A new approach for solving nonlinear algebraic systems with complementarity conditions. Application to compositional multiphase equilibrium problems

Duc Thach Son Vu, Ibtihel Ben Gharbia, Mounir Haddou, Quang Huy Tran

A new approach for solving nonlinear algebraic systems with complementarity conditions. Application to compositional multiphase equilibrium problems

Duc Thach Son Vu* Ibtihel Ben Gharbia* Mounir Haddou†
Quang-Huy Tran*

January 28, 2021

Abstract

We present a new method to solve general systems of equations containing complementarity conditions, with a special focus on those arising in the thermodynamics of multicomponent multiphase mixtures at equilibrium. Indeed, the unified formulation introduced by Lauser et al. [*Adv. Water Res.* **34** (2011), 957–966] has recently emerged as a promising way to automatically handle the appearance and disappearance of phases in porous media compositional multiphase flows. From a mathematical viewpoint and after discretization in space and time, this leads to a system consisting of algebraic equations and nonlinear complementarity equations. Such a system exhibit serious convergence difficulties for the existing semismooth and smoothing methods. This observation led us to design a new strategy called NPIPM (*Non-Parametric Interior-Point Method*). Inspired from interior-point methods in optimization, the technique we propose avoids any parameter management while ensuring good theoretical convergence results. These are validated by extensive numerical tests, in which we compare NPIPM to the Newton-min method.

Keywords

complementarity condition, unified formulation, phase equilibrium, Newton's method, interior-point methods

Mathematics subject classification

80M25, 90C33, 90C51

1 Introduction

1.1 Motivation and objectives

In many applied scientific fields such as mechanics, electronics or chemical kinetics [1,16] we often encounter models of the form

$$\Lambda(X) = 0, \qquad \in \mathbb{R}^{\ell-m}, \tag{1.1a}$$

$$\min(G(X),\, H(X)) = 0, \qquad \in \mathbb{R}^{m}, \tag{1.1b}$$

*IFP Energies nouvelles, 1 et 4 avenue de Bois Préau, 92852 Rueil-Malmaison Cedex, France. thachsonnt94@gmail.com, ibtihel.ben-gharbia@ifpen.fr, quang-huy.tran@ifpen.fr

†IRMAR, INSA-Rennes, 20 Avenue des Buttes de Coësmes, 35708 Rennes, France. mounir.haddou@insa-rennes.fr

where the unknown $X \in \mathcal{D}$ is to be found in some open domain $\mathcal{D} \subset \mathbb{R}^\ell$ and where the given functions $\Lambda : \mathcal{D} \subset \mathbb{R}^\ell \to \mathbb{R}^{\ell-m}$ and $G, H : \mathcal{D} \subset \mathbb{R}^\ell \to \mathbb{R}^m$, with $0 < m \leqslant \ell$, are assumed to be continuously differentiable on $\mathcal{D}$. The first $\ell - m$ equations (1.1a) are "ordinary" algebraic equations. By contrast, the last m equations (1.1b) are rather "special" in that they involve the componentwise minimum function and are therefore nondifferentiable. They represent the so-called *complementarity conditions* (or *unilateral conditions*), the exact significance of which is $0 \leqslant G(X) \perp H(X) \geqslant 0$, or equivalently,

$$G(X) \geqslant 0, \qquad H(X) \geqslant 0, \qquad \langle G(X),\, H(X) \rangle = 0. \tag{1.2}$$

This name is justified by the observation that for each index $\alpha \in \{1, \ldots, m\}$, at least one of the two quantities $G_\alpha(X)$ and $H_\alpha(X)$ vanishes while the other remains nonnegative. Each complementarity condition $\min(G_\alpha(X),\, H_\alpha(X)) = 0$ expresses two possible functioning regimes by a same single equation. The m complementarity conditions thus enable us to conveniently envision the 2^m configurations of the physical system in a unified manner. By "unified" we mean that a fixed set of equations and unknowns is assembled throughout the simulation, namely,

$$F(X) = 0, \qquad \text{with} \quad F(X) = \begin{bmatrix} \Lambda(X) \\ \min(G(X), H(X)) \end{bmatrix}. \tag{1.3}$$

Complementarity conditions were introduced in thermodynamics by Lauser et al. [23] for the phase equilibrium problem of compositional multiphase mixtures, with an aim to speed up the simulation of flows in porous media. The difficulty of this problem lies in the appearance and the disappearance of various phases (liquid, gas, oil), the handling of which is quite delicate. Reservoir engineers traditionally use the *variable-switching formulation* [13], in which only those unknowns and equations that correspond to a currently present phase are considered. This has the advantage of keeping the system small-sized, but is cumbersome and costly to implement, insofar as "switching" may occur all the time. In this respect, Lauser et al.'s *unified formulation* is a tremendous progress, not only for the practical comfort it offers but also for the theoretical properties that it encapsulates, as emphasized in [9].

Subsequent works by Ben Gharbia [6], Ben Gharbia and Jaffré [10], Masson et al. [26, 27] and Beaude et al. [4] testified to the superiority of the unified formulation over the variable-switching one regarding computational time in simple cases, i.e., with Henry's law for fugacity coeffients. Another series of works at IFPEN [7, 8, 24, 31] seem to reach the opposite conclusion for realistic fugacity coefficients given by cubic equations of state, such as Peng-Robinson's law. All these works use the *Newton-min* method [2, 22] to solve the nondifferentiable algebraic system of equations (1.1). It was observed that Newton-min may suffer from periodic oscillations or converge to a wrong solution. There are two possible explanations to this lack of robustness:

1. System (1.1) is ill-posed for some data and thermodyanimic laws. It may not have a solution or some components of Λ are not well-defined over the whole domain of interest $\mathcal{D}$. This issue pertains to physical modelling and was addressed in [9].

2. Despite its popularity among numericist, thanks to its simplicity and the semismooth local convergence theory of [29, 32], Newton-min may not be well-suited to the specific nature of our problem. We need another algorithm. This issue is the subject of this article.

This work is a first step in the direction of an alternative numerical method to solve (1.1), with a better guarantee of convergence and a greater robustness with respect to the parameters of the problem. Our quest is deeply rooted in the belief that the unified formulation has a strong potential to improve the performance of compositional multiphase flow simulators. After all, it is already a major advance to have succeeded in recasting the continuous problem under a unified language. It would be a pity not to "convert the try" for want of an adequate numerical method.

1.2 Main results and outline of the paper

First of all, it is essential to review the existing classical methods that could be considered for our problem. This is the purpose of §2. In addition to Newton-min (§2.1), which is part of the semismooth family, we also revisit the smoothing family with the θ-smoothing methods (§2.2) and the interior-point methods (§2.3). Although we prefer this second category for its conceptual elegance, we insist on its main shortcoming, which is the absence of a reliable strategy to drive the regularization parameter towards zero.

The identification of this difficulty prompts us to propose a new interior-point method (§3), in which the smoothing parameter becomes a full-fledged unknown. The latter is governed by a new equation coupling the new variable with the remaining ones. The augmented system is then solved with the ordinary smooth Newton method, with an optional Armijo-type line search to ensure global convergence. Since the parameter is now updated "automatically" and no longer "manually," the new algorithm is given the name *NonParametric Interior-Point Method* (NPIPM). This is our main contribution.

In preparation for the numerical tests, we consider two models giving rise to systems of the form (1.3). Both are reduced versions of more realistic models for two-phase compositional mixtures. The first one, expounded in §4.1, is the stationary phase equilibrium problem at a given global composition. A thorough analysis of it was conducted in [9], where we highlighted its good mathematical properties and clarified the assumptions ensuring the existence of a solution, detailed in §4.2. In §4.3, we prove the regularity of all solutions barring two degenerate cases, which is a favorable feature for quadratic convergence of NPIPM. This model is deployed in combination with Henry's and Peng-Robinson's fugacity laws. Thanks to the simplicity of this model, we were able to diagnose [9] an intrinsic deficiency of Peng-Robinson's law that makes it impossible for F to be well-defined over the whole domain $\mathcal{D}$, thus jeopardizing the unified formulation. The domain extension procedure proposed in [9] and recalled in §4.4 are therefore systemically applied in order to prevent the numerical methods from stopping prematurely.

The second model, which includes an evolution in time, is built on top of the first one by prescribing an ordinary differential equation on the global mass fraction (§5.1). Far from being arbitrary, this differential equation can be heuristically derived from a one-dimensional two-phase binary flow model discretized in space by a one-cell mesh. The interest of this second model lies in the new difficulty associated with the time-step. This is why it is deployed in combination with only Henry's law for fugacities, which also allows the exact solution to be determined (§5.2) and proven to be regular (§5.3).

Finally, extensive numerical results are provided in §6, where NPIPM is systematically compared to Newton-min. For each model, we sweep over the initial points and the parameters. This test campaign clearly demonstrates NPIPM's excellence in comparison with Newton-min: most of the time, NPIMP achieves a 100% convergence ratio.

2 Standard approaches for solving systems with complementarity conditions

For a general function $F : \mathcal{D} \in \mathbb{R}^\ell \to \mathbb{R}^\ell$ that is not differentiable everywhere, the numerical methods available in the literature to solve $F(X) = 0$ can be grossly divided into two categories: semismooth methods and smoothing methods.

2.1 Newton-min, a popular semismooth method

The semismooth approach generalizes the classical smooth Newton method by resorting to a broader notion of Jacobian matrix. Actually, the semismooth Newton theory [29, 32] is a concrete

embodiment of the nonsmooth Newton theory [17, §7.2], an abstract framework that was developed much earlier for locally Lipschitz functions.

By Rademacher's theorem [14, §3.4.1], every locally Lipschitz-continuous function is continuously differentiable almost everywhere. Put another way, the set $\mathcal{C}_F$ of points $X \in \mathcal{D}$ where $\nabla F(X)$ exists in the classical sense is non-empty and its complement $\mathcal{D}\backslash\mathcal{C}_F$ has measure zero. This enables two notions of subdifferential to be introduced.

Definition 2.1 (Bouligand and Clarke subdifferentials)**.** Let $F : \mathcal{D} \subseteq \mathbb{R}^\ell \to \mathbb{R}^\ell$ be a locally Lipschitz-continuous function and $\mathcal{C}_F \subset \mathcal{D}$ be the set of points at which F is differentiable.

1. The *B-subdifferential* or the *limiting Jacobian* of F at X is the set-valued mapping $\partial_B F : \mathcal{D} \rightrightarrows \mathbb{R}^{\ell\times\ell}$ defined as

$$\partial_B F(X) = \big\{M \in \mathbb{R}^{\ell\times\ell} \mid \exists (X^n)_{k\in\mathbb{N}} \subset \mathcal{C}_F,\ X^n \to X,\ \nabla F(X^n) \to M\big\}. \tag{2.1a}$$

   In other words, the Bouligand subdifferential $\partial_B F(X)$ is the set of all matrices M are the limits of the Frechet differentials $\nabla F(X^n)$ for a sequence X^n converging to X.

2. The *C-subdifferential* or the *generalized Jacobian* of F at X is the set-valued mapping $\partial F : \mathcal{D} \rightrightarrows \mathbb{R}^{\ell\times\ell}$ given by

$$\partial F(X) = \operatorname{conv}(\partial_B F(X)). \tag{2.1b}$$

   In other words, the Clarke subdifferential $\partial F(X)$ is the convex hull of the Bouligand subdifferential $\partial_B F(X)$.

The idea is then to prescribe the local linear approximation $d \mapsto F(X^k) + M^k d$ around X^k with $M^k \in \partial F(X^k)$ and to look for a zero of this approximation scheme. This defines the sequence $X^{k+1} = X^k + d^k$ after having solved $M^k d^k = -F(X^k)$ for d^k. Unfortunately, in general this linear local model does not satisfy the technical conditions required by the nonsmooth framework to ensure convergence [33, Definition 4.5]. This is why we have to restrict ourselves to a subclass of those locally Lipschitz functions that comply exactly with these conditions.

Definition 2.2 (Semismooth function)**.** Let $F : \mathcal{D} \subseteq \mathbb{R}^\ell \to \mathbb{R}^\ell$ be a locally Lipschitz-continuous function. We say that F is *semismooth* at $\bar{X} \in \mathcal{D}$ if

$$\limsup_{\substack{X\to\bar{X}\\ M\in\partial F(X)}} \frac{\|F(X) + M(\bar{X} - X) - F(\bar{X})\|}{\|X - \bar{X}\|} = 0. \tag{2.2}$$

The subclass of semismooth functions is rich enough to cover all functions of interest in real applications. It happens, however, that the generic element of $\partial F(X^k)$ is difficult to identify. This is precisely the case for the functions F having the form (1.3). It is then recommended [21] to pick $M^k \in \partial_B F(X^k)$ instead, which might be easier to determine. The following statement provides the generic element of $\partial_B F(X^k)$ when F is of the type (1.3).

Proposition 2.1. *If Λ, G, $H : \mathcal{D} \subset \mathbb{R}^\ell \to \mathbb{R}^\ell$ are continuously differentiable, then F is semismooth. Its B-subdifferential consists of all matrices $M \in \mathbb{R}^{\ell\times\ell}$ of the form*

$$\partial_B F(X) = \left\{ M = \begin{bmatrix} \nabla\Lambda(X) \\ \nabla\!\!\!\nabla \end{bmatrix},\ \nabla\!\!\!\nabla \in \mathbb{R}^{m\times\ell} \right\}, \tag{2.3a}$$

where the α-th row of $\nabla\!\!\!\nabla$ for $\alpha \in \{1, \ldots, m\}$ is

$$\nabla\!\!\!\nabla_\alpha = \begin{cases} \nabla G_\alpha(X) & \text{if } G_\alpha(X) < H_\alpha(X), \\ \nabla G_\alpha(X) \text{ or } \nabla H_\alpha(X) & \text{if } G_\alpha(X) = H_\alpha(X), \\ \nabla H_\alpha(X) & \text{if } G_\alpha(X) > H_\alpha(X). \end{cases} \tag{2.3b}$$

Algorithm 1 Newton-min algorithm

1. Choose $X^0 \in \mathcal{D} \subset \mathbb{R}^\ell$. Set $k = 0$.
2. If $F(X^k) = 0$, stop.
3. Select an element $M^k \in \partial_B F(X^k)$ as in (2.3). Find a direction $d^k \in \mathbb{R}^\ell$ such that
$$F(X^k) + M^k d^k = 0. \tag{2.4}$$
4. Set $X^{k+1} = X^k + d^k$ and $k \leftarrow k + 1$. Go to step 2.

Proof. The proof makes use of general results on the semismoothness of the componentwise minimum of two semismooth functions and on the B-subdifferential of a vector-valuded function [21, §1.75, §1.54]. See details in [33, Proposition 4.2]. □

Algorithm 1 summarizes the successive steps of the Newton-min method for problem (1.3). Historically, the first report on Newton-min seems to date back to Aganagić [2]. Instead of the minimum function, we could have expressed componentwise complementarity by means of another *C-function*, that is, a function $\psi : \mathbb{R}^2 \to \mathbb{R}$ such that
$$\psi(a, b) = 0 \quad \Leftrightarrow \quad a \geqslant 0, \quad b \geqslant 0, \quad ab = 0. \tag{2.5}$$
Besides $\psi_{\min}(a, b) = \min(a, b)$, two other C-functions are worth knowing of:

- the Fischer-Burmeister function $\psi_{\mathrm{FB}}(a, b) = \sqrt{a^2 + b^2} - (a + b)$. This C-function is differentiable everywhere except at $(0, 0)$. In addition, its square $\psi^2_{FB}(a, b)$ is continuously differentiable on the entire plane. Introduced in [18], the Fischer-Burmeister function played a central role in the early development of algorithms. The corresponding semi-smooth method is called *Newton-FB*.
- the Mangasarian function $\psi_{\mathrm{M}}(a, b) = \zeta(|a-b|)-\zeta(a)-\zeta(b)$, where $\zeta : \mathbb{R} \to \mathbb{R}$ is an increasing function and $\zeta(0) = 0$. It can be made differentiable everywhere by an appropriate choice of ζ, e.g., $\zeta(t) = t^3$ as advocated in [25]. The coresponding *Newton-M* method is nevertheless not very successful, due to a flaw shared by all smooth C-functions, that is, $\nabla\psi(0, 0) = (0, 0)$. This makes the Jacobian matrix singular at the solution.

Remark 2.1. It was observed [6] that the direction d^k computed by (2.4) is not always a descent direction for the least-squares merit function $\Theta(X) = \frac{1}{2}\|F(X)\|^2$. Consequently, globalization of Newton-min by means of a line search (cf. §3.3) along d^k remains a delicate issue. □

2.2 θ-smoothing, a typical smoothing method

In contrast to semismooth methods, smoothing methods first try to regularize the F function, which introduces a *regularization parameter* that must be progressively pushed to zero. A regularization of F is the choice of a family of functions
$$\left\{ \widetilde{F}(\cdot; \nu) : \ \mathcal{D} \subset \mathbb{R}^\ell \to \mathbb{R}^\ell, \ \ \nu > 0 \right\} \tag{2.6}$$
such that: (i) $\widetilde{F}(\cdot; \nu)$ is a smooth (continuously differentiable) function of X, for all $\nu > 0$; (ii) $\widetilde{F}(\cdot; \nu)$ is continuous with respect to ν, in some functional sense; (iii) $\lim_{\nu \downarrow 0} \widetilde{F}(\cdot; \nu) = F(\cdot)$, in some functional sense. Starting from a current pair of values (X^k, ν^k), the overall strategy of a smoothing method is to

1. solve $\widetilde{F}(X^{k+1}; \nu^k) = 0$ in the unknown X^{k+1} by means of the smooth Newton method, using X^k as the initial point;

2. decrease the regularization parameter from ν^k to ν^{k+1} by some "rule of thumb." Start over the process until $F(X^{k+1}) = 0$.

