Reactive flash for ideal multiphase mixtures: Unified formulation and efficient computation

Maxime Jonval, Ibtihel Ben Gharbia, Clément Cancès, Thibault Faney, Quang Huy Tran

Reactive flash for ideal multiphase mixtures: Unified formulation and efficient computation

Maxime Jonval* Ibtihel Ben Gharbia† Clément Cancès‡ Thibault Faney†
Quang-Huy Tran†

Abstract

Multiphase chemical equilibrium problems lead to nonlinear systems with complementarity constraints, which become particularly challenging when phases may vanish. We introduce a new algebraic formulation of the equilibrium problem based on *extended mole fractions*, derived from the subdifferential of the Gibbs free energy, and establish its equivalence with the classical minimization problem. Our analysis provides new conditions ensuring the uniqueness of solutions, even when some phases disappear. Building on this formulation, we propose two parametrized Newton-based strategies: one reformulates the relation between species quantities and chemical potentials, while the other parametrizes the complementarity conditions directly. Numerical experiments on a system with 72 species and 22 phases confirm the robustness and efficiency of the proposed methods. In tests with randomized inputs, both strategies achieve success rates above 90% with moderate iteration counts, outperforming established approaches such as the Newton–min and Fischer–Burmeister complementarity functions, and interior-point methods.

1 Introduction

Simulating reactive transport remains a central challenge in many areas of science and engineering, including porous media flow, combustion, and chemical reactor design. In porous media in particular, accurate modeling of reactive transport is essential for applications such as CO_2 and H_2 storage and geothermal energy production. A key computational bottleneck in these simulations lies in solving nonlinear chemical equilibrium equations, which must be evaluated at each time step and in every mesh cell. Even modest gains in the robustness and efficiency of equilibrium solvers can therefore yield significant improvements in overall simulation performance.

Newton's method is the standard approach for solving chemical equilibrium equations, but it faces significant challenges in multiphase settings. In earlier work [18], we investigated the single-phase case and showed that the parametrization technique proposed by Brenner and Cancès [6] can enhance robustness. In the present article, we extend this approach to the more complex case of multiphase systems.

Unlike single-phase systems, multiphase equilibrium problems are complicated by the nonnegativity constraints on species quantities, which allow certain phases to vanish at equilibrium. As a result, the set of present phases is not known a priori. Two main families of methods have been developed to address this issue. Combinatorial approaches compute equilibria for all possible subsets of phases [21, 9], but become computationally prohibitive as the number of candidate phases grows, due to the exponential increase in subsets to evaluate. Unified formulations, by contrast, introduce extended mole fractions and complementarity constraints to account for all phases simultaneously [19, 4]. By treating all phases within a single framework, unified formulations avoid combinatorial explosion and provide a more scalable approach for systems with many potential phases.

The concept of extended mole fractions was first introduced by Lauser *et al.* [19] in the context of multiphase flow in porous media, where complementarity constraints were used to handle phase appearance and disappearance within a unified system of equations. This formulation was later applied to non-reactive phase equilibrium problems by Vu *et al.* [4], who provided a detailed mathematical analysis.

*Inria, Univ. Côte d'Azur, CNRS, UMR 7351 – Laboratoire J.A. Dieudonné, 06108 Nice, France. maxime.jonval@inria.fr

†IFP Energies nouvelles, 1 et 4 avenue de Bois Préau, 92852 Rueil-Malmaison Cedex, France. ibtihel.ben-gharbia@ifpen.fr, thibault.faney@ifpen.fr, quang-huy.tran@ifpen.fr

‡Univ. Lille, CNRS, Inria, UMR 8524 – Laboratoire Paul Painlevé, 59000 Lille, France. clement.cances@inria.fr

Coatléven and Michel [8] extended this approach to multiphase chemical equilibria by embedding the optimality criterion of Smith *et al.* [27] into the Karush–Kuhn–Tucker (KKT) conditions.

An alternative strategy is the species-quantity formulation proposed by Leal *et al.* [20], in which species quantities serve as the primary variables and a complementarity equation is written for each species. However, as observed by Shapiro and Shapley [26], if a species is present in a given phase, then the entire phase must also be present. This observation supports the formulation of complementarity conditions at the phase level rather than at the species level. More broadly, complementarity problems have been extensively studied, both theoretically and numerically, and provide a robust framework for modeling phase transitions and their interactions [5, 3, 2, 11].

Our contributions are threefold. First, we establish a new condition guaranteeing uniqueness even when phases vanish. Second, we derive a unified Gibbs-based formulation of multiphase chemical equilibrium and prove its equivalence with the classical minimization problem. Third, we develop two parametrized Newton strategies, including a novel complementarity parametrization method, which achieves both robustness and efficiency improvements. Numerical experiments confirm the effectiveness of these strategies across systems of up to 72 species and 22 phases.

