinson fugacity coefficients phase β are*

$$\begin{aligned}\ln\Phi_\beta^i = {}& \frac{B + \big\langle \nabla B,\, \boldsymbol{\delta}^i - \boldsymbol{x}\big\rangle}{B}\,[W_\beta - 1] - \ln[W_\beta - B] \\ &+ \left[\frac{B + \big\langle \nabla B,\, \boldsymbol{\delta}^i - \boldsymbol{x}\big\rangle}{B} - \frac{2A + \big\langle \nabla_{\boldsymbol{x}} A,\, \boldsymbol{\delta}^i - \boldsymbol{x}\big\rangle}{A}\right]\frac{A}{2\sqrt{2}B}\ln\left[\frac{W_\beta + \big(\sqrt{2}+1\big)B}{W_\beta - \big(\sqrt{2}-1\big)B}\right] \\ &+ \left[\frac{\big\langle \nabla W_\beta,\, \boldsymbol{\delta}^i - \boldsymbol{x}\big\rangle}{W_\beta} - \frac{\big\langle \nabla B,\, \boldsymbol{\delta}^i - \boldsymbol{x}\big\rangle}{B}\right]\frac{W_\beta\,\Upsilon_{A,B}(W_\beta)}{(W_\beta - B)\Big(W_\beta^2 + 2BW_\beta - B^2\Big)}\end{aligned} \tag{6.5}$$

for all $i \in \mathcal{K}$, with $\Upsilon_{A,B}(W) = W^3 + (B-1)W^2 + \big(A - 2B - 3B^2\big)W + \big(B^2 + B^3 - AB\big)$.

Proof. The proof is similar to that of Theorem 5.1. □

The gradient of W_β required by (6.5) can be computed by $\nabla W_\beta = -\frac{1}{2}(\nabla B + \nabla Z_\alpha)$, in which ∇Z_α comes from differentiating Peng–Robinson's cubic equation with respect to $\boldsymbol{x}$, *i.e.*,

$$\begin{aligned}\big[3Z_\alpha^2 + 2(B-1)Z_\alpha + \big(A - 2B - 3B^2\big)\big]\nabla Z_\alpha = {}& (B - Z_\alpha)\nabla A \\ &+ \big(A - 2B - 3B^2 + 6BZ_\alpha + 2Z_\alpha - Z_\alpha^2\big)\nabla B.\end{aligned} \tag{6.6}$$

6.2.2. Alteration in the three-root region

In the one-root region, the gradient $\nabla W_\beta = -\frac{1}{2}(\nabla B + \nabla Z_\alpha)$ remains bounded. If we start from the three-root region and move towards the transition boundary where Z_β disappears, the gradient ∇Z_β does not remain bounded. Indeed, Z_β also obeys (6.6) (just replace Z_α by Z_β), and as Z_β gets closer to being a double root, ∇Z_β blows up. However, we need a finite gradient ∇Z_β for the numerical resolution of system (2.18) by, say, the Newton method.

To achieve a smooth junction, we introduce a further approximation on a tiny portion of the three-root region. Assuming that we are in the three-root region, with $B < Z_L < Z_I < Z_G$, we introduce

$$\vartheta = \frac{Z_I - Z_L}{Z_G - Z_L} \in [0,1] \tag{6.7}$$

as an indicator of the closeness to the transition boundary. Indeed, the cubic equation has double roots when $\vartheta = 0$ or $\vartheta = 1$. Let $\varepsilon \in (0, 1/4)$ be a small threshold.

- If $\vartheta \in [2\varepsilon, 1-2\varepsilon]$, we apply the usual formulas for the case of three real-roots.
- If $\vartheta \in (1-2\varepsilon, 1]$, the two roots Z_I and Z_G are close to each other. We keep Z_L but progressively replace Z_G by $W_G = \frac{1}{2}(1 - B - Z_L) = \frac{1}{2}(Z_I + Z_G)$ whose gradient is bounded. Instead of Z_G, we plug $\widetilde{Z}_G = [1 - \nu_G(\vartheta)]Z_G + \nu_G(\theta)W_G$ into (5.8), where

$$\nu_G(\vartheta) = \begin{cases} 0 & \text{if } \vartheta \leqslant 1-2\varepsilon, \\ q((\vartheta - (1-2\varepsilon))/\varepsilon) & \text{if } \vartheta \in (1-2\varepsilon, 1-\varepsilon), \\ 1 & \text{if } \vartheta \geqslant 1-\varepsilon, \end{cases} \tag{6.8}$$