The regularized system in Step 1 is supposedly easier to solve and may benefit from the enhanced properties of the smooth Newton method. Moreover, it is customary to do just one iteration in Step 1, with a view to lower the computational cost. This paradigm, known as the *diagonal Newton*, induces more approximation error but is of great practical interest.

To smooth out the scalar complementarity equation $0 \leqslant a \perp b \geqslant 0$, let us first rewrite it under the plainly equivalent form

$$a \geqslant 0, \qquad b \geqslant 0, \qquad \mathfrak{S}(a) + \mathfrak{S}(b) \leqslant 1, \tag{2.7}$$

where

$$\mathfrak{S}(t) = \begin{cases} 0 & \text{if } \ t = 0, \\ 1 & \text{if } \ t > 0 \end{cases} \tag{2.8}$$

is the *step function*. The latter serves as an indicator of positive arguments $t > 0$ over $\mathbb{R}_+$. Since the step function is discontinuous, let us regularize it using the notion of a *smoothing function* introduced by Haddou and his coauthors [3, 19].

Definition 2.3 (θ-smoothing function)**.** A function $\theta : \mathbb{R}_+ \to [0, 1)$ is said to be a *θ-smoothing function* if it is continuous, nondecreasing, concave, and

$$\theta(0) = 0, \qquad \lim_{t \to +\infty} \theta(t) = 1. \tag{2.9a}$$

Furthermore, if θ can be defined for negative arguments $t \in (-T, 0)$, with $T > 0$, while remaining continuous, nondecreasing and concave, it is required that

$$\theta(t) < 0 \quad \text{ for } \ t \in (-T, 0). \tag{2.9b}$$

The smoothing function is a "father" function, by compression of which regularized step functions will be generated. The two most common examples of smoothing functions are:

- the rational function $\theta^1 : (-1, +\infty) \to (-\infty, 1)$ defined by

$$\theta^1(t) = \frac{t}{t+1}. \tag{2.10a}$$

- the exponential function $\theta^2 : \mathbb{R} \to (-\infty, 1)$ defined by

$$\theta^2(t) = 1 - \exp(-t). \tag{2.10b}$$

Definition 2.4 (θ-smoothing family)**.** Let θ be a θ-smoothing function. The family of functions

$$\big\{ \theta_\nu(t) := \theta(t/\nu), \ \ \nu > 0 \big\} \tag{2.11}$$

is said to be the θ-smoothing family associated with θ.

Obviously, θ_ν is a smooth function of $t \geqslant 0$ for all $\nu > 0$. It is also continuous with respect to ν at each fixed $t \geqslant 0$. From the defining properties (2.9), it can be readily shown that

$$\lim_{\nu \downarrow 0} \theta_\nu(t) = \mathfrak{S}(t), \quad \forall t \geqslant 0. \tag{2.12}$$

In other words, $\mathfrak{S}$ is the limit of θ_ν in the sense of pointwise convergence. The equivalence between $0 \leqslant a \perp b \geqslant 0$ and (2.7) suggests us to impose

$$a \geqslant 0, \qquad b \geqslant 0, \qquad \theta_\nu(a) + \theta_\nu(b) = 1 \tag{2.13}$$

for $\nu > 0$, as a smooth approximation of (2.7). Replacing $\mathfrak{S}$ by θ_ν in (2.7) is logical. Replacing "$\leqslant$" by "$=$" in (2.7) is motivated by the fact that we want an equality to be mounted into the system of equations. Let us examine the impact of (2.13) on the examples (2.10).

- For the rational function (2.10a), we can easily prove the remarkable equivalence

$$\theta^1_\nu(a) + \theta^1_\nu(b) = 1 \;\Leftrightarrow\; ab = \nu^2. \tag{2.14a}$$

  The equality $ab = \nu^2$ appears to be a natural relaxation of $ab = 0$. This smoothing paradigm will prevail in interior-point methods, with ν in the right-hand side instead of ν^2.

- For the exponential function (2.10b), we readily check the equivalence

$$\theta^2_\nu(a) + \theta^2_\nu(b) = 1 \;\Leftrightarrow\; -\nu\ln[\exp(-a/\nu) + \exp(-b/\nu)] = 0, \tag{2.14b}$$

  In the left-hand side, the function $\min_\nu(a,b) := -\nu\ln[\exp(-a/\nu) + \exp(-b/\nu)]$ can be regarded as a smooth approximation of $\min(a,b)$. This relaxation of $\min(a,b) = 0$ could have been worked out from the more well-known fact [5] that $\max_\nu(a,b) := \nu\ln[\exp(a/\nu) + \exp(b/\nu)]$ is a smooth approximation of $\max(a,b)$.

For F of the form (1.3), we consider the regularization family $\{\widetilde{F}(\cdot,\nu),\ \nu > 0\}$, where

$$\widetilde{F}(X\,,\nu) = \begin{bmatrix} \Lambda(X) \\ \nu\,(\theta_\nu(G(X)) + \theta_\nu(H(X)) - \mathbf{1}) \end{bmatrix}. \tag{2.15}$$

Here, it is understood that θ_ν operates componentwise on $G(X)$ and $H(X)$, while $\mathbf{1} \in \mathbb{R}^m$ is the vector whose entries are all equal to 1. The premultiplication by ν of the second line in (2.15) is aimed at controlling the magnitude of their partial derivatives. Indeed, for all $t \geqslant 0$, $\theta'_\nu(t) = \nu^{-1}\theta'(t/\nu)$ can be seen to blow up when $\nu \downarrow 0$, while $\nu\theta'_\nu(t)$ tends to the finite limit $\theta'(0)$.

To complete the two-step procedure described at the beginning of §2.2, we also need some "rule of thumb" to steer ν to 0. This issue will be discussed at the end of §2.3.

2.3 Interior-point, a reference in optimization

Renowned for their efficiency in linear programming thanks to their polynomial complexity, interior-point methods [34] can be interpreted as regularization methods. We are mostly interested in *primal-dual* methods [35], in which primal variables (initial unknowns) and duals (Lagrange multipliers) enjoy the same status. When one dissects a primal-dual inner-point method, one realizes that it is basically a method for solving the algebraic system of Karush-Kuhn-Tucker (KKT) optimality conditions. The fact that the system comes from a constrained minimization problem is ultimately of little importance in the method. This opens up the prospect of transposing these methods to the case of a general system containing complementarity conditions.

Let us consider the family of regularized problems

$$\widetilde{F}(X;\nu) = 0, \qquad \text{with} \quad \widetilde{F}(X;\nu) = \begin{bmatrix} \Lambda(X) \\ G(X) \odot H(X) - \nu\mathbf{1} \end{bmatrix} \in \mathbb{R}^\ell, \tag{2.16}$$

where $\nu \geqslant 0$ is the smoothing parameter, $\mathbf{1} \in \mathbb{R}^m$ is the vector whose components are all equal to 1, and $\odot$ denotes Hadamard's componentwise product. It is usually more convenient to cast the previous sytem under the form

$$\mathbf{F}(\mathbf{X}\,;\nu) = 0, \qquad \text{with} \quad \mathbf{F}(\mathbf{X}\,;\nu) = \begin{bmatrix} \Lambda(X) \\ G(X) - V \\ H(X) - W \\ V \odot W - \nu\mathbf{1} \end{bmatrix} \in \mathbb{R}^{\ell+2m}, \tag{2.17}$$

where $\mathbf{X} = [X^T; V^T; W^T]^T \in \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^m \subset \mathbb{R}^{\ell+2m}$ is the augmented unknown and where the *slack variables* $(V, W) \in \mathbb{R}^m \times \mathbb{R}^m$ are subject to the componentwise positivity conditions

$$V \geqslant 0, \qquad W \geqslant 0. \tag{2.18}$$

Enlarging the size of the system and the number of unknowns does not change the determinant of the Jacobian matrix at the corresponding solution. Let us state this result for later use. Due to defintions (2.16) and (2.17), the Jacobian matrices $\nabla_X \widetilde{F}(X;\nu)$ and $\nabla_{\mathbf{X}}\mathbf{F}(\mathbf{X};\nu)$ do not depend on ν. For short, they will be denoted by $\nabla\widetilde{F}(X)$ and $\nabla\mathbf{F}(\mathbf{X})$.

Lemma 2.1. *Let $X \in \mathcal{D}$ and $\mathbf{X} = [X^T; V^T; W^T]^T \in \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^m$ such that $V = G(X)$ and $W = H(X)$. Then,*

$$\det \nabla\mathbf{F}(\mathbf{X}) = \det \nabla\widetilde{F}(X). \tag{2.19}$$

Proof. The determinant of the Jacobian matrix of $\mathbf{F}(\mathbf{X};\nu)$ is equal to

$$\det \nabla\mathbf{F}(\mathbf{X}) = \begin{vmatrix} \nabla\Lambda(X) & 0 & 0 \\ \nabla G(X) & -I_m & 0 \\ \nabla H(X) & 0 & -I_m \\ 0 & I_m \odot W & I_m \odot V \end{vmatrix},$$

where the Hadamard product between a matrix and a vector is defined as the matrix whose each column is the Hadamard product between the corresponding column of the matrix and the vector. By linear combination of the last (block)-row with the second and third (block)-rows, we obtain

$$\det \nabla\mathbf{F}(\mathbf{X}) = \begin{vmatrix} \nabla\Lambda(X) & 0 & 0 \\ \nabla G(X) & -I_m & 0 \\ \nabla H(X) & 0 & -I_m \\ \nabla G(X) \odot W + \nabla H(X) \odot V & 0 & 0 \end{vmatrix}.$$

By means of $2m$ row permutations, we end up with

$$\det \nabla\mathbf{F}(\mathbf{X}) = \begin{vmatrix} \nabla\Lambda(X) & 0 & 0 \\ \nabla G(X) \odot W + \nabla H(X) \odot V & 0 & 0 \\ \nabla G(X) & -I_m & 0 \\ \nabla H(X) & 0 & -I_m \end{vmatrix} = \begin{vmatrix} \nabla\Lambda(X) \\ \nabla G(X) \odot W + \nabla H(X) \odot V \end{vmatrix}.$$

For $V = G(X)$, $W = H(X)$, the last determinant coincides $\det \widetilde{F}(\cdot;\nu)$. Note that the Lemma does not require X to be a solution of (2.16). □

A primal-dual interior-point method strives to generate a sequence

$$\{\mathbf{X}^k\}_{k\in\mathbb{N}} \subset \mathfrak{I} := \big\{\mathbf{X} = (X,\, V,\, W) \in \mathbb{R}^{\ell+2m} \;\mid\; V > 0,\;\; W > 0\big\}, \tag{2.20}$$

as well as an auxiliary sequence $\{\nu^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^*_+$ such that $(X^k,\, V^k,\, W^k) \to (\bar{X},\, G(\bar{X}),\, H(\bar{X}))$ and $\nu^k \to 0$, where $\bar{X} \in \mathcal{D}$ is a zero of $F = \widetilde{F}(\cdot;0)$. Algorithm 2 describes a *single-stage* interior-point method which consists of one Newton iteration (Step 3), followed by a truncation (Step 4) and an update for the regularization parameter (Step 6).

Below are a few common empirical ways to progressively drive ν^k to 0:

$$\nu^{k+1} = (\nu^k)^2; \tag{2.23a}$$

$$\nu^{k+1} = (\nu^k)^2; \tag{2.23b}$$

$$\nu^{k+1} = \min\big(0.5\,\nu^k,\, (\nu^k)^2\big); \tag{2.23c}$$

$$\nu^{k+1} = \min\big(0.5\,\nu^k,\; (\nu^k)^2,\, \langle V^{k+1},\, W^{k+1}\rangle / m\big). \tag{2.23d}$$

The geometric sequence (2.23a) has the advantage of going slowly to zero, which is recommended when ν^k is still large. The power sequence (2.23b) goes quickly to zero, which is relevant when ν^k is already small. The hybrid geometric-power sequence (2.23c) combines the advantages of the first two strategies. The hybrid geometric-power sequence involving a duality measure (2.23d)

Algorithm 2 Single-stage interior-point algorithm

1. Choose $\mathbf{X}^0 \in \mathfrak{I}$. Set $k = 0$, $\nu^0 = \langle V^0, W^0 \rangle / m$, $\gamma = 0.99$.
2. If $\mathbf{F}(\mathbf{X}^k ; 0) = 0$, stop.
3. Find a direction $\mathbf{d}^k = (\mathrm{d}X^k, \mathrm{d}V^k, \mathrm{d}W^k) \in \mathbb{R}^{\ell+2m}$ such that
$$\mathbf{F}(\mathbf{X}^k ; \nu^k) + \nabla \mathbf{F}(\mathbf{X}^k)\, \mathbf{d}^k = 0. \tag{2.21}$$
4. Compute $\varsigma^k \in (0,1)$ such that $\mathbf{X}^k + \varsigma^k \mathbf{d}^k \in \mathfrak{I}$ by
$$\varsigma^k = \gamma \cdot \arg\max \{ \varsigma \in [0,1] \mid V^k + \varsigma\, \mathrm{d}V^k \geqslant 0,\ W^k + \varsigma\, \mathrm{d}W^k \geqslant 0 \}. \tag{2.22}$$
5. Set $\mathbf{X}^{k+1} = \mathbf{X}^k + \varsigma^k \mathbf{d}^k$.
6. Set $\nu^{k+1} =$ one of the heuristic strategies (2.23).
7. Set $k \leftarrow k+1$. Go to step 2.

allows the sequence to be reconnected to some current "reality" [20, §5]. Unfortunately, there is no universal magic formula to monitor the sequence of regularization parameter $\{\nu^k\}$. A heuristic strategy that works fine with one problem may fail with another. We have to try several sequences $\{\nu^k\}$ before knowing which one is best suited to the problem at hand.

Remark 2.2. Most of today's interior-point general-purpose softwares for linear programming are based on Mehrotra's predictor-corrector algorithm [28]. This celebrated two-stage algorithm rests upon a subtle tuning of the regularization parameter. It can be similarly extended to an equation solver [33, §4.3.3.3], but the numerical results are not always as good as the other methods [33, §6.1.1]. □

3 Design of a new nonparametric interior-point method

To compensate for the weakness of smoothing methods regarding the lack of a good strategy to decrease ν to 0, we undertake to look for another way of dealing with the regularization parameter. Our guiding principle is to treat it as an unknown in its own right.

3.1 NPIPM in a nutshell

In system (2.17), the status of the parameter ν is very distinct from that of the variable $\mathbf{X}$. While $\mathbf{X}$ is computed by a Newton iteration, ν has to be updated in an ad hoc manner. Usually, progress occurs when two objects of different natures are put on an equal footing. This is why we feel that it would be judicious to incorporate the parameter ν into the variables $\mathbf{X}$.

Let us consider the enlarged vector $\mathbb{X} = [\mathbf{X}^T ; \nu]^T \in \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}_+ \subset \mathbb{R}^{\ell+2m+1}$. We seek a system of $\ell + 2m + 1$ equations
$$\mathbb{F}(\mathbb{X}) = 0, \qquad \text{with} \quad \mathbb{F}(\mathbb{X}) = \begin{bmatrix} \mathbf{F}(\mathbf{X} ; \nu) \\ f(\mathbf{X}, \nu) \end{bmatrix} \tag{3.1}$$
to be prescribed on $\mathbb{X}$. The last equation $f(\mathbf{X}, \nu) = 0$ must be devised in such a way that every solution of (2.17) with $\nu = 0$ is also a solution of the enlarged system (3.1). To enforce a few "desirable" properties that will be enumerated later, we advocate
$$f(\mathbf{X}, \nu) = \frac{1}{2}\|V^-\|^2 + \frac{1}{2}\|W^-\|^2 + \frac{u}{2m^2}(\langle V, W \rangle^+)^2 + \eta\nu + \nu^2, \tag{3.2a}$$

where u and η are two positive parameters and

$$\|V^-\|^2 = \sum_{\alpha=1}^{m} (V_\alpha^-)^2, \qquad \|W^-\|^2 = \sum_{\alpha=1}^{m} (W_\alpha^-)^2. \tag{3.2b}$$

To explain the rationale behind the choice (3.2), let us remind ourselves that our ultimate goal is to make ν equal 0, while ensuring the inequalities (2.18). Thus, it seems really natural to first consider $f(\mathbf{X}, \nu) = \nu$. This construction turns out to be too naive. Indeed, if we start from some $\nu^0 > 0$ and solve the smooth system (3.1) by the smooth Newton method, since the last equation is linear, we end up with $\nu^1 = 0$ at the first iteration. Once the boundary of the interior region is reached, we are "stuck" there.

To prevent ν from rushing to zero in just one iteration, we could set $f(\mathbf{X}, \nu) = \nu^2$, which is equivalent at the continuous level. At the level of Newton iterates, there is still a flaw: since $\nu = 0$ is now a double root of the last equation, quadratic convergence will be lost when ν^k approaches 0! A remedy to this is to add a small linear term, that is, $f(\mathbf{X}, \nu) = \eta\nu + \nu^2$, where $\eta > 0$ is a small parameter. The price to be paid for recovering quadratic convergence is that there is now a spurious negative solution $\nu = -\eta < 0$. This should not be a problem, however, if we start from a positive value for ν.