Section 2 introduces the chemical system, defines the mass conservation laws and the Gibbs free energy, and formulates the equilibrium problem as a constrained minimization problem.

In Section 3, we extend classical results on existence and uniqueness of minimizers from Shapiro and Shapley [26], who showed that uniqueness is guaranteed when all phases are present, but not necessarily when some phases vanish. We generalize the notion of azeotropy to propose a new condition that ensures uniqueness even in the presence of phase disappearance, thereby strengthening the analytical foundation of the model. We then derive a unified formulation based on extended mole fractions and establish the uniqueness of its solution under reasonable assumptions. This formulation involves one complementarity problem per phase, which requires an adapted approach for its numerical resolution.

Section 4 is devoted to the numerical resolution of the unified formulation using Newton's method. We employ the parametrization techniques introduced in [18] to reformulate the relation between extended mole fractions and chemical potentials. In addition, we propose a novel method for addressing the complementarity problem, the complementarity parametrization technique, which parametrizes the relationship between the complementary variables.

Finally, Section 5 presents numerical experiments validating our proposed methods. We begin with two simple test cases, followed by a more demanding benchmark involving 22 phases. This benchmark is designed to evaluate the robustness and performance of the algorithm under complex conditions. The results demonstrate the effectiveness of our numerical strategy, in particular the complementarity parametrization approach, in resolving phase presence conditions both accurately and efficiently.

2 Mathematical formulation for multiphase chemical equilibria

A chemical species is a group of atoms characterized by its molecular formula and the phase to which it belongs. Typical phases include aqueous (aq) solutions such as diluted electrolytes or oils, gaseous (g) phases, and solid (s) phases such as pure minerals. For instance $H_2O(aq)$, $H^+(aq)$, $NaCl(aq)$, $CO_2(g)$, $H_2O(g)$, $CaCO_3(s)$ are all distinct species. Note that $H_2O(aq)$ and $H_2O(g)$ are considered different species, since they belong to different phases. Similarly, $SiO_2(s)$ can occur as quartz or amorphous silica, which correspond to two distinct solid phases, denoted $SiO_2(quartz)$ and $SiO_2(amorphous - silica)$.

For a given temperature T and pressure P, a chemical system is defined as

$$\mathcal{S}_{\mathrm{P,T}} = \{\mathcal{P}, \mathcal{E}, \mathcal{C}, \mathcal{R}\}$$

where

- $\mathcal{P} = \{P_1, \ldots, P_{N_{\mathrm{ph}}}\}$ is the set of N_{ph} phases;
- $\mathcal{E} = \{E_1, \ldots, E_M\}$ is the set of M chemical elements (typically atoms);
- $\mathcal{C} = \{C_1, \ldots, C_N\}$ is the set of N chemical species, with $N > M$;
- $\mathcal{R} = \{R_1, \ldots, R_{N-M}\}$ is the set of $N - M$ chemical reactions among these species.

The set $\mathcal{E}$ contains all the elements that compose the species of the set $\mathcal{C}$, while the reactions in $\mathcal{R}$ describe how these species interact. A chemical reaction $R_j \in \mathcal{R}$ can be written as

$$\sum_{i=1}^{N} s_{ij} C_i = 0,$$

where the s_{ij} are the stoichiometric coefficients indicating the number of molecules of C_i involved in reaction R_j.

The set $\mathcal{P}$ contains the phases of the system. To link each species C_i to its phase α, we introduce the mapping

$$\sigma : i \in \{1, \ldots, N\} \mapsto \alpha \in \{1, \ldots, N_{Ph}\} \tag{1}$$

so that $\sigma^{-1}(\alpha)$ is the set of all species in phase α. We denote by $N^\alpha = \#\sigma^{-1}(\alpha)$ the number of species in phase α.

2.1 Mass conservation

The systems studied in this article are closed, meaning that no exchange of matter occurs with the surroundings. In such systems, the principle of mass conservation ensures that the total amount $\mathbf{b} = (\mathrm{b}_1, \ldots, \mathrm{b}_M)$ of each element of $\mathcal{E}$ remains constant. Each species C_i in the set of species $\mathcal{C}$ has an associated *formula vector* $\mathbf{a}_i$ in the basis of elements $\mathcal{E}$. For example, if $\mathcal{E}$=(H, C, O) and C_i =$\mathrm{HCO_3^-}$, then $\mathbf{a}_i = (1, 1, 3)^T$. Based on these vectors, the set of species $\mathcal{C}$ is partitioned into two subsets:

- Primary species $\mathcal{C}_{Pr} = \{C_1, \ldots, C_M\}$, whose formula vectors are linearly independent. They form the primary basis of the system. The number of primary species is equal to the number of elements M.
- Secondary species $\mathcal{C}_{Sd} = \{C_{M+1}, \ldots, C_N\}$, whose formula vectors are linear combinations of the primary species. Their number is $N - M$, corresponding to the $N - M$ reactions in the set $\mathcal{R}$.