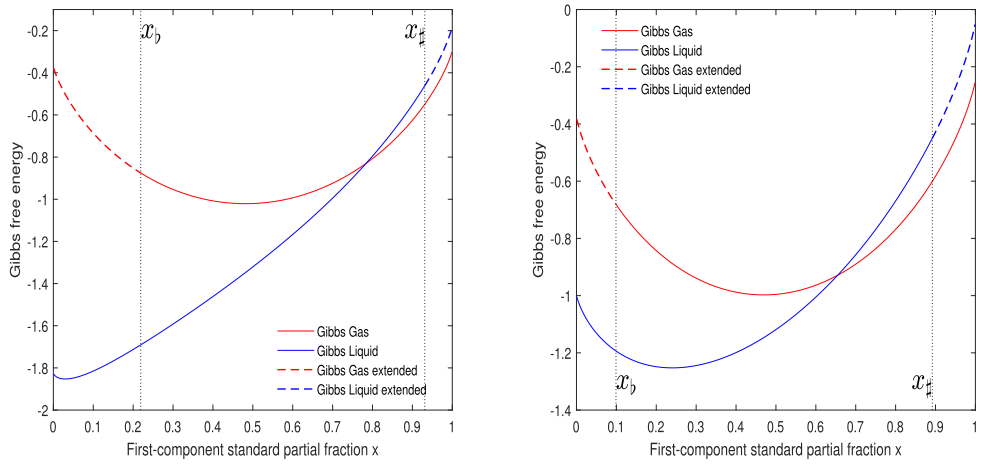


FIGURE 6. Extended Gibbs energy functions g_L (blue) and g_G (red) for Peng–Robinson's law by the indirect method, with $\varepsilon = 0.001$. *Left panel*: $\left(A^{\mathrm{I}}, B^{\mathrm{I}}\right) = (0.322, 0.053)$ and $\left(A^{\mathrm{II}}, B^{\mathrm{II}}\right) = (0.33, 0.03)$. *Right panel*: $\left(A^{\mathrm{I}}, B^{\mathrm{I}}\right) = (0.275, 0.045)$ and $\left(A^{\mathrm{II}}, B^{\mathrm{II}}\right) = (0.35, 0.04)$.