At this stage, system (3.1) is still not yet adequate. Indeed, the last equation is totally decoupled from the others. Everything happens as if ν is governed by a pre-determined sequence, generated a priori by the Newton iterates of the scalar equation $\eta\nu + \nu^2 = 0$, regardless of $\mathbf{X}$. It is desirable to couple ν and $\mathbf{X}$ in a tighter way. The choice (3.2) does achieve this purpose:

- As long as $\nu \geqslant 0$, the cancellation of $f(\mathbf{X}, \nu)$ implies $V^- = W^- = 0$, which amounts to saying that $V \geqslant 0$ and $W \geqslant 0$. Should a component of V or W become negative during the iteration, the penalty terms $\frac{1}{2}\|V^-\|^2$ and $\frac{1}{2}\|W^-\|^2$ are strong incentives for it to return into the interior domain $\mathfrak{I}$.
- Thanks to $u(\langle V, W\rangle^+)^2/2m^2$, where $u > 0$ is a small parameter, the update of ν depends on V and W. Without this term, we run again into the previous problem of ν "living its own separate life" inside $\mathfrak{I}$. Note that at the continuous level, we have $\langle V, W\rangle/m = \nu$ as a consequence of $V \odot W = \nu\mathbf{1}$, but at the level of iterates, things can be very different.

The idea is now to apply the standard Newton method to the smooth system (3.1)–(3.2), which updates $\mathbf{X}$ and ν simultaneously. To enforce a globally convergent behavior, we also opt for Armijo's line search.

3.2 Determinant of the Jacobian matrix and parameter increment

Before writing down the new algorithm, let us mention some insightful properties regarding the determinant of the Jacobian matrix and the parameter increment when the current iterate lies in the interior region.

Lemma 3.1. *Let $\mathbf{X} \in \overline{\mathfrak{I}}$, where $\mathfrak{I}$ is the interior region defined in (2.20). Let $\mathbb{X} = [\mathbf{X}^T; \nu]^T$ for some $\nu \in \mathbb{R}$. Then,*

$$\det \nabla \mathbb{F}(\mathbb{X}) = (\eta + 2\nu + u\langle V, W\rangle/m) \det \nabla \mathbf{F}(\mathbf{X}). \tag{3.3}$$

If $\nu > -\eta/2$, the two Jacobian matrices are singular or nonsigular at the same time.

Proof. In the same fashion as in the proof of Lemma 2.1, the Jacobian matrix of the enlarged system (3.1)–(3.2) can be decomposed blockwise as

$$\nabla \mathbb{F}(\mathbb{X}) = \begin{bmatrix} \nabla\Lambda(X) & 0 & 0 & 0 \\ \nabla G(X) & -I_m & 0 & 0 \\ \nabla H(X) & 0 & -I_m & 0 \\ 0 & I_m \odot W & I_m \odot V & -\mathbf{1} \\ 0 & (V^-)^T + u\langle V, W\rangle^+ W^T/m^2 & (W^-)^T + u\langle V, W\rangle^+ V^T/m^2 & \eta + 2\nu \end{bmatrix} \tag{3.4}$$

where V^- is the vector of components $V_\alpha^- = \min(V_\alpha, 0)$ and similarly for W^-. If $\mathbf{X} \in \overline{\mathfrak{I}}$, then $V^- = W^- = 0$. Subtracting the product of $u\langle V, W\rangle^+ \mathbf{1}^T/m^2$ with the fourth (block)-row from the last row, we can eliminate the scalar products $\langle V, W\rangle \geqslant 0$ in the second and third columns, so that

$$\det \nabla \mathbb{F}(\mathbb{X}) = \begin{vmatrix} \nabla\Lambda(X) & 0 & 0 & 0 \\ \nabla G(X) & -I_m & 0 & 0 \\ \nabla H(X) & 0 & -I_m & 0 \\ 0 & I_m \odot W & I_m \odot V & -\mathbf{1} \\ 0 & 0 & 0 & \eta + 2\nu + u\langle V, W\rangle/m \end{vmatrix}. \tag{3.5}$$

Expanding (3.5) with respect to the last row yields the desired result. Note that this Lemma does not require $\mathbb{X}$ to solve (3.1)–(3.2). □

The interest of $u(\langle V, W\rangle^+)^2/2m$ in (3.2) can also be revealed by inspecting the increment $\mathrm{d}\nu^k = \nu^{k+1} - \nu^k$ of the Newton method (without truncation nor line search). The following statement shows that whenever ν^k is lower than the central value $\langle V^k, W^k\rangle/m$, either ν^k increases or ν^k decreases with a smaller magnitude than it would have done without $u(\langle V, W\rangle^+)^2/2m$. Put another way, the parameter ν^k cannot go too fast to zero without "waiting" for $\langle V^k, W^k\rangle/m$.

Proposition 3.1. *For $\mathbf{X}^k \in \mathfrak{I}$, the Newton increment for the parameter is*

$$\mathrm{d}\nu^k = -\frac{\eta\nu^k + (\nu^k)^2 - u(\langle V^k, W^k\rangle/2m - \nu^k)\langle V^k, W^k\rangle/m}{\eta + 2\nu^k + u\langle V^k, W^k\rangle/m}. \tag{3.6}$$

In comparison with the value

$$\mathrm{d}\nu_0^k = -\frac{\eta\nu^k + (\nu^k)^2}{\eta + 2\nu^k} \tag{3.7}$$

that corresponds to $u = 0$, we have either $\mathrm{d}\nu_0^k \leqslant 0 < \mathrm{d}\nu^k < |\mathrm{d}\nu_0^k|$ or $\mathrm{d}\nu_0^k < \mathrm{d}\nu^k \leqslant 0$ as soon as $\nu^k < \langle V^k, W^k\rangle/m$.

Proof. Applying the row transformations described the proof of Lemma 3.1 to the linear system $\nabla\mathbb{F}(\mathbb{X}^k)\mathrm{d}^k = -\mathbb{F}(\mathbb{X}^k)$ and taking into account the right-hand side, we end up with

$$\mathrm{d}^k = \begin{bmatrix} \mathrm{d}\mathbf{X}^k \\ \mathrm{d}\nu^k \end{bmatrix} = -\begin{bmatrix} \nabla\mathbf{F}(\mathbf{X}^k) & -\partial_\nu\mathbf{F} \\ 0 & \eta + 2\nu^k + u\langle V^k, W^k\rangle/m \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{F}(\mathbf{X}^k; \nu^k) \\ \eta\nu^k + (\nu^k)^2 - a^k \end{bmatrix},$$

where $a^k = u(\langle V^k, W^k\rangle/2m - \nu^k)\langle V^k, W^k\rangle/m$, hence (3.6). When $u = 0$, the increment (3.6) degenerates to (3.7). Let us introduce

$$A^k = \eta\nu^k + (\nu^k)^2, \qquad a^k = u(\langle V^k, W^k\rangle/2m - \nu^k)\langle V^k, W^k\rangle/m, \tag{3.8a}$$

$$B^k = \eta + 2\nu^k, \qquad b^k = u\langle V^k, W^k\rangle/m, \tag{3.8b}$$

so that $\mathrm{d}\nu^k = -(A^k - a^k)/(B^k + b^k)$ and $\mathrm{d}\nu_0^k = -A^k/B^k$. Then, since $b^k > 0$ and $B^k > 0$, the inequality $|\mathrm{d}\nu^k| < |\mathrm{d}\nu_0^k|$ occurs if and only if $(A^k - a^k)B^k < A^k(B^k + b^k)$, which boils down to

$$\frac{a^k}{b^k} > -\frac{A^k}{B^k}. \tag{3.9}$$

This holds true, e.g., for $a^k \geqslant 0$, that is, $\nu^k \leqslant \langle V^k, W^k\rangle/2m$. But this also holds true when a^k is negative but "not too much." The exact condition (3.9) reads

$$\langle V^k, W^k\rangle/2m - \nu^k > -\frac{\nu^k(\eta + \nu^k)}{\eta + 2\nu^k} = \frac{(\nu^k)^2}{\eta + 2\nu^k} - \nu^k$$

and reduces to $\langle V^k, W^k\rangle/m > 2(\nu^k)^2/(\eta + 2\nu^k)$. Since $2(\nu^k)^2/(\eta + 2\nu^k) < \nu^k$, the condition $\nu^k < \langle V^k, W^k\rangle/m$ becomes sufficient for $|\mathrm{d}\nu^k| < |\mathrm{d}\nu_0^k|$ to take place. □

3.3 Globalized algorithm

Introduce the least-squares potential

$$\Theta(\mathbb{X}) = \frac{1}{2}\|\mathbb{F}(\mathbb{X})\|^2.$$

A detailed description of NPIPM is given in Algorithm 3. The initial point $\mathbb{X}^0 = (\mathbf{X}^0, \nu^0)$ must be an interior point, i.e., $\mathbf{X}^0 \in \mathfrak{I}$. Furthermore, it is often taken at equilibrium, that is, $V^0 = G(X^0)$ and $W^0 = H(X^0)$, so that the initial parameter $\nu^0 = \langle V^0, W^0\rangle/m$ has the correct order of magnitude.

Algorithm 3 Nonparametric interior point algorithm with Armijo line search

1. Choose $\mathbb{X}^0 = (\mathbf{X}^0, \nu^0)$, $\mathbf{X}^0 \in \mathfrak{I}$, $\nu^0 = \langle V^0, W^0\rangle/m$, $\kappa \in (0, 1/2)$, $\varrho \in (0, 1)$. Set $k = 0$.
2. If $\mathbb{F}(\mathbb{X}^k) = 0$, stop.
3. Find a direction $\mathbb{d}^k \in \mathbb{R}^{\ell+2m+1}$ such that

$$\mathbb{F}(\mathbb{X}^k) + \nabla\mathbb{F}(\mathbb{X}^k)\mathbb{d}^k = 0. \tag{3.10}$$

4. Choose $\varsigma^k = \varrho^{j_k} \in (0, 1)$, where $j_k \in \mathbb{N}$ is the smallest integer such that

$$\Theta(\mathbb{X}^k + \varrho^{j_k}\mathbb{d}^k) \leqslant (1 - 2\kappa\varrho^{j_k})\,\Theta(\mathbb{X}^k). \tag{3.11}$$

5. Set $\mathbb{X}^{k+1} = \mathbb{X}^k + \varsigma^k\mathbb{d}^k$ and $k \leftarrow k + 1$. Go to step 2.

Remark 3.1. There is no need to truncate the Newton direction $\mathbb{d}^k$ to preserve positivity for V^{k+1} and W^{k+1}, since nonnegativity is guaranteed at convergence. However, the possibility for the iterates to get out of the interior region makes this method not strictly "interior-point." A truly interior-point variant can be cooked up by carrying out the damping (2.22) before Step 4.

Remark 3.2. The positive parameters η and u are chosen once and for all. They not need to be dynamically adapted in some "smart" way during the iterations. It is in this sense that the adjective *nonparametric* is to be understood.

The convergence of Algorithm 3 is governed by the general theory for the smooth Newton method [11, §6]. This global convergence theory heavily relies on the regularity of zeros.

Definition 3.1 (Regular zero)**.** Let $\bar{X} \in \mathcal{D} \subset \mathbb{R}^\ell$ be a zero of F, that is, $F(\bar{X}) = 0$. If the Jacobian matrix $\nabla F(\bar{X})$ is nonsingular, $\bar{X}$ is said to be a *regular zero* of F.

The three items of the upcoming Theorem illustrate the conditions and the qualities of convergence of the algorithm. Item (i) corresponds to the behavior of the algorithm near a regular zero. Item (ii) states the rate of convergence in some particular situations. Item (iii) summarizes all of the possible scenarios when running the algorithm.

Theorem 3.1. *Let $\mathbb{F} : \mathbb{R}^{\ell+2m+1} \to \mathbb{R}^{\ell+2m+1}$ be a continuously-differentiable function.*

(i) [Local analysis] *Let $\bar{\mathbb{X}}$ be a regular zero of $\mathbb{F}$. If $\mathbb{X}^0$ is close enough to $\bar{\mathbb{X}}$, then $\varsigma_k = 1$ for all k, and $\mathbb{X}^k \to \bar{\mathbb{X}}$ superlinearly (and we recover the standard Newton method).*

(ii) [Limit point] *Let $\tilde{\mathbb{X}}$ be a limit point of sequence $\{\mathbb{X}^k\}$. If $\nabla\mathbb{F}(\tilde{\mathbb{X}})$ is invertible, then $\tilde{\mathbb{X}}$ is a regular zero of $\mathbb{F}$. If $\tilde{\mathbb{X}}$ is a regular zero of F, then $\varsigma_k = 1$ for k big enough and $\mathbb{X}^k \to \tilde{\mathbb{X}}$ superlinearly.*

(iii) [General behavior] *At least one of three possibilities below holds:*

(a) $\mathbb{F}(\mathbb{X}^k) \to 0$.

(b) $\|d(\mathbb{X}^k)\|$ *is unbounded.*

(c) *The sequence* $\{\mathbb{X}^k\}$ *converges to* $\widetilde{\mathbb{X}}$ *where* $\nabla\mathbb{F}(\widetilde{\mathbb{X}})$ *is not invertible.*

Proof. See [11, Theorem 6.9] or the condensed exposition of [33, Theorem 5.2]. □

4 Stationary model for compositional two-phase mixtures

We are going to compare Newton-min and NPIPM on two models involving complementarity conditions for multicomponent two-phase mixtures. Designed as reduced versions of those commonly encountered in reservoir simulations, they enable us to gain valuable insights into what happens to the algorithms and to conduct a thorough numerical study with respect to a full range of parameters (§6). The first model is the unified formulation for the stationary phase equilibrium problem investigated in [33, §2] and [9].

4.1 Phase equilibria in the unified formulation

We consider a mixture consisting of $\mathrm{K} \geqslant 2$ distinct *components* or *species*, labeled by the elements of the set $\mathcal{K} = \{\mathrm{I}, \mathrm{II}, \ldots, \mathrm{K}\}$. Each component $i \in \mathcal{K}$ may be present under at least one phase but at most two phases. To fix ideas, the phases are labeled by the elements of $\mathscr{P} = \{G, L\}$, which respectively stand for Gas and Liquid. Let $\{c^i\}_{i\in\mathcal{K}}$ be nonnegative numbers such that $\sum_{i\in\mathcal{K}} c^i = 1$. Each c^i represents the global fraction of species i. The vector $\boldsymbol{c} = (c^{\mathrm{I}}, \ldots, c^{\mathrm{K}-1}) \in \Omega$, called *global composition*, lies in

$$\Omega = \big\{\boldsymbol{x} = (x^{\mathrm{I}}, \ldots, x^{\mathrm{K}-1}) \in \mathbb{R}^{\mathrm{K}-1} \mid x^{\mathrm{I}} > 0,\ \ldots,\ x^{\mathrm{K}-1} > 0,\ 1 - x^{\mathrm{I}} - \ldots - x^{\mathrm{K}-1} > 0\big\}. \tag{4.1}$$

For any vector $\boldsymbol{x} = (x^{\mathrm{I}}, \ldots, x^{\mathrm{K}-1}) \in \bar{\Omega}$, it is understood that $x^{\mathrm{K}} = 1 - x^{\mathrm{I}} - \ldots - x^{\mathrm{K}-1} \in [0, 1]$ is the fraction corresponding to the last component.

Each phase $\alpha \in \mathscr{P}$ is characterized by a 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

$$\mu_\alpha^j(\boldsymbol{x}) = g_\alpha(\boldsymbol{x}) + \langle \nabla g_\alpha(\boldsymbol{x}),\, \boldsymbol{\delta}^j - \boldsymbol{x}\rangle \tag{4.2}$$

for $\boldsymbol{x} \in \Omega$, where the vector $\boldsymbol{\delta}^j = (\delta_{j,1},\, \delta_{j,2},\, \ldots,\, \delta_{j,\mathrm{K}-1}) \in \mathbb{R}^{\mathrm{K}-1}$ is made up of Kronecker's symbols. The following statement summarizes some helpful identities between g_α and μ_α^i.

Lemma 4.1 (Connection between Gibbs energy and chemical potentials)**.** *For all* $\boldsymbol{x} \in \Omega$*:*

$$g_\alpha(\boldsymbol{x}) = \sum_{j=\mathrm{I}}^{\mathrm{K}} x^j\, \mu_\alpha^j(\boldsymbol{x}); \tag{4.3a}$$

$$\frac{\partial g_\alpha}{\partial x^j}(\boldsymbol{x}) = \mu_\alpha^j(\boldsymbol{x}) - \mu_\alpha^{\mathrm{K}}(\boldsymbol{x}), \qquad \forall j \in \mathcal{K}\backslash\{\mathrm{K}\}. \tag{4.3b}$$

$$0 = \sum_{i=\mathrm{I}}^{\mathrm{K}} x_\alpha^i\, \nabla_{\boldsymbol{x}_\alpha} \mu_\alpha^i(\boldsymbol{x}_\alpha). \tag{4.3c}$$

Proof. See [33, Lemma 2.1]. □

The first equation (4.3a) relates to Gibbs energy to the potentials. The second formula (4.3b) provides the gradient of the Gibbs energy from the potentials. The last identity (4.3c) is known as the *Gibbs-Duhem condition*. The potentials usually take the form

$$\mu_\alpha^i(\boldsymbol{x}) = \ln(x^i \Phi_\alpha^i(\boldsymbol{x})), \tag{4.4}$$

in which Φ_α^i is called the *fugacity coefficient* (or *activity coefficients*) of component i in phase α. Substituting the form (4.4) into (4.3a), we obtain

$$g_\alpha(\boldsymbol{x}) = \sum_{i=\mathrm{I}}^{\mathrm{K}} x^i \ln x^i + \Psi_\alpha(\boldsymbol{x}), \tag{4.5}$$

where the first sum $\sum_{j=\mathrm{I}}^{\mathrm{K}} x^j \ln x^j$ is the *ideal* part and

$$\Psi_\alpha(\boldsymbol{x}) = \sum_{i=\mathrm{I}}^{\mathrm{K}} x^i \ln \Phi_\alpha^i(\boldsymbol{x}), \tag{4.6}$$

is the *excess* part. The relations between Ψ_α and $\ln \Phi_\alpha^i$ are similar to those between g_α and μ_α^i.