The choice of the primary species is not unique. For clarity, we order the species so that primary species come first, followed by secondary species. This leads to the *formula matrix* $\mathbf{A} = [\mathbf{A}_{Pr}, \mathbf{A}_{Sd}]$ where $\mathbf{A}_{Pr}$ is an $M \times M$ invertible matrix formed from the primary species formula vectors, and $\mathbf{A}_{Sd}$ is an $M \times (N-M)$ matrix formed from the secondary species formula vectors.

Example. Consider a system consisting of one aqueous phase and one pure quartz mineral:

$$\begin{aligned}
\mathcal{P} &= \{\text{aqueous, quartz}\} \\
\mathcal{E} &= \{\mathrm{H, O, Si}\} \\
\mathcal{C} &= \{\mathrm{H_2O}, \mathrm{H^+}, \mathrm{OH^-}, \mathrm{SiO_2(aq)}, \mathrm{SiO_2(quartz)}\} \\
\mathcal{R} &= \{\mathrm{H_2O = H^+ + OH^-}, \mathrm{SiO_2(aq) = SiO_2(quartz)}\}
\end{aligned}$$

If we choose the primary species $\mathcal{C}_{Pr} = \{\mathrm{H^+}, \mathrm{OH^-}, \mathrm{SiO_2(aq)}\}$ and the secondary species $\mathcal{C}_{Sd} = \{\mathrm{H_2O}, \mathrm{SiO_2(quartz)}\}$, the corresponding formula matrix is

$$\mathbf{A} = \begin{array}{c} \begin{matrix} \mathrm{H^+} & \mathrm{OH^-} & \mathrm{SiO_2(aq)} & \mathrm{H_2O} & \mathrm{SiO_2(quartz)} \end{matrix} \\ \begin{bmatrix} 1 & 1 & 0 & 2 & 0 \\ 0 & 1 & 2 & 1 & 2 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix} \end{array} \begin{matrix} \mathrm{H} \\ \mathrm{O} \\ \mathrm{Si} \end{matrix}.$$

Let $\mathbf{n} = (n_1, \ldots, n_N)$ denote the molar quantities of the species in $\mathcal{C}$. The conservation of elements is expressed as

$$\mathbf{A}\mathbf{n} = \mathbf{b}. \tag{2}$$

Since we study multiphase equilibria, it is convenient to partition $\mathbf{n}$ and $\mathbf{A}$ into sub-vectors $\mathbf{n}^\alpha$ and sub-matrices $\mathbf{A}^\alpha$, each corresponding to the species belonging to a given phase α. In the example above, $\mathbf{n}^{\mathrm{aq}} = (n_{\mathrm{H^+}}, n_{\mathrm{OH^-}}, n_{\mathrm{SiO_2(aq)}}, n_{\mathrm{H_2O}})$, $\mathbf{n}^{quartz} = (n_{\mathrm{SiO_2(quartz)}})$,

$$\mathbf{A}^{\mathrm{aq}} = \begin{bmatrix} 1 & 1 & 0 & 2 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{A}^{\mathrm{quartz}} = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}.$$

Equation (2) then becomes

$$\sum_{\alpha=1}^{N_{Ph}} \mathbf{A}^{\alpha}\mathbf{n}^{\alpha} = \mathbf{b}.$$

We define the *stoichiometry matrix*

$$\mathbf{S} = \begin{bmatrix} \mathbf{A}_{Pr}^{-1}\mathbf{A}_{Sd} \\ -\mathbf{I}_{Sd} \end{bmatrix}. \tag{3}$$

By construction, $\mathbf{AS} = \mathbf{0}$ and $\operatorname{rank}\mathbf{S} = N - M$. The entries of $\mathbf{S}$ coincide with the stoichiometric coefficients of the reactions in $\mathcal{R}$. For the example system,

$$\mathbf{S} = \begin{array}{c} \begin{matrix} \mathrm{R}_1 & \mathrm{R}_2 \end{matrix} \\ \left[\begin{array}{cc} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ -1 & 0 \\ 0 & -1 \end{array}\right] \end{array} \begin{matrix} \mathrm{H}^+ \\ \mathrm{OH}^- \\ \mathrm{SiO_2(aq)} \\ \mathrm{H_2O} \\ \mathrm{SiO_2(quartz)} \end{matrix}.$$

The fundamental link between $\mathbf{A}$ and $\mathbf{S}$ is

$$\ker \mathbf{S}^T = (\ker \mathbf{A})^{\perp} = \operatorname{Im} \mathbf{A}^T. \tag{4}$$

Moreover, the components of an element in $\ker \mathbf{A} \setminus \{\mathbf{0}\}$ do not all have the same sign. In particular

$$\ker \mathbf{A} \cap \mathbb{R}_+^N = \{\mathbf{0}\}. \tag{5}$$

This means that no strictly positive combination of species can vanish under the conservation laws. Proofs of these properties can be found in [18], and further details on relationship between $\mathbf{A}$ and $\mathbf{S}$ can be found in the monograph of Smith and Missen [28].