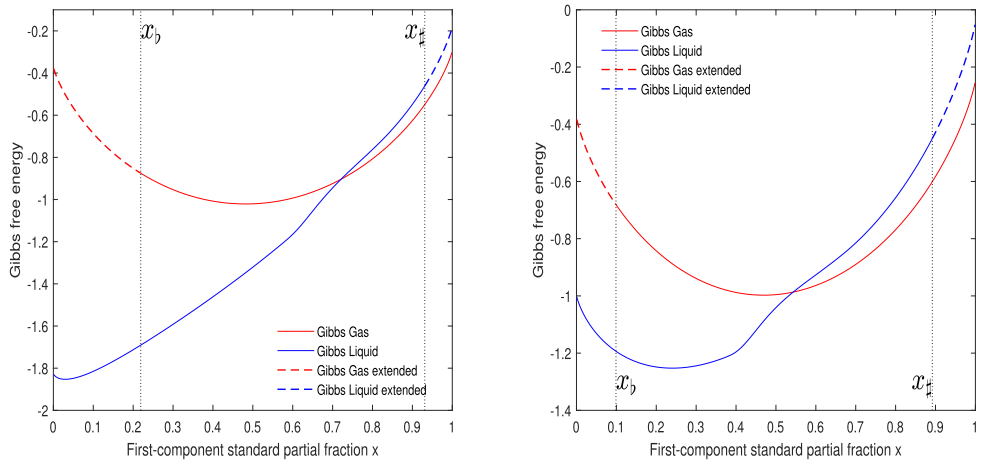


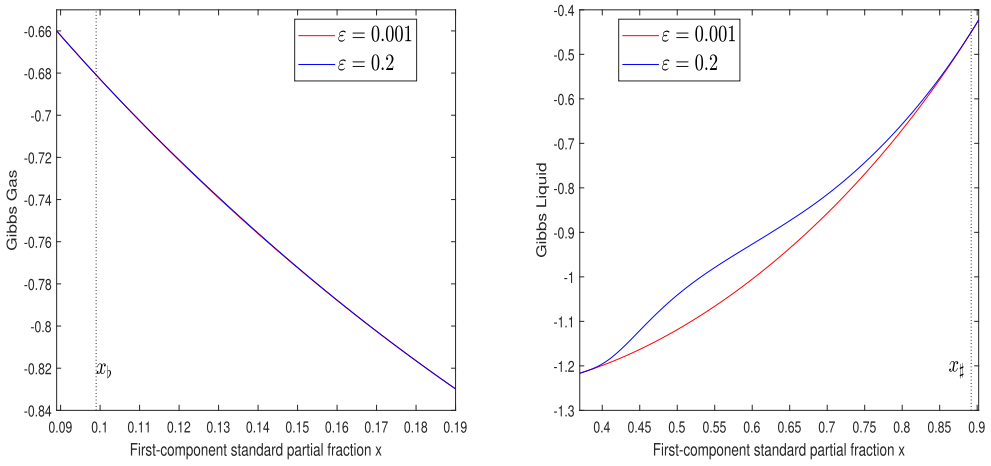
FIGURE 7. Extended Gibbs energy functions g_L (blue) and g_G (red) for Peng–Robinson's law by the indirect method, with $\varepsilon = 0.2$. *Left panel*: $\left(A^{\mathrm{I}}, B^{\mathrm{I}}\right) = (0.322, 0.053)$ and $\left(A^{\mathrm{II}}, B^{\mathrm{II}}\right) = (0.33, 0.03)$. *Right panel*: $\left(A^{\mathrm{I}}, B^{\mathrm{I}}\right) = (0.275, 0.045)$ and $\left(A^{\mathrm{II}}, B^{\mathrm{II}}\right) = (0.35, 0.04)$.

and $q(y) = y^2(3-2y)$. The rescaled function $y \mapsto q(y/\varepsilon)$ serves as a C^1 step function over the interval $[0, \varepsilon]$. We note that $q(0) = 0$, $q(1) = 1$ and $q'(0) = q'(1) = 0$. From the modified excess Gibbs energy

$$\Psi_G = \widetilde{Z}_G - 1 - \ln\left[\widetilde{Z}_G - B\right] - \frac{A}{2\sqrt{2}B} \ln\left[\frac{\widetilde{Z}_G + \left(\sqrt{2}+1\right)B}{\widetilde{Z}_G - \left(\sqrt{2}-1\right)B}\right] \tag{6.9a}$$

and from the rule (2.14a), the fugacity coefficients are inferred as

$$\ln \Phi_G^i = \frac{B + \left\langle \nabla B, \boldsymbol{\delta}^i - \boldsymbol{x} \right\rangle}{B} \left[\widetilde{Z}_G - 1\right] - \ln\left[\widetilde{Z}_G - B\right]$$



(A) (B)

FIGURE 8. Close-up comparison of the extended Gibbs functions between $\varepsilon = 0.001$ and $\varepsilon = 0.2$ for Peng–Robinson' law with the indirect method. $(A^{\mathrm{I}}, B^{\mathrm{I}}) = (0.275, 0.045)$ and $(A^{\mathrm{II}}, B^{\mathrm{II}}) = (0.35, 0.04)$. (A) g_G. (B) g_L.

$$+\left[\frac{B+\langle\nabla B,\,\boldsymbol{\delta}^i-\boldsymbol{x}\rangle}{B}-\frac{2A+\langle\nabla_{\boldsymbol{x}}A,\,\boldsymbol{\delta}^i-\boldsymbol{x}\rangle}{A}\right]\frac{A}{2\sqrt{2}B}\ln\left[\frac{\widetilde{Z}_G+\left(\sqrt{2}+1\right)B}{\widetilde{Z}_G-\left(\sqrt{2}-1\right)B}\right]$$
$$+\left[\frac{\langle\nabla\widetilde{Z}_G,\,\boldsymbol{\delta}^i-\boldsymbol{x}\rangle}{\widetilde{Z}_G}-\frac{\langle\nabla B,\,\boldsymbol{\delta}^i-\boldsymbol{x}\rangle}{B}\right]\frac{\widetilde{Z}_G\,\Upsilon_{A,B}\left(\widetilde{Z}_G\right)}{\left(\widetilde{Z}_G-B\right)\left(\widetilde{Z}_G^2+2B\widetilde{Z}_G-B^2\right)}. \tag{6.9b}$$