Lemma 4.2 (Connection between excess energy and fugacity coefficients)**.** *For all* $\boldsymbol{x} \in \Omega$*:*

$$\ln \Phi_\alpha^j(\boldsymbol{x}) = \Psi_\alpha(\boldsymbol{x}) + \langle \nabla \Psi_\alpha(\boldsymbol{x}),\, \boldsymbol{\delta}^j - \boldsymbol{x} \rangle, \qquad \forall j \in \mathcal{K}; \tag{4.7a}$$

$$\frac{\partial \Psi_\alpha}{\partial x_\alpha^j}(\boldsymbol{x}) = \ln \Phi_\alpha^j(\boldsymbol{x}) - \ln \Phi_\alpha^{\mathrm{K}}(\boldsymbol{x}), \qquad \forall j \in \mathcal{K}\backslash\{\mathrm{K}\}; \tag{4.7b}$$

$$0 = \sum_{i=\mathrm{I}}^{\mathrm{K}} x_\alpha^i \, \nabla_{\boldsymbol{x}_\alpha} \{\ln \Phi_\alpha^i\}(\boldsymbol{x}_\alpha). \tag{4.7c}$$

Proof. See [33, Lemma 2.2]. □

A 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 such that they are the fugacity coefficients. This implies, in particular, that the functions Φ_α^i satisfy the Gibbs-Duhem condition (4.7c). Given $\boldsymbol{c} \in \Omega$ and an admissible family of fuagacity coefficients, the phase equilibrium problem in the unified formulation amounts to solving the nonlinear system

$$Y\xi_G^i + (1-Y)\xi_L^i - c^i = 0, \qquad \forall i \in \mathcal{K}\backslash\{\mathrm{K}\}, \tag{4.8a}$$

$$\xi_G^i \Phi_G^i(\boldsymbol{x}_G) - \xi_L^i \Phi_L^i(\boldsymbol{x}_L) = 0, \qquad \forall i \in \mathcal{K}, \tag{4.8b}$$

$$\min\left(Y,\, 1 - \textstyle\sum_{j\in\mathcal{K}} \xi_G^j\right) = 0, \tag{4.8c}$$

$$\min\left(1-Y,\, 1 - \textstyle\sum_{j\in\mathcal{K}} \xi_L^j\right) = 0, \tag{4.8d}$$

in the unknowns $(Y, \{\xi_G^i\}_{i\in\mathcal{K}}, \{\xi_L^i\}_{i\in\mathcal{K}}) \in \mathbb{R} \times \mathbb{R}^{\mathrm{K}} \times \mathbb{R}^{\mathrm{K}}$. for $\alpha \in \{G, L\}$. Here, the fraction $Y \in [0,1]$ measures the overall importance of phase G in the mixture. As a result, $1-Y$ reflects the global presence of phase L. The quantities $\boldsymbol{\xi}_\alpha = (\xi_\alpha^{\mathrm{I}}, \ldots, \xi_\alpha^{\mathrm{K}}) \in \mathbb{R}_+^{\mathrm{K}}$ are called *extended fractions* of the components in phase α. If a phase $\alpha \in \{G, L\}$ is present, that is, $Y > 0$ or $1 - Y > 0$, then the corresponding extended fractions sum to 1, as can be seen from the complementarity conditions (4.8c)–(4.8d). In such a case, they coincide with the familiar *partial fractions*.

The $\mathrm{K} - 1$ equations (4.8a) are material balance of component $i \in \mathcal{K}\backslash\{K\}$. It seems that the material balance for the last species K is missing, but in fact it can be readily recovered by combining (4.8a) and (4.8c)–(4.8d). Therefore, it is redundant and must be left out of the problem statement. The K equalities (4.8b) express the extended thermodynamic equilibrium of

each component across the phases. In (4.8b), it is very important to be aware of the fact that $\boldsymbol{x}_\alpha$ is the *renormalized* fraction vector defined as

$$x_\alpha^i = \frac{\xi_\alpha^i}{\sum_{j\in\mathcal{K}} \xi_j^\alpha}. \tag{4.8e}$$

System (4.8) is of the form (1.3) with $\ell = 2\mathrm{K}+1$, $m = 2$, $X = (Y, \xi_G^{\mathrm{I}}, \dots, \xi_G^{\mathrm{K}}, \xi_L^{\mathrm{I}}, \dots, \xi_L^{\mathrm{K}})$,

$$\Lambda(X) = \begin{bmatrix} Y\xi_G^{\mathrm{I}} + (1-Y)\xi_L^{\mathrm{I}} - c^{\mathrm{I}} \\ \vdots \\ Y\xi_G^{\mathrm{K}-1} + (1-Y)\xi_G^{\mathrm{K}-1} - c^{\mathrm{K}-1} \\ \xi_G^{\mathrm{I}}\Phi_G^{\mathrm{I}}(\boldsymbol{x}_G) - \xi_L^{\mathrm{I}}\Phi_L^{\mathrm{I}}(\boldsymbol{x}_L) \\ \vdots \\ \xi_G^{\mathrm{K}}\Phi_G^{\mathrm{K}}(\boldsymbol{x}_G) - \xi_L^{\mathrm{K}}\Phi_L^{\mathrm{K}}(\boldsymbol{x}_L) \end{bmatrix} \in \mathbb{R}^{2\mathrm{K}-1} \tag{4.9a}$$

and

$$G(X) = \begin{bmatrix} Y \\ 1-Y \end{bmatrix} \in \mathbb{R}^2, \qquad H(X) = \begin{bmatrix} 1-\xi_G^1-\dots-\xi_G^{\mathrm{K}} \\ 1-\xi_L^1-\dots-\xi_L^{\mathrm{K}} \end{bmatrix} \in \mathbb{R}^2, \tag{4.9b}$$

REMARK 4.1. The thermodynamic functions g_α, Ψ_α and Φ_α^i also depend on the pressure P and the temperature T. We voluntarily omitted to indicate this dependence, since the phase equilibrium problem (4.8) is set at fixed (P, T).

4.2 Existence and construction of a solution

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. In [9], we showed that under the following additional hypotheses, an explicit solution of (4.8) can be worked out.

Hypotheses 4.1. The gradient map $\nabla g_\alpha : \Omega \to \mathbb{R}^{\mathrm{K}-1}$ 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 [12] for a twice differentiable function:

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

$$\big\langle \nabla g_\alpha(\boldsymbol{x}) - \nabla g_\alpha(\boldsymbol{y}),\, \boldsymbol{x}-\boldsymbol{y}\big\rangle > 0. \tag{4.10}$$

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

This solution, which may not be unique, is inspired from Gibbs' geometric construction for the binary case [15] ($\mathrm{K} = 2$), which is depicted in Figure 1. Let $g = \min\{g_G, g_L\}$ and let $\breve{g}$ be the lower convex envelope of g on $\bar{\Omega}$. By design, $\breve{g}$ is a convex and continuous function. It can be shown [9, Lemma 3.2] that under Hypotheses 4.1, $\breve{g}$ is differentiable at all interior point.

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

$$\bar{\Gamma}(\boldsymbol{c}) = \big\{\alpha\in\mathscr{P} \mid \exists\, \bar{\boldsymbol{x}}_\alpha \in \Omega,\ \ g_\alpha(\bar{\boldsymbol{x}}_\alpha) = T_{\boldsymbol{c}}\breve{g}(\bar{\boldsymbol{x}}_\alpha),\ \ \nabla g_\alpha(\bar{\boldsymbol{x}}_\alpha) = \nabla\breve{g}(\boldsymbol{c})\big\} \tag{4.11}$$

as the set of thoses phases whose Gibbs function g_α is tangent to the hyperplane $T_{\boldsymbol{c}}\breve{g}$. It can then be proven [9, Lemma 3.3] that: (i) $\bar{\Gamma}(\boldsymbol{c}) \neq \varnothing$; (ii) for each $\alpha \in \bar{\Gamma}(\boldsymbol{c})$, the contact point $\bar{\boldsymbol{x}}_\alpha$ is unique; (iii) if $P \leqslant \mathrm{K}$, then $\boldsymbol{c} \in \mathrm{conv}\{\bar{\boldsymbol{x}}_\alpha\}_{\alpha\in\bar{\Gamma}(\boldsymbol{c})}$.

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 $\sum_{\alpha\in\bar{\Gamma}(\boldsymbol{c})}\bar{Y}_\alpha\bar{\boldsymbol{x}}_\alpha = \boldsymbol{c}$. The remaining unknowns are determined as follows.

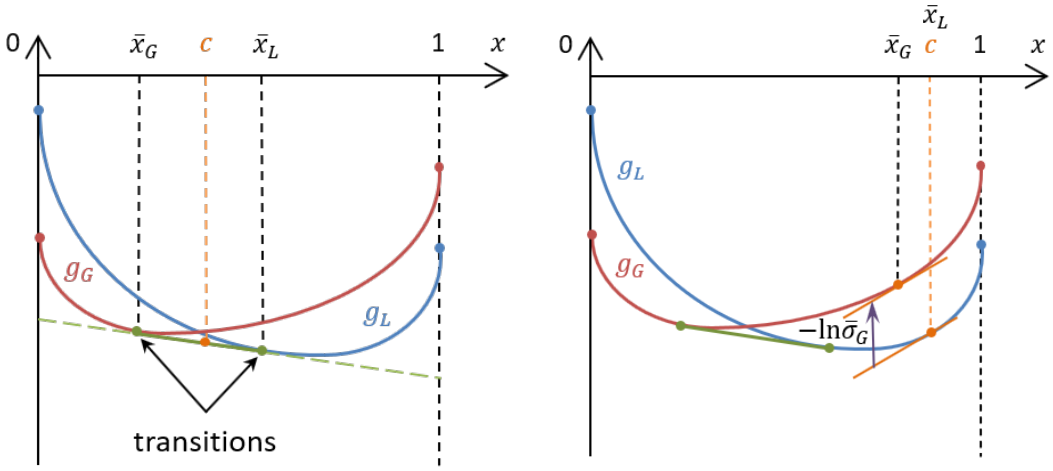


Figure 1: Gibbs' geometric construction for the phase equilibrium of a two-phase binary mixture. Left: two-phase solution. Right: single-phase solution.

Theorem 4.1 (Geometric solution)**.** *Assume* $\mathrm{K} \geqslant 2$, $\boldsymbol{c} \in \Omega$. *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*

$$\bar{\xi}^i_\alpha = \bar{x}^i_\alpha, \qquad \text{for } (\alpha, i) \in \bar{\Gamma}(\boldsymbol{c}) \times \mathcal{K}. \tag{4.12a}$$

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

$$\bar{Y}_\beta = 0, \qquad \bar{\boldsymbol{x}}_\beta = [\nabla g_\beta]^{-1}(\nabla \breve{g}(\boldsymbol{c})), \qquad \bar{\xi}^i_\beta = \exp[T_{\boldsymbol{c}}\breve{g}(\bar{\boldsymbol{x}}_\beta) - g_\beta(\bar{\boldsymbol{x}}_\beta)]\bar{x}^i_\beta. \tag{4.12b}$$

This procedure supplies us with a solution of (4.8).

Proof. See [9, Theorem 3.6]. Note that $[\nabla g_\beta]^{-1}$ is well-defined thanks to Hypotheses 4.1. □

4.3 Regularity of zeros

According to Theorem 3.1, the promise of global convergence for the NPIPM algorithm hinges on the regularity of the zeros of the system at hand. Put another way, if we could prove that the Jacobian matrix $\nabla\mathbb{F}(\bar{\mathbb{X}})$ at a solution $\bar{\mathbb{X}}$ is nonsingular, this would be an auspicious sign of the adequacy of the NPIPM algorithm to the problem. To study the regularity of a zero of (4.8), we need to pay attention to two kinds of "singular" solution.

Definition 4.1 (Transition point)**.** A solution $(\bar{Y}, \bar{\boldsymbol{\xi}}_G, \bar{\boldsymbol{\xi}}_L) \in \mathbb{R} \times \mathbb{R}^{\mathrm{K}} \times \mathbb{R}^{\mathrm{K}}$ of (4.8) is said to be a *transition* point when both arguments of one of the complementarity conditions vanish simultaneously, that is,

$$\left\{\bar{Y} = 0, \quad 1 - \sum_{i\in\mathcal{K}} \bar{\xi}^i_G = 0\right\} \qquad \text{or} \qquad \left\{\bar{Y} = 1, \quad 1 - \sum_{i\in\mathcal{K}} \bar{\xi}^i_L = 0\right\}. \tag{4.13}$$

In the two-phase framework, such a point marks the change in the nature of the solution, from a two-phase regime to a single-phase regime or vice-versa.

Definition 4.2 (Azeotropic composition)**.** A global composition $\boldsymbol{c} \in \Omega$ is said to be *azeotropic* if the graphs of g_G and g_L are tangent to each other at $\boldsymbol{c}$. In other words,

$$g_G(\boldsymbol{c}) = g_L(\boldsymbol{c}), \qquad \nabla g_G(\boldsymbol{c}) = \nabla g_L(\boldsymbol{c}). \tag{4.14}$$

Figure 2 illustrates two azeotropic situations in the binary case. Note that $\boldsymbol{c}$ alone is not responsible for azeotropy. It takes the two Gibbs functions to behave in a peculiar way to satisfy (4.14). If azeotropy occurs at $\boldsymbol{c} \in \Omega$, the solution of (4.8) is not unique: since the two contact points $\bar{\boldsymbol{x}}_\alpha$, $\alpha \in \{G, L\}$, coincide with each other, $\bar{Y}$ could be replaced by any other value $Y \in [0, 1]$.

The following Theorem tells us that, under Hypotheses 4.1, a solution of (4.8) is regular except for the two previous pathological cases.

Theorem 4.2. *Let $\bar{\mathbb{X}} = (\bar{X}, \bar{V}, \overline{W}, \bar{\nu}) \in \mathbb{R}^{2\mathrm{K}+6}$ be a solution of (3.1), (3.2) using the functions (4.9). Assume that $\bar{\nu} = 0$ and that the Gibbs energy functions g_G and g_L meet Hypotheses 4.1. Then, $\bar{\mathbb{X}}$ is a regular zero if and only if $\bar{X}$ is neither a transition point nor an azeotropic point.*

Proof. The proof is based on a brute-force calculation. We first invoke Lemma 3.1 and Lemma 2.1 successively to get

$$\det \nabla \mathbb{F}(\bar{\mathbb{X}}) = (\eta + 2\bar{\nu} + \langle \bar{V}, \overline{W} \rangle / m) \left| \begin{array}{c} \nabla \Lambda(\bar{X}) \\ \nabla G(\bar{X}) \odot H(\bar{X}) + \nabla H(\bar{X}) \odot G(\bar{X}) \end{array} \right| =: \eta\, \bar{\mathfrak{d}}, \tag{4.15}$$

with $\bar{\nu} = \langle \bar{V}, \overline{W} \rangle / m = 0$. Next, we observe that the determinant $\bar{\mathfrak{d}}$ in the right-hand side has the same sign as

$$\bar{\mathfrak{d}}^\bullet = \left| \begin{array}{c} \nabla \Lambda^\bullet(\bar{X}) \\ \nabla G(\bar{X}) \odot H(\bar{X}) + \nabla H(\bar{X}) \odot G(\bar{X}) \end{array} \right|,$$

where

$$\Lambda^\bullet(X) = \begin{bmatrix} Y \xi_G^{\mathrm{I}} + (1-Y) \xi_L^{\mathrm{I}} - c^{\mathrm{I}} \\ \vdots \\ Y \xi_G^{\mathrm{K}-1} + (1-Y) \xi_G^{\mathrm{K}-1} - c^{\mathrm{K}-1} \\ \ln(\xi_G^{\mathrm{I}} \Phi_G^{\mathrm{I}}(\boldsymbol{x}_G)) - \ln(\xi_L^{\mathrm{I}} \Phi_L^{\mathrm{I}}(\boldsymbol{x}_L)) \\ \vdots \\ \ln(\xi_G^{\mathrm{K}} \Phi_G^{\mathrm{K}}(\boldsymbol{x}_G)) - \ln(\xi_L^{\mathrm{K}} \Phi_L^{\mathrm{K}}(\boldsymbol{x}_L)) \end{bmatrix}.$$

This is because the logarithm is an increasing function and because we are at a solution, which allows us to take out the same positive factor on each line. To compute $\bar{\mathfrak{d}}^\bullet$, we go through a long sequence of row and column linear combinations. A lot of simplifications occur thanks to the Gibbs-Duhem condition (4.3c). At the end of the process, we obtain the following cases:

- If $\bar{Y} = 1$, i.e., the solution is single-phase in G, then

$$\bar{\mathfrak{d}}^\bullet = \frac{1 - \bar{\sigma}_L}{(\bar{\sigma}_L)^{\mathrm{K}}} \det \nabla^2 g_L(\bar{\boldsymbol{x}}_L),$$

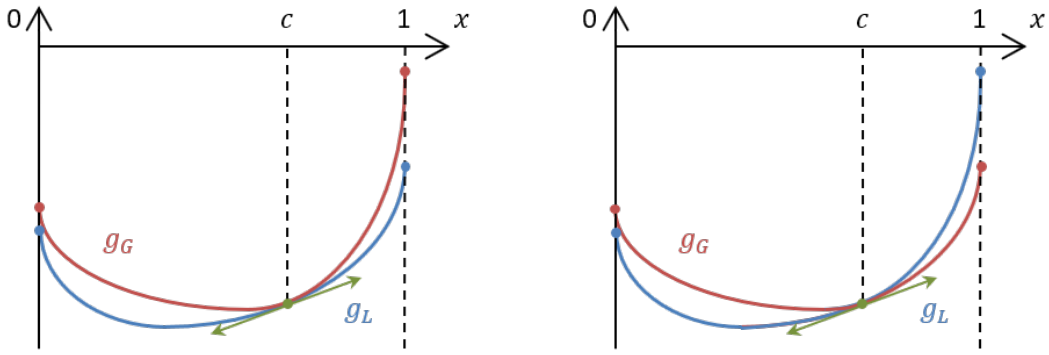


Figure 2: Azeotropic compositions for a two-phase two-component mixture.

where $\bar\sigma_L = \sum_{i\in\mathcal{K}} \bar\xi^i_L$. Because of strict convexity of g_L, this quantity vanishes only when $\bar\sigma_L = 1$, which implies that we are at a transition point.