2.2 Gibbs free energy and chemical potentials

The state of a closed system $\mathcal{S}_{\mathrm{P},\,\mathrm{T}}$ at fixed pressure P and temperature T is characterized by the Gibbs free energy function $G : \mathbb{R}_+^N \to \mathbb{R}$.

The Gibbs free energy function decomposes into phase-specific functions $G_\alpha : \mathbb{R}_+^{N^\alpha} \to \mathbb{R}$ that depend solely on the species in the phase α. Each G_α is extensive, i.e. homogeneous of degree one in the molar quantities, and therefore by Euler's theorem on homogeneous functions admits the representation:

$$G(\mathbf{n}) = \sum_{\alpha=1}^{N_{Ph}} G_\alpha(\mathbf{n}^\alpha) \quad \text{where} \quad G_\alpha(\mathbf{n}^\alpha) = \sum_{i \in \sigma^{-1}(\alpha)} n_i \frac{\partial G_\alpha(\mathbf{n}^\alpha)}{\partial n_i}. \tag{6}$$

The chemical potential of the species C_i is defined as $\mu_i(\mathbf{n}^\alpha) = \partial G_\alpha(\mathbf{n}^\alpha)/\partial n_i = \partial G(\mathbf{n})/\partial n_i$. In practice, chemical potentials are commonly expressed as

$$\mu_i = \mu_i^\circ(\mathrm{P},\,\mathrm{T}) + \mathrm{RT} \ln a_i(\mathbf{n}^\alpha). \tag{7}$$

where $\mu_i^\circ(\mathrm{P},\,\mathrm{T})$ is the chemical potential of species C_i in phase α in its standard state at pressure P and temperature T, to be computed from thermodynamic tables, and a_i is the activity of species C_i that depends in general on the concentration of all the species in phase α.

The activity of a species C_i is written as $a_i = \gamma_i x_i$, where γ_i is the activity coefficient, accounting for non-ideal interactions, and x_i denotes the mole fraction of C_i in phase α, defined by

$$x_i := n_i / \textstyle\sum_{j \in \sigma^{-1}(\alpha)} n_j = n_i / \langle \mathbf{n}^\alpha,\, \mathbf{1} \rangle, \quad \text{where} \quad \mathbf{1} := (1,\,\ldots,\,1)^T.$$

Numerous, increasingly complex activity models for γ_i are available in the literature (see *e.g.* [22, 31]). The simplest is the ideal activity model in which $\gamma_i = 1$. This model corresponds to an ideal mixture in which all intermolecular interactions are equivalent. In this case, the activity reduces to the mole fraction. The resulting ideal Gibbs energy in (6) is convex on $\mathbb{R}_+^N$ (see [26, Theorem 8.13]). We restrict our attention to this case in the remainder of the article.

2.3 Multiphase chemical equilibrium as a minimization problem

At fixed temperature T, pressure P, and element quantities $\mathbf{b}$, the equilibrium state of the system is characterized by the minimization of the Gibbs free energy G. In other words, the vector of species quantities $\mathbf{n}$ is obtained by solving a constrained optimization problem:

$$\mathbf{n} \in \arg\min \{G(\mathbf{n}) \mid \mathbf{A}\mathbf{n} = \mathbf{b} \,\text{and}\, \mathbf{n} \geq \mathbf{0}\}. \tag{8}$$

This formulation reflects two fundamental physical principles: conservation of elements, encoded by the linear constraints $An = b$, and non-negativity of molar quantities. In the ideal activity model, G is convex on $\mathbb{R}^N_+$, which ensures that (8) is a convex optimization problem. This classical formulation serves as the foundation for the existence and uniqueness analysis developed in Section 3.

3 On the minimizers of the Gibbs free energy

The current section gathers mathematical results on the optimization problem (8). We first establish in Section 3.1 a well-posedness result for the problem in close connection to the seminal work by Shapiro and Shapley [26]. Section 3.2 is then devoted to the thorough mathematical derivation of a unified formulation in the spirit of [19]. In such a formulation, a suitable notion of extended mole fractions is introduced for all the phases, including the absent ones. These persistent variables can then be used to solve the nonlinear system in an efficient way (see Section 4). We comment in Section 3.3 on the uniqueness of the solutions of the unified formulation. The loss of uniqueness due to azeotropy is described, as well as a possible lack of uniqueness for the extended mole fractions of the absent phases. The latter difficulty can however only happen for very specific right-hand sides, and uniqueness can be established as long as the two aforementioned identified difficulties do not occur.