– If $\vartheta \in [0, 2\varepsilon)$, we proceed in a similar and symmetric fashion to replace Z_L by $\widetilde{Z}_L = [1-\nu_L(\vartheta)]Z_L + \nu_L(\vartheta)W_L$ in the expression of Ψ_L, while preserving Z_G.

Figures 6 and 7 display a few examples of the extension method for Peng–Robinson's law in the binary case. Figures 8 and 9 provide a close-up comparison between two choices of ε.

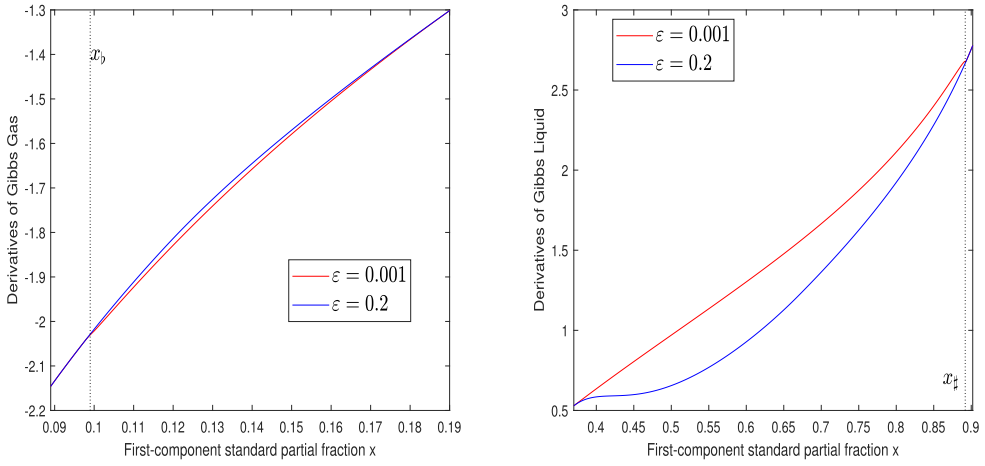
6.2.3. Numerical validation of the extension method

Extensive numerical simulations are provided in [35], a companion paper to the present one, to demonstrate the relevance of the above extension method. In [35], we considered various systems of equations in the unified form (2.18) with a wide range of physical parameters and initial points. A careful comparison is carried out between two numerical methods used to solve these systems: the Newton-min method and the *Non-Parametic Interior Point Method* (NPIPM) that we designed on purpose to deal with such systems.

It turned out that very good results (nearly 100% convergence) can be achieved thanks to the combination of both ingredients, *i.e.*, the extension of Gibbs functions and the NPIPM algorithm. A single ingredient alone is not enough in the following sense: unextended Gibbs functions always cause divergence of both numerical methods (Newton-min and NPIPM), but extended Gibbs functions combined with Newton-min does not bring significant improvement.

7. CONCLUSION

Beyond implementational advantages, the unified formulation has been shown to be able to recover all the properties known to physicists on phase equilibrium. Indeed, the complementary equations do encapsulate the tangent plane criterion (Thm. 3.4), as the sign information is related to some stability condition. The unified



(A) (B)

FIGURE 9. Close-up comparison of the derivative of the extended Gibbs functions between $\varepsilon = 0.001$ and $\varepsilon = 0.2$ for Peng–Robinson' law with the indirect method. $(A^{\mathrm{I}}, B^{\mathrm{I}}) = (0.275, 0.045)$ and $(A^{\mathrm{II}}, B^{\mathrm{II}}) = (0.35, 0.04)$. (A) g'_G. (B) g'_L.

formulation can also be regarded as a solution of some constrained minimization problem (Thms. 3.2 and 3.3) in which the objective function is a Gibbs energy of the mixture. This solution is characterized by a set of equations that is slightly stronger than the KKT optimality conditions when a phase vanishes.