- If $\bar Y = 0$, i.e., the solution is single-phase in L, and the discussion is similar.
- If $\bar Y \in (0,1)$, i.e., the solution is two-phase, then

$$\bar{\mathfrak{d}}^\bullet = \bar Y(1-\bar Y)\det\big[(1-\bar Y)\nabla^2 g_G(\bar{\boldsymbol{x}}_G) + \bar Y\nabla^2 g_L(\bar{\boldsymbol{x}}_L)\big]\,\Delta\bar{\boldsymbol{x}}^T(\mathsf{M}_G(\bar{\boldsymbol{x}}_G)+\mathsf{M}_L(\bar{\boldsymbol{x}}_L))\Delta\bar{\boldsymbol{x}},$$

where $\Delta\bar{\boldsymbol{x}} = \bar{\boldsymbol{x}}_G - \bar{\boldsymbol{x}}_L$ and where both symmetric matrices

$$\mathsf{M}_G = (\nabla^2 g_G)^{1/2}\Big\{I_{\mathrm{K}-1} - \Big[I_{\mathrm{K}-1} + \frac{1-\bar Y}{\bar Y}(\nabla^2 g_G)^{-1/2}\,\nabla^2 g_L\,(\nabla^2 g_G)^{-1/2}\Big]^{-1}\Big\}(\nabla^2 g_G)^{1/2}$$
$$\mathsf{M}_L = (\nabla^2 g_L)^{1/2}\Big\{I_{\mathrm{K}-1} - \Big[I_{\mathrm{K}-1} + \frac{1-\bar Y}{\bar Y}(\nabla^2 g_L)^{-1/2}\,\nabla^2 g_G\,(\nabla^2 g_L)^{-1/2}\Big]^{-1}\Big\}(\nabla^2 g_L)^{1/2}$$

can be shown to be positive definite owing to strict convexity of g_α. To cancel $\bar{\mathfrak{d}}^\bullet$ requires $\Delta\bar{\boldsymbol{x}} = 0$, i.e., $\bar{\boldsymbol{x}}_G = \bar{\boldsymbol{x}}_L$. This is precisely the characterization of an azeotropic solution.

Full details of the calculation can be found in [33, Theorem 5.3]. □

4.4 Laws for fugacity coefficients

We now give two examples of fugacity coefficients and their associated Gibbs functions. The first one, due to Henry, is very simple and complies with Hypotheses 4.1. The second one, due to Peng-Robinson, is one of the most physically advanced laws by current standards and does not necessarily satisfy these assumptions.

4.4.1 Henry's law

If phase $\alpha \in \{G, L\}$ is an ideal fluid, for which $\Psi_\alpha \equiv 0$, then it can be checked to fulfill Hypotheses 4.1. Henry's law is a straightforward generalization, with

$$\Psi_\alpha(\boldsymbol{x}) = \sum_{i=1}^{\mathrm{K}} x^i \ln k^i \tag{4.16}$$

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 (4.7a) are then

$$\ln \Phi^j_\alpha(\boldsymbol{x}) = \ln k^i, \qquad \text{for all } \ j \in \mathcal{K}. \tag{4.17}$$

This is why this law is also called the *constant coefficients* law. Again, it is not difficult to show [33, Proposition 3.1] that the Gibbs energy function g_α associated with Henry's law fulfills Hypotheses 4.1 for all $(k^{\mathrm{I}}, \dots, k^{\mathrm{K}}) \in (\mathbb{R}^*_+)^{\mathrm{K}}$.

4.4.2 Peng-Robinson's law

In Peng-Robinson's law, the fugacity coefficients of the two phases G, L are coupled with each other and, in a sense, are computed "simultaneously." 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 to a pair of postive dimensionless parameters

$$A^i = \frac{\mathsf{P}a^i}{(\mathsf{RT})^2}, \qquad B^i = \frac{\mathsf{P}b^i}{\mathsf{RT}}. \tag{4.18}$$

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 (A^i, B^i)'s and the current partial fractions by means of a *mixing rule* $\boldsymbol{x} \mapsto (A(\boldsymbol{x}), B(\boldsymbol{x}))$. There can be found [30] a wide variety of mixing rules . We choose the most common one, namely,

$$A(\boldsymbol{x}) = \left(\sum_{j \in \mathcal{K}} x^j \sqrt{A^j} \right)^2, \qquad B(\boldsymbol{x}) = \sum_{j \in \mathcal{K}} x^j B^j. \tag{4.19}$$

The next step is to consider the cubic equation

$$\begin{aligned} &Z^3(\boldsymbol{x}) + (B(\boldsymbol{x}) - 1)Z^2(\boldsymbol{x}) \\ &\quad + [A(\boldsymbol{x}) - 2B(\boldsymbol{x}) - 3B^2(\boldsymbol{x})]Z(\boldsymbol{x}) + [B^2(\boldsymbol{x}) + B^3(\boldsymbol{x}) - A(\boldsymbol{x})B(\boldsymbol{x})] = 0 \end{aligned} \tag{4.20}$$

in the variable $Z(\boldsymbol{x})$. In the most favorable situation, there are three real roots, all greater than $B(\boldsymbol{x})$. These are then named $B(\boldsymbol{x}) < Z_L(\boldsymbol{x}) < Z_I(\boldsymbol{x}) < Z_G(\boldsymbol{x})$. In other words, the smallest root is associated with the liquid phase L, while the largest one is associated with the gas phase G. As for the intermediate root $Z_I(\boldsymbol{x})$, it does not have any physical meaning.

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

$$\Psi_\alpha(\boldsymbol{x}) = Z_\alpha(\boldsymbol{x}) - 1 - \ln\left[Z_\alpha(\boldsymbol{x}) - B(\boldsymbol{x})\right] - \frac{A(\boldsymbol{x})}{2\sqrt{2}B(\boldsymbol{x})} \ln\left[\frac{Z_\alpha(\boldsymbol{x}) + (1+\sqrt{2})B(\boldsymbol{x})}{Z_\alpha(\boldsymbol{x}) - (\sqrt{2}-1)B(\boldsymbol{x})}\right]. \tag{4.21}$$

From this, the fugacity coefficients can be deduced with the help of (4.7a). Combining the result with the mixing rule (4.19), we end up with [33, Corollary 3.2]

$$\begin{aligned} \ln \Phi^i_\alpha(\boldsymbol{x}) = \; & \frac{B^i}{B(\boldsymbol{x})}\left[Z_\alpha(\boldsymbol{x}) - 1\right] - \ln\left[Z_\alpha(\boldsymbol{x}) - B(\boldsymbol{x})\right] \\ & + \left[\frac{B^i}{B(\boldsymbol{x})} - \frac{2A^i(\boldsymbol{x})}{A(\boldsymbol{x})}\right] \frac{A(\boldsymbol{x})}{2\sqrt{2}B_(\boldsymbol{x})} \ln\left[\frac{Z_\alpha(\boldsymbol{x}) + (1+\sqrt{2})B(\boldsymbol{x})}{Z_\alpha(\boldsymbol{x}) - (\sqrt{2}-1)B(\boldsymbol{x})}\right]. \end{aligned} \tag{4.22}$$

using the "matrix-vector" product

$$A^i(\boldsymbol{x}) = \sqrt{A^i} \left(\sum_{j \in \mathcal{K}} x^j \sqrt{A^j} \right). \tag{4.23}$$

In the unfavorable situation when equation (4.20) has only on real root greater than $B(\boldsymbol{x})$, 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 (4.21), leaving its counterpart in the other phase undefined. Otherwise, Ψ_α is undefined in both phases. This process raises two serious questions:

1. When does the cubic equation has 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 $B(\boldsymbol{x})$ and when is it impossible?

These questions were addressed at length in [33, §3.2.3] (see also [9] for a shorter presentation), where we stressed the critical issue of the Gibbs functions g_G, g_L not being defined on the whole domain Ω. To sketch out the difficulty, let us consider the binary case where $\boldsymbol{x} = x \in (0, 1) = \Omega$. Then, for some data $(A^{\mathrm{I}}, B^{\mathrm{I}})$ and $(A^{\mathrm{II}}, B^{\mathrm{II}})$, there exist $0 < x_\flat < x_\sharp < 1$ such that the quantities $Z_L(x)$, $\Psi_L(x)$, $g_L(x)$ are well-defined only for $x \in [0, x_\sharp]$, while the quantities $Z_G(x)$, $\Psi_G(x)$, $g_G(x)$ are well-defined only for $x \in [x_\flat, 1]$. Furthermore, since $g'_G(x_\flat^+)$ and $g'_L(x_\sharp^-)$ are finite, the image

sets $g'_G([x_\flat, 1))$ and $g'_L((0, x_\sharp])$ are not equal to $\mathbb{R}$. This prevents the unified formulation from being always able to assigning a well-defined extended fraction to a vanishing phase by inverting g'_G and g'_L.

To circumvent this obstacle, we proposed [33, §3.3] a procedure to extend the domains of defintion of the Gibbs functions to $\bar{\Omega}$. When the cubic equation does not have three real roots, the 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

$$Z^3 + (B-1)Z^2 + (A - 2B - 3B^2)Z + (B^2 + B^3 - AB) = 0.$$

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

$$W_\beta = \frac{1 - B - Z_\alpha}{2}. \tag{4.24}$$

It can then be proven that for physically reasonable values of B, we have $W_\beta > B$ and

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

These properties of W_β are the constraints to which Z_β would have been subject, had it existed. They speak in favor of the enrollment of W_β as a substitute for Z_β. Doing so yields the Gibbs function

$$\Psi_\beta = W_\beta - 1 - \ln\big[W_\beta - B\big] - \frac{A}{2\sqrt{2}B}\,\ln\left[\frac{W_\beta + (\sqrt{2}+1)B}{W_\beta - (\sqrt{2}-1)B}\right]. \tag{4.25}$$

for the missing phase β. By (4.7a), we can derive the corresponding fugacity coefficients. In view of the mixing rule (4.19), these are given by

$$\begin{aligned}
\ln \Phi^i_\beta(\boldsymbol{x}) = \;& \frac{B^i}{B}\,[W_\beta - 1] - \ln\big[W_\beta - B\big] \\
&+ \left[\frac{B^i}{B} - \frac{2A^i(\boldsymbol{x})}{A}\right]\frac{A}{2\sqrt{2}B}\,\ln\left[\frac{W_\beta + (\sqrt{2}+1)B}{W_\beta - (\sqrt{2}-1)B}\right] \\
&+ \left[\frac{\langle \nabla W_\beta,\, \boldsymbol{\delta}^i - \boldsymbol{x}\rangle}{W_\beta} - \frac{\langle \nabla B,\, \boldsymbol{\delta}^i - \boldsymbol{x}\rangle}{B}\right]\frac{W_\beta\,\Upsilon_{A,B}(W_\beta)}{(W_\beta - B)(W_\beta^2 + 2BW_\beta - B^2)}
\end{aligned} \tag{4.26}$$

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

The last step of this procedure is a "sacrifice" of a tiny portion of the three-root region in order for ∇Z_β not to blow up when approaching the one-root region. To this end, we introduce

$$\vartheta = \frac{Z_I - Z_L}{Z_G - Z_L} \;\in [0, 1] \tag{4.27}$$

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 $\widetilde{Z}_G = [1 - \nu_G(\vartheta)]Z_G + \nu_G(\vartheta)W_G$ into (4.21), where

$$\nu_G(\vartheta) = \begin{cases} 0 & \text{if } \vartheta \leqslant 1 - 2\varepsilon, \\ q((\vartheta - (1 - 2\varepsilon))/\varepsilon) & \text{if } \vartheta \in (1 - 2\varepsilon, 1 - \varepsilon), \\ 1 & \text{if } \vartheta \geqslant 1 - \varepsilon, \end{cases} \tag{4.28}$$

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

$$\Psi_G = \tilde{Z}_G - 1 - \ln\left[\tilde{Z}_G - B\right] - \frac{A}{2\sqrt{2}B} \ln\left[\frac{\tilde{Z}_G + (\sqrt{2}+1)B}{\tilde{Z}_G - (\sqrt{2}-1)B}\right] \tag{4.29a}$$

and from the rule (4.7a), the fugacity coefficients are inferred as

$$\begin{aligned} \ln \Phi_G^i = \; & \frac{B^i}{B}\left[\tilde{Z}_G - 1\right] - \ln\left[\tilde{Z}_G - B\right] \\ & + \left[\frac{B^i}{B} - \frac{2A^i(\boldsymbol{x})}{A}\right] \frac{A}{2\sqrt{2}B} \ln\left[\frac{\tilde{Z}_G + (\sqrt{2}+1)B}{\tilde{Z}_G - (\sqrt{2}-1)B}\right] \\ & + \left[\frac{\langle \nabla \tilde{Z}_G,\, \boldsymbol{\delta}^i - \boldsymbol{x}\rangle}{\tilde{Z}_G} - \frac{\langle \nabla B,\, \boldsymbol{\delta}^i - \boldsymbol{x}\rangle}{B}\right] \frac{\tilde{Z}_G\, \Upsilon_{A,B}(\tilde{Z}_G)}{(\tilde{Z}_G - B)(\tilde{Z}_G^2 + 2B\tilde{Z}_G - B^2)}. \end{aligned} \tag{4.29b}$$

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

To conclude this section, we lay emphasis on the fact the domain extension procedure is aimed at improving the unified formulation's chance of "survival," which will be corroborated by the numerical experiements of §6. We do not strive to fulfill Hypotheses 4.1, since these assumptions may already be violated for the original unextended Gibbs functions.

5 Evolutionary model for binary two-phase mixtures

The second model on which comparisons between Newton-min and NPIPM are carried out is a simplified system with a time evolution. To build this model, we start from the stationary two-phase binary equilibrium problem and impose a differentiel equation for the global fraction on the top of it.

5.1 A simplified ODE system

In the binary case of the stationary model (4.8), the global fraction of the first component c is a given data. Let us make it depend on time by considering the algebro-differential system

$$\frac{\mathrm{d}c}{\mathrm{d}t} - Y(1-Y)\xi_G^{\mathrm{I}}(1 - 1/k^{\mathrm{I}}) = 0, \tag{5.1a}$$

$$Y\xi_G^{\mathrm{I}} + (1-Y)\xi_G^{\mathrm{I}}/k^{\mathrm{I}} - c = 0, \tag{5.1b}$$

$$\min(Y,\, 1 - \xi_G^{\mathrm{I}} - \xi_G^{\mathrm{II}}) = 0, \tag{5.1c}$$

$$\min(1-Y,\, 1 - \xi_G^{\mathrm{I}}/k^{\mathrm{I}} - \xi_G^{\mathrm{II}}/k^{\mathrm{II}}) = 0 \tag{5.1d}$$

in the four unknowns $(c,\, Y,\, \xi_G^{\mathrm{I}},\, \xi_G^{\mathrm{II}})$, equipped with the intial condition

$$(c,\, Y,\, \xi_G^{\mathrm{I}},\, \xi_G^{\mathrm{II}})(t=0) = (c_0, Y_0, (\xi_G^{\mathrm{I}})_0, (\xi_G^{\mathrm{II}})_0) \tag{5.2}$$

subject to the equilibrium relations

$$Y_0(\xi_G^{\mathrm{I}})_0 + (1-Y_0)(\xi_G^{\mathrm{II}})_0/k^{\mathrm{I}} - c_0 = 0, \tag{5.3a}$$

$$\min(Y_0,\, 1 - (\xi_G^{\mathrm{I}})_0 - (\xi_G^{\mathrm{II}})_0) = 0, \tag{5.3b}$$

$$\min(1-Y_0,\, 1 - (\xi_G^{\mathrm{I}})_0/k^{\mathrm{I}} - (\xi_G^{\mathrm{II}})_0/k^{\mathrm{II}}) = 0. \tag{5.3c}$$

At fixed c, the last three equations of (5.1) are in fact the stationary equilibrium problem (4.8) with K = 2, in which the variables ξ_L^{I} and ξ_L^{II} have been eliminated with the help of the equilibrum relations

$$\xi_G^{\mathrm{I}} = k^{\mathrm{I}}\xi_L^{\mathrm{I}}, \qquad \xi_G^{\mathrm{II}} = k^{\mathrm{II}}\xi_L^{\mathrm{II}}.$$

The latter mean that phase G is an ideal gas and phase L is governed by Henry's law.

For simplicity, we restrict ourselves to $k^{\mathrm{I}} > 1 > k^{\mathrm{II}} > 0$. Let K_G, K_L be the two constants

$$\frac{k^{\mathrm{I}}(1-k^{\mathrm{II}})}{k^{\mathrm{I}}-k^{\mathrm{II}}} =: K_G > K_L := \frac{1-k^{\mathrm{II}}}{k^{\mathrm{I}}-k^{\mathrm{II}}} > 0. \tag{5.4}$$

It can then be proven that the exact solution of (5.1)–(5.3) is given by

$$c(t) = \begin{cases} c_0 & \text{if } c_0 \in [0, K_L], \\ \dfrac{K_G\gamma_0\exp(t) + K_L}{\gamma_0\exp(t)+1} & \text{if } c_0 \in (K_L, K_G), \\ c_0 & \text{if } c_0 \in [K_G, 1], \end{cases} \tag{5.5}$$

where $\gamma_0 = (c_0 - K_L)/(K_G - c_0)$. The values of $Y(t)$, $\xi_G^{\mathrm{I}}(t)$ and $\xi_G^{\mathrm{II}}(t)$ are deduced from $c(t)$ by the formulas

$$(\bar{Y},\, \bar{\xi}_G^{\mathrm{I}},\, \bar{\xi}_G^{\mathrm{II}}) = \begin{cases} (0,\, k^{\mathrm{I}}c,\, k^{\mathrm{II}}(1-c)) & \text{if } c \in [0, K_L], \\ \left(\dfrac{c-K_L}{K_G-K_L},\, K_G,\, k^{\mathrm{II}}(1-K_L)\right) & \text{if } c \in (K_L, K_G), \\ (1,\, c,\, 1-c) & \text{if } c \in [K_G, 1], \end{cases} \tag{5.6}$$

which are none other than the solution of the stationary two-phase binary equilibrium problem (4.8) at fixed c.