3.1 Existence and uniqueness of the minimizers

The existence and uniqueness of a minimizer for an ideal multiphase chemical equilibrium have been rigorously examined by Shapiro and Shapley in [26]. In their fundamental contribution, they provide a comprehensive characterization of the minimizers set for the multiphase chemical equilibrium problem (8) in Theorem 9.2. Furthermore, they prove in Theorem 12.1 that if the configuration of the chemical system necessitates the presence of all phases at equilibrium, then the minimizer is unique. In cases where one or more phases are absent, uniqueness does not hold for the overall solution. However, Shapiro and Shapley demonstrate in Theorem 12.4 that the mole fractions of the present phases are uniquely defined. A last important result is Theorem 9.8, which states that if a species is present in a phase, then the entire phase must also be present.

In this section, we extend the work of Shapiro and Shapley by presenting a novel condition that ensures the uniqueness of a minimizer, even in cases where certain phases may be absent. This condition is derived from their characterization of the set of minimizers and provides a more comprehensive analysis framework.

Let $\mathbb{R}^N_{>0}$ and $\mathbb{R}^N_{\geq 0}$ be the set of positive and nonnegative vectors of $\mathbb{R}^N$, respectively. We define the set of vectors satisfying the constraints of conservation of elements by

$$\mathcal{M}_{\mathbf{A},\mathbf{b}} := \{\mathbf{n} \in \mathbb{R}^N \mid \mathbf{A}\mathbf{n} = \mathbf{b}\}.$$

We then make the following assumption:

(H1) $\mathcal{M}_{\mathbf{A},\mathbf{b}} \cap \mathbb{R}^N_{>0}$ is nonempty.

This assumption corresponds to the existence of a strictly feasible point, which is known as the Slater condition in optimization. This assumption is physically reasonable, as it only requires that the element amounts $\mathbf{b}$ do not a priori exclude any species from being present at equilibrium. Shapiro and Shapley have proved in [26, Theorem 9.2] that if the minimizers set for the multiphase chemical equilibrium problem (8) is nonempty, which is the case thanks to (H1) and since $\mathcal{M}_{\mathbf{A},\mathbf{b}} \cap \mathbb{R}^N_{\geq 0}$ is compact as shown in [18], then there exists a minimizer $\tilde{\mathbf{n}}$ such that:

$$\arg\min \left\{G(\mathbf{m}) \;\middle|\; \mathbf{m} \in \mathcal{M}_{\mathbf{A},\mathbf{b}} \cap \mathbb{R}^N_{\geq 0}\right\} = \mathcal{C}(\tilde{\mathbf{n}}) \cap \mathcal{M}_{\mathbf{A},\mathbf{b}} \cap \mathbb{R}^N_{\geq 0}, \tag{9}$$

where

$$\mathcal{C}(\tilde{\mathbf{n}}) = \left\{\mathbf{m} \in \mathbb{R}^N \;\middle|\; \mathbf{m}^\alpha = \lambda_\alpha \tilde{\mathbf{n}}^\alpha,\, \lambda_\alpha \in \mathbb{R},\, \forall\, \alpha = 1, \ldots, N_{Ph}\right\}. \tag{10}$$

It follows from the definition of $\mathcal{C}(\tilde{\mathbf{n}})$ that if $\tilde{\mathbf{n}}^\alpha = \mathbf{0}$ then $\mathbf{m}^\alpha = \mathbf{0}$, for each $\mathbf{m}$ belonging to $\mathcal{C}(\tilde{\mathbf{n}}) \cap \mathcal{M}_{\mathbf{A},\mathbf{b}} \cap \mathbb{R}^N_{\geq 0}$. We introduce the set of present phases in $\tilde{\mathbf{n}}$, sometimes referred to as the *context* in the literature, by

$$\Gamma_{\tilde{\mathbf{n}}} := \{\alpha \in \{1, \ldots, N_{Ph}\} \mid \tilde{\mathbf{n}}^\alpha \neq \mathbf{0}\}.$$

We can then establish the necessary and sufficient condition for the uniqueness of $\tilde{\mathbf{n}}$ in the subsequent theorem.

Theorem 3.1. *Let $\tilde{\mathbf{n}}$ the minimizer defined in* (9) *and* (10) *and let $\mathbf{A}^\alpha$ be the $M \times N^\alpha$ matrix corresponding to the part of $\mathbf{A}$ related to the phase α, then the minimizer $\tilde{\mathbf{n}}$ is unique if and only if the set of vectors $\{\mathbf{A}^\alpha \tilde{\mathbf{n}}^\alpha\}_{\alpha \in \Gamma_{\tilde{\mathbf{n}}}}$ is linearly independent.*