The possibility of assigning well-defined values to the extended fractions of an absent phase can only be achieved if the Gibbs functions meet some restrictive requirements. Strictly convexity and surjectivity of the gradient over the whole domain of fractions (Hypotheses 3.7) are sufficient for this purpose. Remarkably, these assumptions also guarantee the existence of a solution to the phase equilibrium problem in the unified formulation (Thm. 3.12).

Unfortunately, Hypotheses 3.7 are not satisfied by realistic Gibbs functions. The obligation of assigning well-defined fraction values to an absent phase then becomes a weakness that jeopardizes the whole unified approach. This is especially true for the Gibbs functions derived from cubic EOS, for which they are not even defined on the whole domain of fractions. The extension procedure of Section 6 is aimed at improving the robustness of the unified formulation. The corresponding numerical results will be the subject of another simulation-oriented article, where we also design an interior-point algorithm ([34], Sect. 5) in order to efficiently cope with complementarity conditions.

Despite its solid theoretical foundation, the current unified formulation is not able to support the phenomenon of phase separation, where the same phase is split into two or several distinct subphases due to the non-convexity of its Gibbs function. Note that the variable-switching formulation cannot do it either. Future works should tackle this question perhaps by combining the unified formulation with some judicious approaches advocated by [11, 25].

REFERENCES

[1] V. Acary and B. Brogliato, Numerical Methods for Nonsmooth Dynamical Systems: Applications in Mechanics and Electronics. Vol. 35 of *Lecture Notes in Applied and Computational Mechanics*. Springer, Berlin (2008).

[2] M. Aganagić, Newton's method for linear complementarity problems. *Math. Program.* **28** (1984) 349–362.

[3] L. Asselineau, G. Bogdanic and J. Vidal, A versatile algorithm for calculating vapour-liquid equilibria. *Fluid Phase Equilib.* **3** (1979) 273–290.

[4] L. Beaude, K. Brenner, S. Lopez, R. Masson and F. Smai, Non-isothermal compositional liquid gas Darcy flow: formulation, soil-atmosphere boundary condition and application to high-energy geothermal simulations. *Comput. Geosci.* **23** (2019) 443–470.

[5] I. Ben Gharbia, *Résolution de problèmes de complémentarité: Application à un écoulement diphasique dans un milieu poreux*. Ph.D. thesis, Université Paris Dauphine (Paris IX) (December 2012) http://tel.archives-ouvertes.fr/tel-00776617.

[6] I. Ben Gharbia and É. Flauraud, Study of compositional multiphase flow formulation using complementarity conditions. *Oil Gas Sci. Technol.* **74** (2019) 43.

[7] I. Ben Gharbia and J. Jaffré, Gas phase appearance and disappearance as a problem with complementarity constraints. *Math. Comput. Simul.* **99** (2014) 28–36.

[8] I. Ben Gharbia, É. Flauraud and A. Michel, Study of compositional multi-phase flow formulations with cubic EOS. In: Vol. 2 of SPE Reservoir Simulation Symposium, 23–25 February. Houston, Texas, USA (2015) 1015–1025.

[9] S. Boyd and L. Vandenberghe, Convex Optimization. Berichte über verteilte messysteme. Cambridge University Press, Cambridge, UK (2004).

[10] K.H. Coats, An equation of state compositional model. *SPE J.* **20** (1980) 363–376.

[11] L. Contento, A. Ern and R. Vermiglio, A linear-time approximate convex envelope algorithm using the double Legendre-Fenchel transform with application to phase separation. *Comput. Optim. Appl.* **60** (2015) 231–261.

[12] U.K. Deiters and T. Kraska, High-pressure Fluid Phase Equilibria: Phenomenology and Computation. Vol. 2 of *Supercritical Fluid Science and Technology*. Elsevier, Amsterdam (2012) http://store.elsevier.com/High-Pressure-Fluid-Phase-Equilibria/isbn-9780444563545/.

[13] R.A. Heidemann, Computation of high pressure phase equilibria. *Fluid Phase Equilib.* **14** (1983) 55–78.