But our primary interest is the algebraic system that arises when we apply the Euler backward scheme to (5.1) with a time-step $\tau > 0$. This system reads

$$c - c_\flat - \tau\left(1 - \frac{1}{k^{\mathrm{I}}}\right)\xi_G^{\mathrm{I}} Y(1-Y) = 0, \tag{5.7a}$$

$$Y\xi_G^{\mathrm{I}} + (1-Y)\xi_G^{\mathrm{I}}/k^{\mathrm{I}} - c = 0, \tag{5.7b}$$

$$\min(Y,\, 1 - \xi_G^{\mathrm{I}} - \xi_G^{\mathrm{II}}) = 0, \tag{5.7c}$$

$$\min(1-Y,\, 1 - \xi_G^{\mathrm{I}}/k^{\mathrm{I}} - \xi_G^{\mathrm{II}}/k^{\mathrm{II}}) = 0, \tag{5.7d}$$

where $c_\flat$ denotes the value of c at the previous time-step. In (5.7), $c_\flat \in [0,1]$, $\tau > 0$ and $k^{\mathrm{I}} > 1 > k^{\mathrm{II}} > 0$ play the role of parameters. System (5.7) is of the form (1.3) with $\ell = 4$, $m = 2$,

$$\Lambda(X) = \begin{bmatrix} c - c_\flat - \tau(1-1/k^{\mathrm{I}})\,\xi_G^{\mathrm{I}}\,Y(1-Y) \\ Y\xi_G^{\mathrm{I}} + (1-Y)\,\xi_G^{\mathrm{I}}/k^{\mathrm{I}} - c \end{bmatrix}, \tag{5.8a}$$

and

$$G(X) = \begin{bmatrix} Y \\ 1-Y \end{bmatrix}, \qquad H(X) = \begin{bmatrix} 1 - \xi_G^{\mathrm{I}} - \xi_G^{\mathrm{II}} \\ 1 - \xi_G^{\mathrm{I}}/k^{\mathrm{I}} - \xi_G^{\mathrm{II}}/k^{\mathrm{II}} \end{bmatrix}. \tag{5.8b}$$

5.2 Reference solution

The upcoming Proposition deals with the exitence and uniqueness of a solution to (5.7).

Proposition 5.1. *Let K_L, K_G be the constants defined by (5.4). Except for the case 3(b) in the enumeration below, system (5.7) has a unique solution $(\bar{c}, \bar{Y}, \bar{\xi}_G^{\mathrm{I}}, \bar{\xi}_G^{\mathrm{II}}) \in [0,1]\times[0,1]\times\mathbb{R}_+\times\mathbb{R}_+$ called* reference solution.

1. *If* $c_\flat \in [K_G, 1]$, *then the reference solution is in the* G *single-phase regime and given by*

$$\bar{c} = c_\flat, \quad \bar{Y} = 1, \quad \bar{\xi}_G^{\mathrm{I}} = c_\flat, \quad \bar{\xi}_G^{\mathrm{II}} = 1 - c_\flat. \tag{5.9}$$

2. *If* $c_\flat \in (K_L, K_G)$, *then the reference solution is in the two-phase regime and given by*

$$\bar{c} = \frac{K_G + K_L}{2} - \frac{K_G - K_L}{2\tau}\left\{1 - \left[1 - 2\frac{K_G + K_L - 2c_\flat}{K_G - K_L}\tau + \tau^2\right]^{1/2}\right\}. \tag{5.10}$$

The values of $\bar{Y}$, $\bar{\xi}_G^{\mathrm{I}}$ *and* $\bar{\xi}_G^{\mathrm{II}}$ *are deduced from* $\bar{c}$ *by formulas* (5.6).

3. *If* $c_\flat \in [0, K_L]$, *then the number*

$$\tau_{\max} = \frac{K_G + K_L - 2c_\flat}{K_G - K_L} + \sqrt{\left(\frac{K_G + K_L - 2c_\flat}{K_G - K_L}\right)^2 - 1} \tag{5.11}$$

is well-defined and greater than or equal to 1.

   (a) *For* $\tau < \tau_{\max}$, *the reference solution is in the* L *single-phase regime and given by*

$$\bar{c} = c_\flat, \quad \bar{Y} = 0, \quad \bar{\xi}_G^{\mathrm{I}} = k^{\mathrm{I}} c_\flat, \quad \bar{\xi}_G^{\mathrm{II}} = k^{\mathrm{II}}(1 - c_\flat); \tag{5.12}$$

   (b) *For* $\tau \geqslant \tau_{\max}$, *in addition to* (5.12) *that we declare to be the reference solution, there are two spurious solutions (counted with multiplicity).*

Proof. The last three equations of model (5.7) are exactly the stationary binary model (4.8). Therefore, $(Y, \xi_G^{\mathrm{I}}, \xi_G^{\mathrm{II}})$ can be expressed as functions of c by means of (5.6). In particular,

$$Y = \frac{c - K_L}{K_G - K_L}\,\mathbf{1}_{(K_L, K_G)}(c)$$

for all phase regimes, using the characteristic function $\mathbf{1}$. Inserting this into the first equation (5.7a) and invoking $k^{\mathrm{I}} = K_G/K_L$, we obtain a scalar equation on c, namely,

$$c - c_\flat + \frac{\tau}{K_G - K_L}(c - K_L)(c - K_G)\,\mathbf{1}_{(K_L, K_G)}(c) = 0. \tag{5.13}$$

The rest of the proof relies on studying the function representing the left-hand side of the above equation. This part is not difficult and is left to the readers. □

5.3 Regularity of zeros

The most significant result for this model is that the reference solution corresponds most of the time to a regular zero.

Theorem 5.1. *For all* $\tau \geqslant 0$, *the reference solution of* (5.7) *defined in Proposition* 5.1 *gives rise to a regular zero for the NPIPM system, except at transitional and azeotropic points.*

Proof. Let $X = (c, Y, \xi_G^{\mathrm{I}}, \xi_G^{\mathrm{II}})$ et recall the notations of (5.8). By Lemma 3.1 and Lemma 2.1, we can study the sign of

$$\bar{\mathfrak{d}} = \begin{vmatrix} \nabla\Lambda(\bar{X}) \\ \nabla G(\bar{X}) \odot H(\bar{X}) + \nabla H(\bar{X}) \odot G(\bar{X}) \end{vmatrix},$$

where $\bar{X} = (\bar{c}, \bar{Y}, \bar{\xi}_G^{\mathrm{I}}, \bar{\xi}_G^{\mathrm{II}})$ is the reference solution, instead of the sign of $\det \nabla\mathbb{F}(\bar{\mathbb{X}})$ or $\det \nabla\mathbf{F}(\bar{\mathbf{X}})$. In this case, we have

$$\bar{\mathfrak{d}} = \begin{vmatrix} 1 & -\tau\Delta\bar{\xi}^{\mathrm{I}}(1 - 2\bar{Y}) & -\tau\Delta\bar{\xi}^{\mathrm{I}}\,\bar{Y}(1 - \bar{Y})/\bar{\xi}_G^{\mathrm{I}} & 0 \\ -1 & \Delta\bar{\xi}^{\mathrm{I}} & \bar{Y} + (1 - \bar{Y})/k^{\mathrm{I}} & 0 \\ 0 & 1 - \bar{\sigma}_G & -\bar{Y} & -\bar{Y} \\ 0 & -1 + \bar{\sigma}_L & (\bar{Y} - 1)/k^{\mathrm{I}} & (\bar{Y} - 1)/k^{\mathrm{II}} \end{vmatrix},$$

with

$$\Delta\bar{\xi}^{\mathrm{I}} = \bar{\xi}^{\mathrm{I}}_G - \bar{\xi}^{\mathrm{I}}_L = \bar{\xi}^{\mathrm{I}}_G(1 - 1/k^{\mathrm{I}}), \qquad \bar{\sigma}_G = \bar{\xi}^{\mathrm{I}}_G + \bar{\xi}^{\mathrm{II}}_G, \qquad \bar{\sigma}_L = \bar{\xi}^{\mathrm{I}}_L + \bar{\xi}^{\mathrm{II}}_L.$$

Expanding the determinant with respect to the first column, we find

$$\bar{\mathfrak{d}} = \bar{\mathfrak{d}}_0 - \tau\,\bar{\mathfrak{d}}_1, \tag{5.14}$$

where

$$\bar{\mathfrak{d}}_0 = \begin{vmatrix} \Delta\bar{\xi}^{\mathrm{I}} & \bar{Y} + (1-\bar{Y})/k^{\mathrm{I}} & 0 \\ 1-\bar{\sigma}_G & -\bar{Y} & -\bar{Y} \\ -1+\bar{\sigma}_L & (\bar{Y}-1)/k^{\mathrm{I}} & (\bar{Y}-1)/k^{\mathrm{II}} \end{vmatrix}$$

is the determinant of the stationary binary model and

$$\bar{\mathfrak{d}}_1 = \begin{vmatrix} \Delta\bar{\xi}^{\mathrm{I}}(1-2\bar{Y}) & \Delta\bar{\xi}^{\mathrm{I}}\,\bar{Y}(1-\bar{Y})/\bar{\xi}^{\mathrm{I}}_G & 0 \\ 1-\bar{\sigma}_G & -\bar{Y} & -\bar{Y} \\ -1+\bar{\sigma}_L & (\bar{Y}-1)/k^{\mathrm{I}} & (\bar{Y}-1)/k^{\mathrm{II}} \end{vmatrix}.$$

Assume $\bar{Y} = 1$, i.e., the solution is in the gas phase. Then, $\bar{\sigma}_G = 1$. We obtain that $\bar{\mathfrak{d}}_0 = 1 - \bar{\sigma}_L$. By a direct computation, we have $\bar{\mathfrak{d}}_1 = 0$. Therefore, $\bar{\mathfrak{d}} = \bar{\mathfrak{d}}_0 = 1 - \bar{\sigma}_L \geqslant 0$. Equality holds at a transition point. The other single-phase case $\bar{Y} = 0$ is similar.

Assume now $\bar{Y} \in (0,1)$, i.e., the solution is in the two-phase regime. Then, $\bar{\sigma}_G = \bar{\sigma}_L = 1$, $\boldsymbol{\xi}_G = \boldsymbol{x}_G$ and $\boldsymbol{\xi}_L = \boldsymbol{x}_L$. We get

$$\bar{\mathfrak{d}}_0 = \Delta\bar{x}^{\mathrm{I}} \begin{vmatrix} -\bar{Y} & -\bar{Y} \\ (\bar{Y}-1)/k^{\mathrm{I}} & (\bar{Y}-1)/k^{\mathrm{II}} \end{vmatrix}$$

can be expressed as a quadratic form and hence $\bar{\mathfrak{d}}_0 \geqslant 0$, with equality if and only if $\boldsymbol{x}_G = \boldsymbol{x}_L$, namely, at an azeotropic point. Let us compute $\bar{\mathfrak{d}}_1$. By expanding with respect to its first column and by noticing that the, we obtain

$$\bar{\mathfrak{d}}_1 = (1 - 2\bar{Y})\,\bar{\mathfrak{d}}_0.$$

Coming back to (5.14), we get

$$\bar{\mathfrak{d}} = \bar{\mathfrak{d}}_0\,[1 - \tau(1 - 2\bar{Y})].$$

From (5.10), we infer by (5.6) that

$$\tau(1-2\bar{Y}) = 1 - \left[1 - 2\frac{K_G + K_L - 2c_\flat}{K_G - K_L}\tau + \tau^2\right]^{1/2} < 1.$$

Consequently, $\bar{\mathfrak{d}}$ has the same sign behavior as $\bar{\mathfrak{d}}_0$. □

6 Numerical results

On the grounds of the previous models, we now compare NPIPM and the Newton-min method. For the stationary model of §4, we consider the binary and ternary cases. For each case, we consider two tests: one with the ideal and Henry's laws, the other with Peng-Robinson's law. The domain extension procedure (4.24)–(4.26) must be activated, otherwise both Newton-min and NPIPM may crash.

For each test, we display two figures: the first one represents the solution for various $\boldsymbol{c} \in \Omega$ using the same initial point in both methods; the second one indicates not only the number of iterations to reach convergence, starting from this initial point, but also the percentage of elements within a generated set of initial points for which convergence occurs.

Unless otherwise specified, the stopping criterion is $\|\mathbb{F}(\mathbb{X})\| < 10^{-7}$ and the maximum number of iterations to be 50 in all tests. With NPIPM, the parameters are $\eta = 0.5$, $u = 1$, $\kappa = 0.4$ and $\varrho = 0.99$.

6.1 Stationary two-phase binary model

In the two-component case, we write c, x_α instead of c^{I}, x_α^{I}. Model (4.8a) then becomes

$$Y\xi_G^{\mathrm{I}} + (1-Y)\xi_L^{\mathrm{I}} - c = 0, \tag{6.1a}$$
$$\xi_G^{\mathrm{I}}\,\Phi_G^{\mathrm{I}}(x_G) - \xi_L^{\mathrm{I}}\,\Phi_L^{\mathrm{I}}(x_L) = 0, \tag{6.1b}$$
$$\xi_G^{\mathrm{II}}\,\Phi_G^{\mathrm{II}}(x_G) - \xi_L^{\mathrm{II}}\,\Phi_L^{\mathrm{II}}(x_L) = 0, \tag{6.1c}$$
$$\min(Y; 1-\xi_G^{\mathrm{I}} - \xi_G^{\mathrm{II}}) = 0, \tag{6.1d}$$
$$\min(1-Y; 1-\xi_L^{\mathrm{I}} - \xi_L^{\mathrm{II}}) = 0, \tag{6.1e}$$

with the implicit renormalization $x_\alpha = \xi_\alpha^{\mathrm{I}}/(\xi_\alpha^{\mathrm{I}} + \xi_\alpha^{\mathrm{II}})$ for $\alpha \in \{G, L\}$.

6.1.1 With Henry's law

Phase G is an ideal gas, phase L obeys Henry's law with $k^{\mathrm{I}} = 2$, $k^{\mathrm{II}} = 0.5$. Thanks to the extreme simplicity of (6.1c), we can even eliminate ξ_L^{I}, ξ_L^{II} from the equations, as was done in §5.1 for the evolutionary model.

In the first test, starting from the same initial point $(Y, \xi_G^{\mathrm{I}}, \xi_G^{\mathrm{II}}) = (0.99, 0.67, 0.327)$ and sweeping over the grid of parameters $c \in \{0.01; 0.02; \ldots; 0.99\}$, we run the two algorithms and observe the computed solution in Figure 3. Barring a single divergence point (red dot at $c \approx 0.4$) for Newton-min, the two methods find the same solution, whose Y-component is drawn. The number of iterations required for each method is shown in panel (b) of Figure 4b.

In the second test, we sweep over the grid of parameters $c \in \{0.0001, 0.0002, ..., 0.9999\}$ and the set of initial points

$$\mathcal{D}^0 = \big\{(Y, \xi_G^{\mathrm{I}}, \xi_G^{\mathrm{II}})^0 \in \mathcal{M}^3 \;\big|\; 1 - (\xi_G^{\mathrm{I}})^0 - (\xi_G^{\mathrm{II}})^0 > 0 \text{ and } 1 - (\xi_G^{\mathrm{I}})^0/k^{\mathrm{I}} - (\xi_G^{\mathrm{II}})^0/k^{\mathrm{II}} > 0\big\},$$

where $\mathcal{M} = \{0.1; 0.2; \ldots; 0.9\}$. The number of initial points used for the tests is $|\mathcal{D}^0| = 216$. For each c, we count the number of initial points for which the method converges and then plot the percentage of success for each algorithm in Figure 4a. The success rate of NPIPM is 100%, while that of Newton-min is around 90%. In this Figure, we also plot the percentage of success corresponding to the θ-smoothing method using θ^1 and the Mehrotra predictor-corrector algorithm. We see that they are significantly worse than Newton-min and NPIPM. This is the reason why we keep only Newton-min as a reference for comparison.

6.1.2 With Peng-Robinson's law

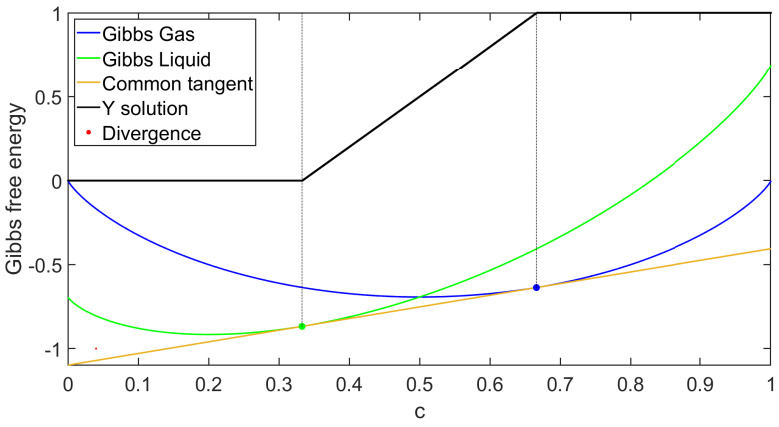
Let us choose $(A^{\mathrm{I}}, B^{\mathrm{I}}) = (0.2153, 0.03)$ and $(A^{\mathrm{II}}, B^{\mathrm{II}}) = (0.1861, 0.02)$ to ensure that the Gibbs functions g_G, g_L are "visually" strictly convex once extended.