Proof. Let $\mathbf{m} \in \mathcal{C}(\tilde{\mathbf{n}}) \cap \mathcal{M}_{\mathbf{A},\mathbf{b}} \cap \mathbb{R}^N_{\geq 0}$ with $\mathbf{m} \neq \tilde{\mathbf{n}}$. Since both $\tilde{\mathbf{n}}$ and $\mathbf{m}$ belong to $\in \mathcal{M}_{\mathbf{A},\mathbf{b}} \cap \mathbb{R}^N_{\geq 0}$, we have

$$\mathbf{A}(\tilde{\mathbf{n}} - \mathbf{m}) = \mathbf{0} \quad \Leftrightarrow \quad \sum_{\alpha \in \Gamma_{\tilde{\mathbf{n}}}} \mathbf{A}^\alpha (\tilde{\mathbf{n}}^\alpha - \mathbf{m}^\alpha) = \mathbf{0}.$$

Since $\mathbf{m} \in \mathcal{C}(\tilde{\mathbf{n}}) \cap \mathbb{R}^N_{\geq 0}$, there exists $\lambda_\alpha \geq 0$ such that $\mathbf{m}^\alpha = \lambda_\alpha \tilde{\mathbf{n}}^\alpha$, $\forall \alpha = 1, \ldots, N_{Ph}$. Therefore

$$\sum_{\alpha \in \Gamma_{\tilde{\mathbf{n}}}} (1 - \lambda_\alpha) \mathbf{A}^\alpha \tilde{\mathbf{n}}^\alpha = \mathbf{0},$$

with $\{1 - \lambda_\alpha\}_{\alpha \in \Gamma_{\tilde{\mathbf{n}}}}$ not all zero, meaning that $\{\mathbf{A}^\alpha \tilde{\mathbf{n}}^\alpha\}_{\alpha \in \Gamma_{\tilde{\mathbf{n}}}}$ is not linearly independent.

Conversely, assume that $\{\mathbf{A}^\alpha \tilde{\mathbf{n}}^\alpha\}_{\alpha \in \Gamma_{\tilde{\mathbf{n}}}}$ is not linearly independent, then there exists a set of non-zero scalars $\{\gamma_\alpha\}_{\alpha \in \Gamma_{\tilde{\mathbf{n}}}}$ such that

$$\sum_{\alpha \in \Gamma_{\tilde{\mathbf{n}}}} \gamma_\alpha \mathbf{A}^\alpha \tilde{\mathbf{n}}^\alpha = \mathbf{0}.$$

Since $\tilde{\mathbf{n}} \geq \mathbf{0}$ and from (5), there exists $\beta \in \Gamma_{\tilde{\mathbf{n}}}$ such that $\max\{\gamma_\alpha\}_{\alpha \in \Gamma_{\tilde{\mathbf{n}}}} =: \gamma_\beta > 0$. We can define a set of non-negative scalars $\{\lambda_\alpha\}_{\alpha \in \Gamma_{\tilde{\mathbf{n}}}}$ as $\lambda_\alpha = 1 - \gamma_\alpha / \gamma_\beta$. Therefore one defines the vector $\mathbf{m}$ such that $\mathbf{m}^\alpha = \lambda^\alpha \tilde{\mathbf{n}}^\alpha$ if $\alpha \in \Gamma_{\tilde{\mathbf{n}}}$ and $\mathbf{m}^\alpha = \mathbf{0}$ otherwise. By construction, $\mathbf{m} \in \mathcal{C}(\tilde{\mathbf{n}}) \cap \mathcal{M}_{\mathbf{A},\mathbf{b}} \cap \mathbb{R}^N_{\geq 0}$ and $\mathbf{m} \neq \tilde{\mathbf{n}}$. □

Theorem 3.1 extends the work of Shapiro and Shapley by guaranteeing the uniqueness of the total quantities in each phase, rather than merely the uniqueness of mole fractions. Specifically, it addresses a well-known issue of non-uniqueness related to azeotropic compositions. In such mixtures, the liquid and gaseous phases maintain the same composition at a fixed temperature corresponding to a constant boiling point, making it impossible to separate the components through simple distillation. This characteristic renders the two phases indistinguishable.

The notion of azeotropic mixture can be extended to our setting: we say that a minimizer $\mathbf{n}$ of (8) has an azeotropic composition if $\{\mathbf{A}^\alpha \mathbf{n}^\alpha\}_{\alpha \in \Gamma_{\mathbf{n}}}$ is linearly dependent, i.e. if

$$\text{there exists } (\lambda_\alpha)_{\alpha \in \Gamma_{\mathbf{n}}} \neq \mathbf{0} \text{ such that } \sum_{\alpha \in \Gamma_{\mathbf{n}}} \lambda_\alpha \mathbf{A}^\alpha \mathbf{n}^\alpha = \mathbf{0}. \tag{11}$$

We will further comment on the lack of uniqueness when (11) is satisfied in Section 3.3 later on.