[14] W. Henry and J. Banks, III. Experiments on the quantity of gases absorbed by water, at different temperatures, and under different pressures. *Phil. Trans. R. Soc. London* **93** (1803) 29–274.

[15] S. Kräutle, The semismooth Newton method for multicomponent reactive transport with minerals. *Adv. Water Res.* **34** (2011) 137–151.

[16] T.C. Lai Nguyen, *Analysis of a nonlinear algebraic system arising in phase equilibria problems*. Master's thesis, INSA Rennes (2018).

[17] A. Lauser, C. Hager, R. Helmig and B. Wohlmuth, A new approach for phase transitions in miscible multi-phase flow in porous media. *Adv. Water Res.* **34** (2011) 957–966.

[18] S. Le Vent, A summary of the properties of van der Waals fluids. *Int. J. Mech. Engrg Edu.* **29** (2001) 257–277.

[19] I. Lusetti, Numerical methods for compositional multiphase flow models with cubic EOS. Tech. report, IFPEN (2016).

[20] R. Masson, L. Trenty and Y. Zhang, Formulations of two phase liquid gas compositional Darcy flows with phase transitions. *Int. J. Finite Vol.* **11** (2014) 1–34.

[21] R. Masson, L. Trenty and Y. Zhang, Coupling compositional liquid gas Darcy and free gas flows at porous and free-flow domains interface. *J. Comput. Phys.* **321** (2016) 708–728.

[22] M.L. Michelsen, The isothermal flash problem. Part I. Stability. *Fluid Phase Equilib.* **9** (1982) 1–19.

[23] M.L. Michelsen, The isothermal flash problem. Part II. Phase-split calculation. *Fluid Phase Equilib.* **9** (1982) 21–40.

[24] M.L. Michelsen and J.M. Mollerup, Thermodynamic Models: Fundamentals & Computational Aspects. Tie-Line Publications, Holte (2007).

[25] A. Mitsos and P.I. Barton, A dual extremum principle in thermodynamics. *AIChE J.* **53** (2007) 2131–2147.

[26] J. Nocedal and S.J. Wright, Numerical Optimization. *Springer Series in Operations Research and Financial Engineering*. Springer, New York (2006).

[27] H. Orbey and S.I. Sandler, Modeling Vapor-Liquid Equilibria: Cubic Equations of State and Their Mixing Rules. *Cambridge Series in Chemical Engineering*. Cambridge University Press (1998).

[28] D.-Y. Peng and D.B. Robinson, A new two-constant equation of state. *Ind. Eng. Chem. Fundam.* **15** (1976) 59–64.

[29] R.H. Perry and D.W. Green, Perry's Chemical Engineers' Handbook. *McGraw-Hill Chemical Engineering Series*. McGraw-Hill (1999).

[30] N. Peton, Comparaison de plusieurs formulations pour les écoulements multiphasiques et compositionnels en milieu poreux. Tech. report, IFPEN (2015).

[31] N. Peton, C. Cancès, D. Granjeon, Q.-H. Tran and S. Wolf, Numerical scheme for a water flow-driven forward stratigraphic model. *Comput. Geosci.* **24** (2020) 37–60.

[32] H.H. Rachford and J.D. Rice, Procedure for use of electronic digital computers in calculating flash vaporization hydrocarbon equilibrium. *J. Petrol. Technol.* **4** (1952) 19.

[33] J. Vidal, Thermodynamics. Applications in Chemical Engineering and The Petroleum Industry. Institut Français du Pétrole Publications, Technip, Paris (2003).

[34] D.T.S. Vu, *Numerical resolution of algebraic systems with complementarity conditions: application to the thermodynamics of compositional multiphase mixtures*. Ph.D. thesis, Université Paris-Saclay (2020) https://tel.archives-ouvertes.fr/tel-02987892.

[35] D.T.S. Vu, I. Ben Gharbia, M. Haddou and Q.H. Tran, A new approach for solving nonlinear algebraic systems with complementarity conditions: application to compositional multiphase equilibrium problems. *Math. Comput. Simul.* **190** (2021) 1243–1274.

[36] C.H. Whitson and M.L. Michelsen, The negative flash. *Fluid Phase Equilib.* **53** (1989) 51–71.