In the first test, we use the same initial point $(Y, \xi_G^{\mathrm{I}}, \xi_G^{\mathrm{II}}, \xi_L^{\mathrm{I}}, \xi_L^{\mathrm{II}})^0 = (0.2, 0.4, 0.4, 0.6, 0.2)$ and sweep over the grid of parameters $c \in \{0.01; 0.02; \ldots; 0.99\}$. The computed solutions are displayed in Figure 5. Note that Newton-min now has many divergence points (red dots) for $c \leqslant 0.5$. The number of iterations required to reach convergence is supplied Figure 6a.

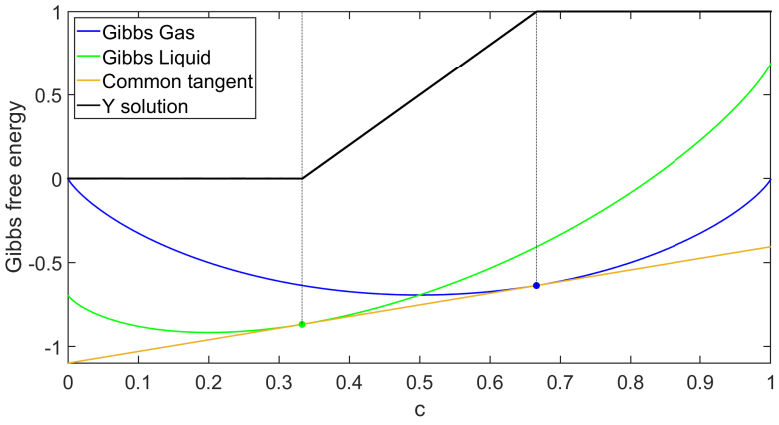
In the second test, we measure the percentage of convergence over the grid of parameters $c \in \{0.0001, 0.0002, ..., 0.9999\}$ and the set of initial points

$$\mathcal{D}^0 = \big\{(Y, \xi_G^{\mathrm{I}}, \xi_G^{\mathrm{II}}, \xi_L^{\mathrm{I}}, \xi_L^{\mathrm{II}})^0 \in \mathcal{M}^5 \;\big|\; 1 - (\xi_G^{\mathrm{I}})^0 - (\xi_G^{\mathrm{II}})^0 > 0 \text{ and } 1 - (\xi_L^{\mathrm{I}})^0 - (\xi_L^{\mathrm{II}})^0 > 0\big\},$$

where $\mathcal{M} = \{0.2; 0.4; 0.6; 0.8\}$. The number of initial points used for the tests is $|\mathcal{D}^0| = 144$. The success rate for each method shown in Figure 6b as a function of c. Again, NPIPM achieves 100%, while Newton-min culminates at about 85%.

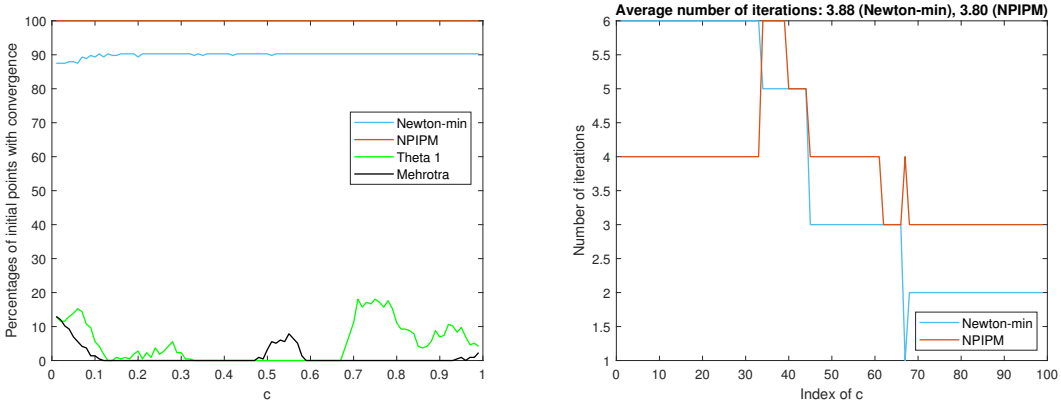


(a) Newton-min



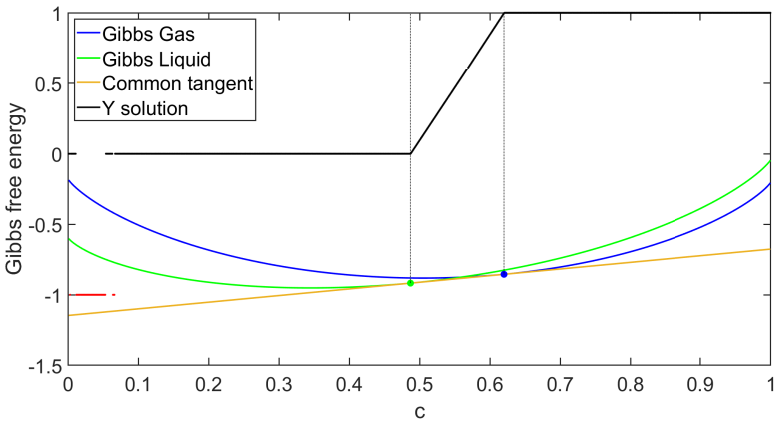
(b) NPIPM

Figure 3: Henry's law, tested with the same initial point.

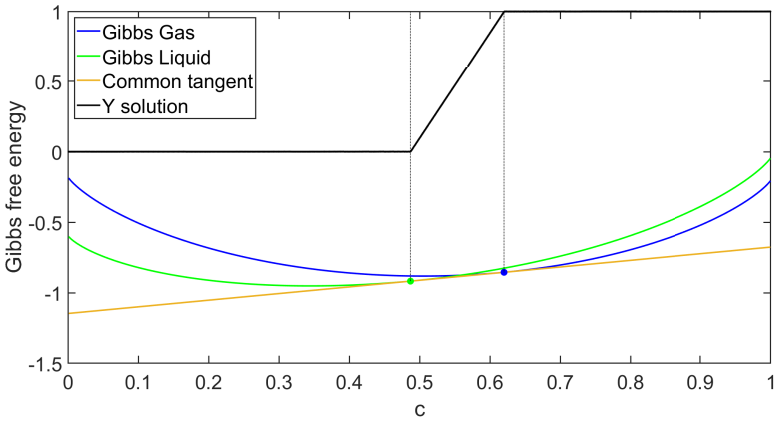


(a) Percentage of convergence over all initial point. (b) Number of iterations with the same initial points.

Figure 4: Henry's law

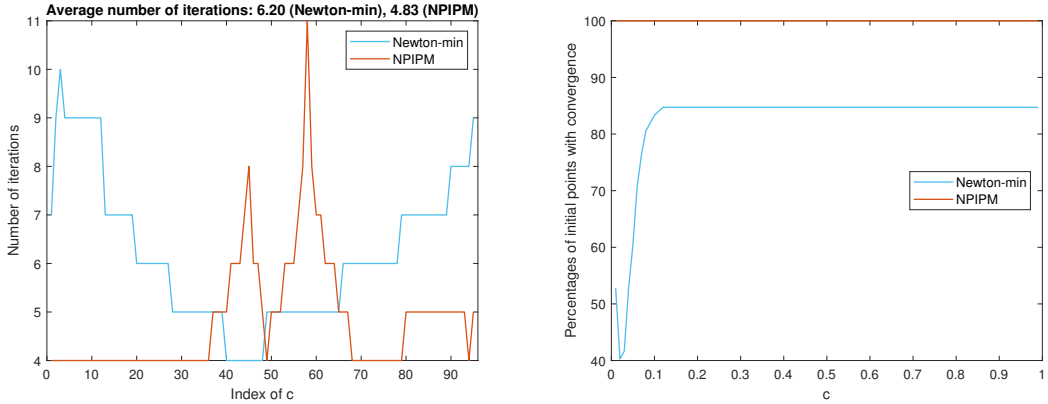


(a) Newton-min



(b) NPIPM

Figure 5: Peng-Robinson's law: one initial point.



(a) Number of iterations with the same initial points. (b) Percentage of convergence over all initial points.

Figure 6: Peng-Robinson's law.

6.2 Stationary two-phase ternary model

In the three-component case, model (4.8) reads

$$Y\xi_G^{\mathrm{I}} + (1-Y)\xi_L^{\mathrm{I}} - c^{\mathrm{I}} = 0, \tag{6.2a}$$
$$Y\xi_G^{\mathrm{II}} + (1-Y)\xi_L^{\mathrm{II}} - c^{\mathrm{II}} = 0, \tag{6.2b}$$
$$\xi_G^{\mathrm{I}}\,\Phi_G^{\mathrm{I}}(x_G^{\mathrm{I}}, x_G^{\mathrm{II}}) - \xi_L^{\mathrm{I}}\,\Phi_L^{\mathrm{I}}(x_L^{\mathrm{I}}, x_L^{\mathrm{II}}) = 0, \tag{6.2c}$$
$$\xi_G^{\mathrm{II}}\,\Phi_G^{\mathrm{II}}(x_G^{\mathrm{I}}, x_G^{\mathrm{II}}) - \xi_L^{\mathrm{II}}\,\Phi_L^{\mathrm{II}}(x_L^{\mathrm{I}}, x_L^{\mathrm{II}}) = 0, \tag{6.2d}$$
$$\xi_G^{\mathrm{III}}\,\Phi_G^{\mathrm{III}}(x_G^{\mathrm{I}}, x_G^{\mathrm{II}}) - \xi_L^{\mathrm{III}}\,\Phi_L^{\mathrm{III}}(x_L^{\mathrm{I}}, x_L^{\mathrm{II}}) = 0, \tag{6.2e}$$
$$\min(Y,\, 1-\xi_G^{\mathrm{I}} - \xi_G^{\mathrm{II}} - \xi_G^{\mathrm{III}}) = 0, \tag{6.2f}$$
$$\min(1-Y,\, 1-\xi_L^{\mathrm{I}} - \xi_L^{\mathrm{II}} - \xi_L^{\mathrm{III}}) = 0, \tag{6.2g}$$

with the implicit renormalization $x_\alpha^i = \xi_\alpha^i/(\xi_\alpha^{\mathrm{I}} + \xi_\alpha^{\mathrm{II}} + \xi_\alpha^{\mathrm{III}})$ for $\alpha \in \{G, L\}$, $i \in \{\mathrm{I}, \mathrm{II}\}$.

6.2.1 With Henry's law

Phase G is an ideal gas, phase L obeys Henry's law with $k^{\mathrm{I}} = 0.2$, $k^{\mathrm{II}} = 6$, $k^{\mathrm{III}} = 2$. The stopping criterion is $\|\mathbb{F}(\mathbb{X})\| < 10^{-12}$. Starting from the same initial point $(Y,\, \xi_G^{\mathrm{I}},\, \xi_G^{\mathrm{II}},\, \xi_G^{\mathrm{III}}) = (0.9,\, 0.1,\, 0.7,\, 0.1)$ and sweeping over the grid of parameters $\mathcal{C} = \{(c^{\mathrm{I}}, c^{\mathrm{II}}) \in \mathcal{P}^2 \;\mid\; c^{\mathrm{I}} + c^{\mathrm{II}} < 1\}$ where $\mathcal{P} = \{0.01;\, 0.02;\, \ldots;\, 0.99\}$, we display the regime type (single-phase G, single-phase L or two-phase) of the computed solution in Figure 7, highlighting divergence in the red color. In Figure 8a, we display the number of iterations after linearly indexing the elements of $\mathcal{C}$.

In the second test, we sweep over the same grid of parameters and the set of initial points

$$\mathcal{D}^0 = \big\{(Y,\, \xi_G^{\mathrm{I}},\, \xi_G^{\mathrm{II}},\, \xi_G^{\mathrm{III}})^0 \in \mathcal{M}^4 \;\mid\; 1 - (\xi_G^{\mathrm{I}})^0 - (\xi_G^{\mathrm{II}})^0 - (\xi_G^{\mathrm{III}})^0 > 0 \text{ and } 1 - (\xi_G^{\mathrm{I}})^0/k^{\mathrm{I}} - (\xi_G^{\mathrm{II}})^0/k^{\mathrm{II}} - (\xi_G^{\mathrm{III}})^0/k^{\mathrm{III}} > 0\big\},$$

where $\mathcal{M} = \{0.1;\, 0.2;\, \ldots;\, 0.9\}$. The number of initial points used for the tests is $|\mathcal{D}^0| = 252$. For each $\boldsymbol{c} \in \mathcal{C}$, we count the number of initial points for which the method converges and then plot the percentage of success for each algorithm in Figures 8b–8c. Figure 8c demonstrates the efficiency of NPIPM relatively to Newton-min, with 100% of convergence.

6.2.2 With Peng-Robinson's law

The stopping criterion is $\|\mathbb{F}(\mathbb{X})\| < 10^{-10}$ and $\eta = 10^{-4}$ in the last equation of the NPIPM system. We select $(A^{\mathrm{I}},\, B^{\mathrm{I}}) = (0.0883,\, 0.01)$, $(A^{\mathrm{II}},\, B^{\mathrm{II}}) = (0.1861,\, 0.02)$, and $(A^{\mathrm{III}},\, B^{\mathrm{III}}) = (0.2153,\, 0.03)$ so that g_G and g_L are "visually" strictly convex once extended.

In the first test, we use the same initial point

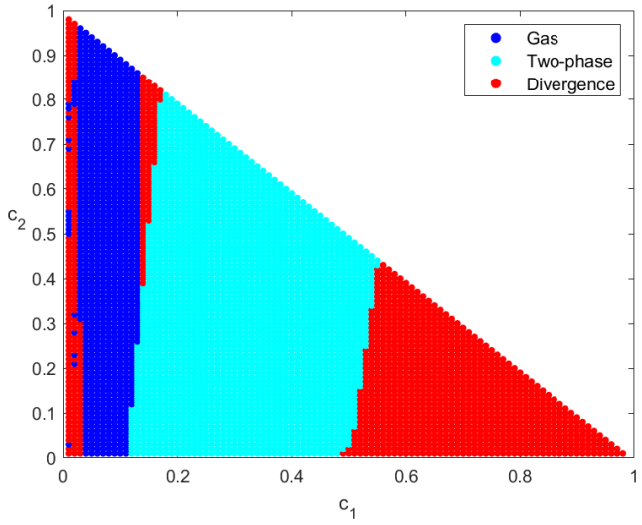
$$(Y,\, \xi_G^{\mathrm{I}},\, \xi_G^{\mathrm{II}},\, \xi_G^{\mathrm{III}},\, \xi_L^{\mathrm{I}},\, \xi_L^{\mathrm{II}},\, \xi_L^{\mathrm{III}})^0 = (0.4,\, 0.3,\, 0.5,\, 0.1,\, 0.325,\, 0.2,\, 0.17)$$

and sweep over the grid of parameters $\mathcal{C} = \{(c^{\mathrm{I}}, c^{\mathrm{II}}) \in \mathcal{P}^2 \;\mid\; c^{\mathrm{I}} + c^{\mathrm{II}} < 1\}$, where $\mathcal{P} = \{0.01;\, 0.02;\, \ldots;\, 0.99\}$. The regime type (single-phase G, single-phase L or two-phase) of the computed solution is represented in Figure 9, where divergence is reported in red. In Figure 10a, we display the number of iterations.

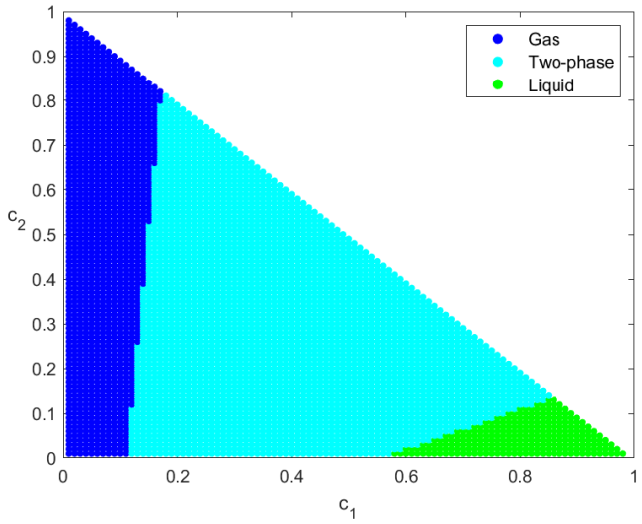
In the second test, we sweep over the grid of parameters $\mathcal{C} = \{\,(c^{\mathrm{I}}, c^{\mathrm{II}}) \in \mathcal{P}^2 \;\mid\; c^{\mathrm{I}} + c^{\mathrm{II}} < 1\,\}$, where $\mathcal{P} = \{0.05;\, 0.10;\, \ldots;\, 0.95\}$, and the set of initial points

$$\mathcal{D}^0 = \big\{(Y,\, \xi_G^{\mathrm{I}},\, \xi_G^{\mathrm{II}},\, \xi_G^{\mathrm{III}},\, \xi_L^{\mathrm{I}},\, \xi_L^{\mathrm{II}},\, \xi_L^{\mathrm{III}})^0 \in \mathcal{M}^7 \;\mid\; 1 - (\xi_G^{\mathrm{I}})^0 - (\xi_G^{\mathrm{II}})^0 - (\xi_G^{\mathrm{III}})^0 > 0 \text{ and } 1 - (\xi_L^{\mathrm{I}})^0 - (\xi_L^{\mathrm{II}})^0 - (\xi_L^{\mathrm{III}})^0 > 0\big\},$$

where $\mathcal{M} = \{0.1;\, 0.2;\, \ldots;\, 0.9\}$. The number of initial points used for the tests is $|\mathcal{D}^0| = 64$. For each $\boldsymbol{c} \in \mathcal{C}$, we plot the percentage of successful initial points for in Figures 10b–10c. Once again, we notice that NPIPM is much more effective than Newton-min.