Moreover, it serves as a clear example of the equality in mole fractions of two minimizers. The condition of Theorem 3.1 ensures that the total quantities of each phase in such cases are the same, thereby making the solution unique.

3.2 Unified formulation for multiphase chemical equilibria

In this section, we present our unified formulation for solving the minimization problem (8) using the concept of the subdifferential of a modified Gibbs energy that incorporates the non-negativity constraints. We establish the equivalence between this new set of equations and the original minimization problem, along with the uniqueness of the solution.

This approach also shows that the formulation of Coatléven and Michel [8] can be recovered directly from the KKT optimality conditions, thereby providing a rigorous theoretical basis for their method. Our formulation, however, adopts a different set of primary variables, offering an alternative perspective for solving the equilibrium equations.

This new formulation is presented in the following proposition.

Proposition 3.1. *Assume that hypothesis (H1) holds true. The vector* $\mathbf{n} \in \mathbb{R}^N$ *is a minimizer for the problem* (8) *if and only if there exists a set of vectors* $(\boldsymbol{\xi}^\alpha,\, s_\alpha,\, r_\alpha)_{\alpha=1,\,\ldots,\,N_{Ph}}$ *such that* $n_i = s_{\sigma(i)}\xi_i$, $\xi_i > 0$, *for all* $i = 1,\, \ldots,\, N$, *and satisfying:*

$$\sum_{\alpha=1}^{N_{Ph}} s_\alpha \mathbf{A}^\alpha \boldsymbol{\xi}^\alpha - \mathbf{b} = 0, \tag{12a}$$

$$\mathbf{S}^T[\mu_i^\circ + \mathrm{RT}\ln \xi_i]_{i=1,\,\ldots,\,N} = \mathbf{0}, \tag{12b}$$

$$\langle \boldsymbol{\xi}^\alpha,\, \mathbf{1}\rangle + r_\alpha - 1 = 0, \qquad (\alpha = 1,\, \ldots,\, N_{Ph}), \tag{12c}$$

$$s_\alpha r_\alpha = 0, \qquad (\alpha = 1,\, \ldots,\, N_{Ph}), \tag{12d}$$

$$s_\alpha \geq 0,\, r_\alpha \geq 0, \qquad (\alpha = 1,\, \ldots,\, N_{Ph}). \tag{12e}$$

In the formulation (12), $\boldsymbol{\xi}^\alpha$ are the extended mole fractions, while s_α represents the total quantities of phase α. Consequently, if a phase is present at equilibrium, we have $s_\alpha > 0$, and $\boldsymbol{\xi}^\alpha$ corresponds to the mole fractions:

$$\xi_i = \frac{n_i}{s_\alpha} = x_i(\mathbf{n}^\alpha).$$

Conversely, if a phase is absent, $r_\alpha \geq 0$, and the vector $\boldsymbol{\xi}^\alpha$ does not necessarily sum to one:

$$\langle \boldsymbol{\xi}^\alpha,\, \mathbf{1}\rangle \leq 1.$$

In this context, $\boldsymbol{\xi}^\alpha$ serves as an extension of the mole fractions for vanishing phases, facilitating the computation of equilibria in a unified framework.

Note that the strict positivity of $\boldsymbol{\xi}^\alpha$ implies that any solution to this system satisfies the result of [26, Theorem 9.8], which states that if a phase is present at equilibrium, then all species within that phase must also be present.

The remainder of this section will be on demonstrating Proposition 3.1.

Proof of Proposition 3.1. We introduce the sets of positive and non-negative vectors in $\mathbb{R}^{N^\alpha}$ by $\mathbb{R}^{N^\alpha}_{>0}$ and $\mathbb{R}^{N^\alpha}_{\geq 0}$ respectively. Using these sets, we rewrite the minimization problem (8) as

$$\mathbf{n} \in \arg\min \left\{ \sum_{\alpha=1}^{N_{Ph}} \mathcal{G}_\alpha(\mathbf{n}^\alpha) \;\middle|\; \mathbf{n} \in \mathcal{M}_{\mathbf{A},\mathbf{b}} \right\}, \tag{13}$$

where

$$\mathcal{G}_\alpha(\mathbf{n}^\alpha) := \begin{cases} G_\alpha(\mathbf{n}^\alpha) & \text{if } \mathbf{n}^\alpha \in \mathbb{R}^{N^\alpha}_{\geq 0}, \\ \\ +\infty & \text{otherwise.} \end{cases} \tag{14}$$

In this way, non-negativity constraints are incorporated into the objective function.