(a) Newton-min



(b) NPIPM

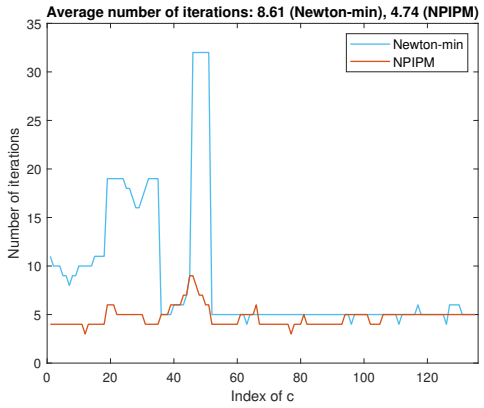
Figure 7: Henry's law: one initial point.

6.3 Evolutionary two-phase binary model

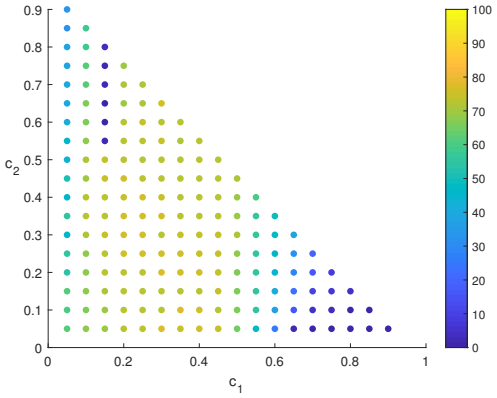
Setting $(k^{\mathrm{I}}, k^{\mathrm{II}}) = (2, 0.5)$, we sweep over the two-dimensional grid of parameters $(c_\flat, \tau) \in \{0.01;\, 0.02;\, \ldots;\, 0.99\} \times \{0.1;\, 0.2;\, \ldots;\, 10\}$ and the set of initial points

$$\mathcal{D}^0 = \left\{ (Y,\, \xi_G^{\mathrm{I}},\, \xi_G^{\mathrm{II}},\, c)^0 \in \mathcal{M}^4 \;\middle|\; 1 - (\xi_G^{\mathrm{I}})^0 - (\xi_G^{\mathrm{II}})^0 > 0 \text{ and } 1 - (\xi_G^{\mathrm{I}})^0/k^{\mathrm{I}} - (\xi_G^{\mathrm{II}})^0/k^{\mathrm{II}} > 0 \right\}$$

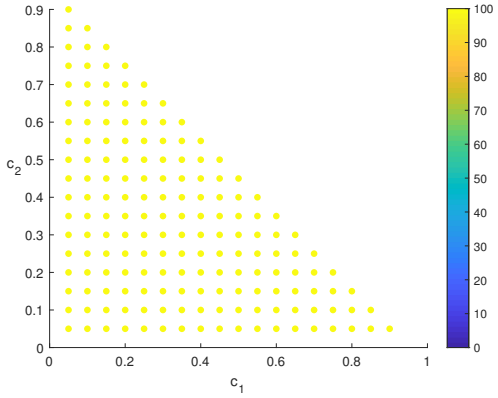
where $\mathcal{M} = \{0.1;\, 0.2;\, \ldots;\, 0.9\}$. The number of initial points used for the test is $|\mathcal{D}^0| = 1944$. For each pair $(c_\flat, \tau)$, we plot the percentage of successful initial points. The results are shown in Figure 11. Although NPIPM no longer claims a 100% rate, especially for large τ, it remains obviously better than Newton-min for $\tau \leqslant 6$ or $c_\flat \leqslant 0.35$.



(a) Number of iterations with the same initial point.



(b) Newton-min: percentage of convergence over all initial points.



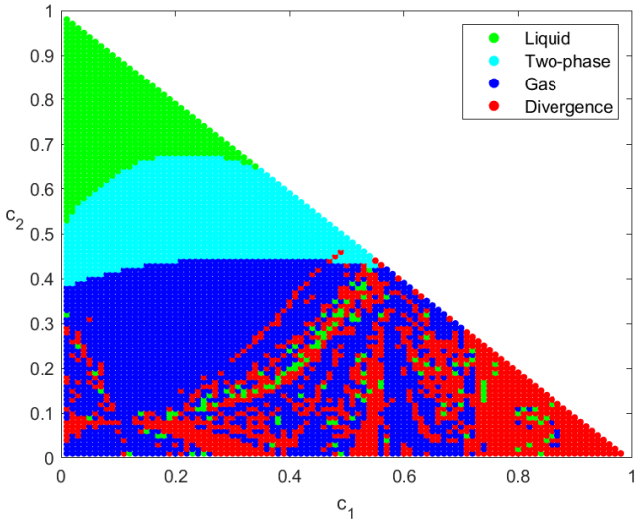
(c) NPIPM: percentage of convergence over all initial points.

Figure 8: Henry's law.

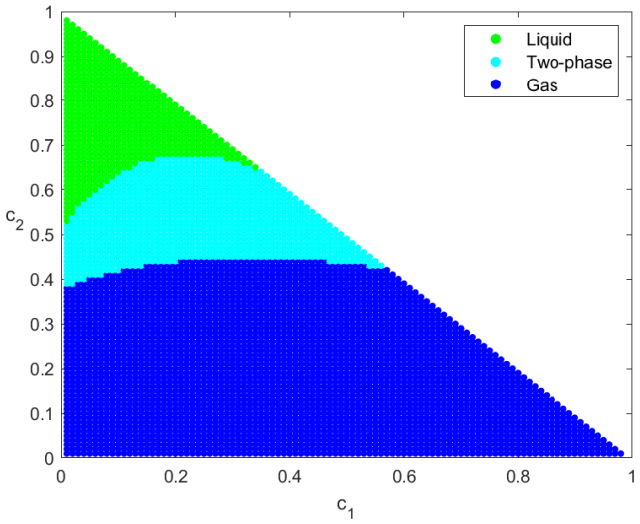
7 Conclusion

In this work, we laid the first stone in the construction of a new method that would be better suited to the resolution of systems containing complementary conditions arising in the unified formulation of thermodynamical equilibria. Numerical results on simple compositional two-phase models revealed an overwhelming superiority of NPIPM over Newton-min. In this sense, they are very promising. The deliberately small size of the systems enabled us to understand in depth the behavior of the algorithms, and most notably the physical obstruction due to Peng-Robinson's law, for which the remedy (4.24)–(4.26) saved us from a compromising situation.

There is still a long way to go. It is of course essential to continue the comparison between NPIPM and Newton-min on the models presented here with a larger number of components, for example a dozen or even a hundred. It is also crucial to try the proposed method on realistic reservoir simulations, in which the system to be solved comes from the finite-volume discretization of the partial differential equations of a porous media flow model. The size of the problem would then be much larger, which could create additional difficulties for the algorithms.



(a) Newton-min



(b) NPIPM

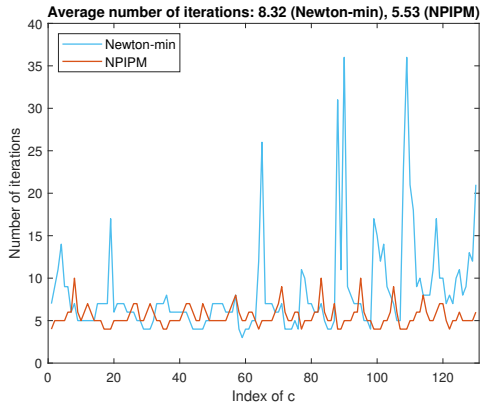
Figure 9: Peng-Robinson's law: one initial point.

References

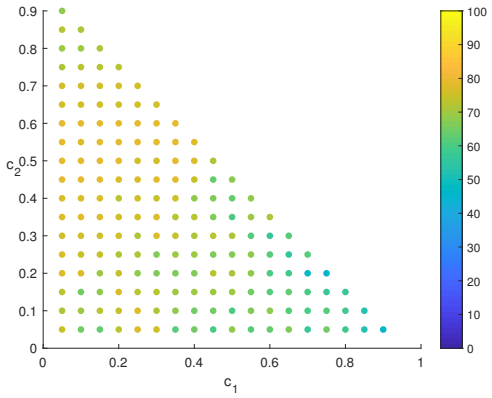
[1] V. Acary and B. Brogliato, *Numerical Methods for Nonsmooth Dynamical Systems: Applications in Mechanics and Electronics*, vol. 35 of Lecture Notes in Applied and Computational Mechanics, Springer, Berlin, 2008.

[2] M. Aganagić, *Newton's method for linear complementarity problems*, Math. Program., 28 (1984), pp. 349–362, https://doi.org/10.1007/BF02612339.

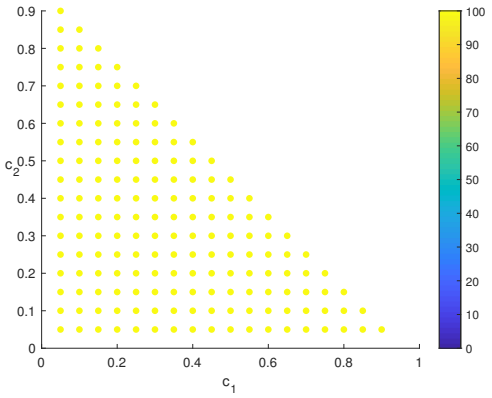
[3] A. Auslender, R. Cominetti, and M. Haddou, *Asymptotic analysis for penalty and barrier methods in convex and linear programming*, Math. Oper. Res., 22 (1997), pp. 43–62, https://doi.org/10.1287/moor.22.1.43.



(a) Number of iterations with the same initial point.



(b) Newton-min: percentage of convergence over all initial points.



(c) NPIPM: percentage of convergence over all initial points.

Figure 10: Peng-Robinson's law.

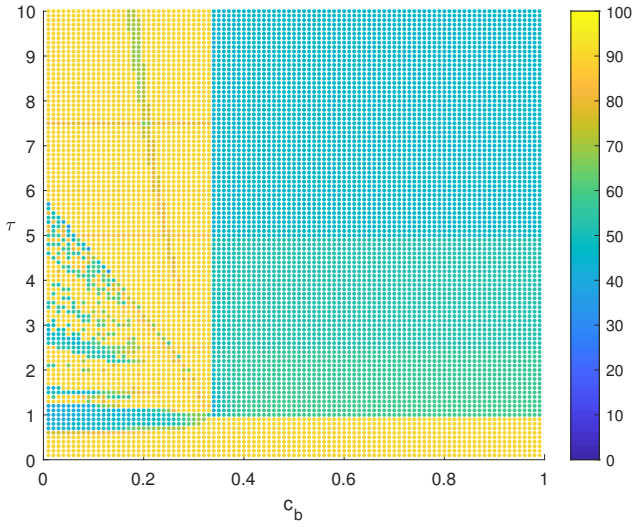
[4] L. Beaude, K. Brenner, S. Lopez, R. Masson, and F. Smai, *Non-isothermal compositional liquid gas Darcy flow: formulation, soil-atmosphere boundary condition and application to high-energy geothermal simulations*, Comput. Geosci., 23 (2019), pp. 443–470, https://doi.org/10.1007/978-3-319-57394-6_34.

[5] A. Beck and M. Teboulle, *Smoothing and first order methods: A unified framework*, SIAM J. Optim., 22 (2012), pp. 557–580, https://doi.org/10.1137/100818327.

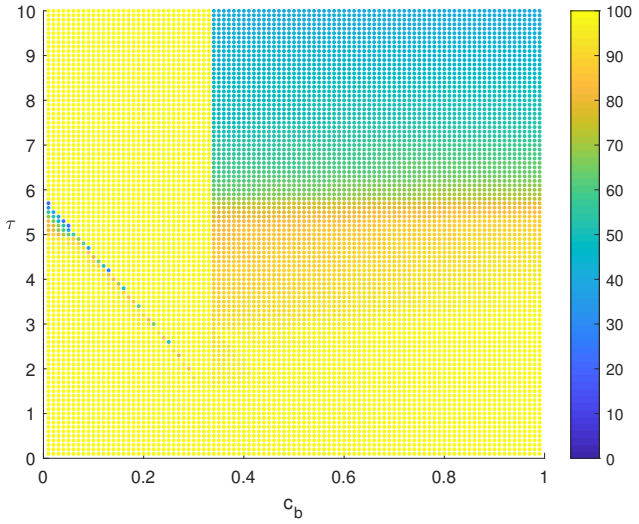
[6] I. Ben Gharbia, *Résolution de problèmes de complémentarité. : Application à un écoulement diphasique dans un milieu poreux*, PhD thesis, Université Paris Dauphine (Paris IX), December 2012, http://tel.archives-ouvertes.fr/tel-00776617.

[7] I. Ben Gharbia and É. Flauraud, *Study of compositional multiphase flow formulation using complementarity conditions*, Oil Gas Sci. Technol., 74 (2019), p. 43, https://doi.org/10.2516/ogst/2019012.

[8] I. Ben Gharbia, É. Flauraud, and A. Michel, *Study of compositional multi-phase flow formulations with cubic EOS*, in SPE Reservoir Simulation Symposium, 23-25 February, Houston, Texas, USA, vol. 2, 01 2015, pp. 1015–1025, https://doi.org/10.2118/173249-MS.



(a) Newton-min



(b) NPIPM

Figure 11: Evolutionary binary model: percentage of convergence over all initial points.

[9] I. Ben Gharbia, M. Haddou, Q. H. Tran, and D. T. S. Vu, *An analysis of the unified formulation for the equilibrium problem of compositional multiphase mixtures*, 2020, https://hal-ifp.archives-ouvertes.fr/hal-03059788. submitted.

[10] I. Ben Gharbia and J. Jaffré, *Gas phase appearance and disappearance as a problem with complementarity constraints*, Math. Comput. Simul., 99 (2014), pp. 28–36, https://doi.org/10.1016/j.matcom.2013.04.021.

[11] F. Bonnans, *Optimisation continue: cours et problèmes corrigés*, Mathématiques appliquées pour le Master, Dunod, 2006.

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

[13] K. H. Coats, *An equation of state compositional model*, SPE Journal, 20 (1980), pp. 363–376, https://doi.org/10.2118/8284-PA.

[14] Ş. Cobzaş, R. Miculescu, and A. Nicolae, *Lipschitz Functions*, vol. 2241 of Lecture Notes in Mathematics, Springer, Cham, Switzerland, 2019, https://doi.org/10.1007/978-3-030-16489-8.

[15] 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.

[16] F. Facchinei and J. S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems, I*, Springer Series in Operations Research, Springer, New York, 2003.

[17] F. Facchinei and J. S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems, II*, Springer Series in Operations Research, Springer, New York, 2003.

[18] A. Fischer, *A special Newton-type optimization method*, Optimization, 24 (1992), pp. 269–284, https://doi.org/10.1080/02331939208843795.

[19] M. Haddou, *A new class of smoothing methods for mathematical programs with equilibrium constraints*, Pacif. J. Optim., 5 (2009), pp. 86–96.

[20] M. Haddou and P. Maheux, *Smoothing methods for nonlinear complementarity problems*, J. Optim. Theory Appl., 160 (2014), pp. 711–729, https://doi.org/10.1007/s10957-013-0398-1.

[21] A. F. Izmailov and M. V. Solodov, *Newton-Type Methods for Optimization and Variational Problems*, Springer Series in Operations Research and Financial Engineering, Springer, Cham, Switzerland, 2014, https://doi.org/10.1007/978-3-319-04247-3.

[22] S. Kräutle, *The semismooth Newton method for multicomponent reactive transport with minerals*, Adv. Water Res., 34 (2011), pp. 137–151, https://doi.org/10.1016/j.advwatres.2010.10.004.

[23] 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), pp. 957–966, https://doi.org/10.1016/j.advwatres.2011.04.021.

[24] I. Lusetti, *Numerical methods for compositional multiphase flow models with cubic EOS*, tech. report, IFPEN, 2016.

[25] O. L. Mangasarian, *Equivalence of the complementarity problem to a system of nonlinear equations*, SIAM J. Appl. Math., 31 (1976), pp. 89–92, https://doi.org/10.1137/0131009.

[26] 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), pp. 1–34, http://ijfv.math.cnrs.fr/IMG/pdf/gazliqcomp-ijfv-1.pdf.

[27] 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), pp. 708–728, https://doi.org/10.1016/j.jcp.2016.06.003.

[28] S. Mehrotra, *On the implementation of a primal-dual interior point method*, SIAM J. Optim., 2 (1992), pp. 575–601, https://doi.org/10.1137/0802028.

[29] R. Mifflin, *Semismooth and semiconvex functions in constrained optimization*, SIAM J. Control Optim., 15 (1977), pp. 959–972, https://doi.org/10.1137/0315061.

[30] 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.

[31] N. Peton, *Comparaison de plusieurs formulations pour les écoulements multiphasiques et compositionnels en milieu poreux*, tech. report, IFPEN, 2015.

[32] L. Qi and J. Sun, *A nonsmooth version of Newton's method*, Math. Program., 58 (1993), pp. 353–367, https://doi.org//10.1007/BF01581275.

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

[34] M. H. Wright, *The interior-point revolution in optimization: history, recent developments, and lasting consequences*, Bull. Amer. Math. Soc., 42 (2005), pp. 39–56, https://doi.org/10.1090/S0273-0979-04-01040-7.

[35] S. J. Wright, *Primal-Dual Interior-Point Methods*, SIAM, Philadelphia, 1997.