The Gibbs function $\mathcal{G}_\alpha$ in (14) is only differentiable in $\mathbb{R}^{N^\alpha}_{>0}$, but it is possible to define its subdifferential for any $\mathbf{n}^\alpha \in \mathbb{R}^{N^\alpha}$. The subdifferential of the convex function $\mathcal{G}_\alpha : \mathbb{R}^{N^\alpha} \to \mathbb{R}$ at a point $\mathbf{n}^\alpha \in \mathbb{R}^{N^\alpha}$ is denoted $\partial\mathcal{G}_\alpha(\mathbf{n}^\alpha)$ and is defined by [25, Section 23]:

$$\boldsymbol{\mu}^\alpha \in \partial\mathcal{G}_\alpha(\mathbf{n}^\alpha) \quad \Leftrightarrow \quad \mathcal{G}_\alpha(\mathbf{m}^\alpha) \geq \mathcal{G}_\alpha(\mathbf{n}^\alpha) + \langle \boldsymbol{\mu}^\alpha,\, \mathbf{m}^\alpha - \mathbf{n}^\alpha\rangle, \qquad \forall \mathbf{m}^\alpha \in \mathbb{R}^{N^\alpha}.$$

The function $\boldsymbol{\mu}^\alpha$ is called a subgradient of $\mathcal{G}_\alpha$. In our particular case, one has

$$\partial\mathcal{G}_\alpha(\mathbf{n}^\alpha) = \begin{cases} \{\nabla\mathcal{G}_\alpha(\mathbf{n}^\alpha)\} & \text{if } \mathbf{n}^\alpha \in \mathbb{R}^{N^\alpha}_{>0}, \\ \emptyset & \text{otherwise.} \end{cases}$$

For $\mathbf{n}^\alpha = \mathbf{0}$, the subdifferential $\partial\mathcal{G}_\alpha(\mathbf{0})$ contains infinitely many elements, as shows the analysis carried out in what follows, cf. Lemma 3.1.

Then, if $\mathbf{n} = (\mathbf{n}^\alpha)_{\alpha=1,\,\ldots,\,N_{Ph}}$ is a solution to the problem defined in (13), it must satisfy the first-order KKT optimality conditions:

$$\mathbf{A}\mathbf{n} - \mathbf{b} = \mathbf{0}, \tag{15a}$$

$$\boldsymbol{\mu} + \mathbf{A}^T\boldsymbol{\Lambda} = \mathbf{0}, \tag{15b}$$

$$\boldsymbol{\mu} = (\boldsymbol{\mu}^\alpha)_{\alpha=1,\,\ldots,\,N_{Ph}} \tag{15c}$$

$$\boldsymbol{\mu}^\alpha \in \partial \mathcal{G}_\alpha(\mathbf{n}^\alpha). \tag{15d}$$

In (15), the subdifferential of $\mathcal{G}_\alpha$ includes the non-negativity constraint, while $\boldsymbol{\Lambda} = (\lambda_1, \ldots, \lambda_M)^T$ represents the vector of Lagrange multipliers associated with the linear constraint (15a).

To establish our formulation, we need to characterize $\partial \mathcal{G}_\alpha(\mathbf{n}^\alpha)$. The case where $\mathbf{n}^\alpha < \mathbf{0}$ is excluded from the minimization problem (13) by the definition of $\mathcal{G}_\alpha$. The following lemma provides a characterization of $\partial \mathcal{G}_\alpha(\mathbf{n}^\alpha)$ for $\mathbf{n}^\alpha \geq \mathbf{0}$.

Lemma 3.1. *Let $\mathbf{n}^\alpha \geq 0$, then $\boldsymbol{\mu}^\alpha \in \partial \mathcal{G}_\alpha(\mathbf{n}^\alpha)$ if and only if there exists s_α, $\boldsymbol{\xi}^\alpha = (\xi_i)_{i \in \sigma^{-1}(\alpha)}$ such that:*

$$s_\alpha \geq 0,\, \boldsymbol{\xi}^\alpha > 0, \quad \langle \boldsymbol{\xi}^\alpha,\, \mathbf{1} \rangle \leq 1, \quad s_\alpha \left(1 - \langle \boldsymbol{\xi}^\alpha,\, \mathbf{1} \rangle\right) = 0,$$

and for all $i \in \sigma^{-1}(\alpha)$,

$$\mu_i = \mu_i^0 + \mathrm{RT} \ln \xi_i, \quad n_i = s_\alpha \xi_i.$$

Proof of Lemma 3.1. In order to prove Lemma 3.1, we will describe the subdifferential of the Legendre transform $\mathcal{G}_\alpha^\star$ of $\mathcal{G}_\alpha$. Indeed, since $\mathcal{G}_\alpha$ is a lower semi-continuous proper convex function, the following equivalence holds:

$$\boldsymbol{\mu}_\alpha \in \partial \mathcal{G}_\alpha(\mathbf{n}^\alpha) \Leftrightarrow \mathbf{n}^\alpha \in \partial \mathcal{G}_\alpha^\star(\boldsymbol{\mu}_\alpha), \tag{16}$$

where $\mathcal{G}_\alpha^\star$ is the Legendre transform of $\mathcal{G}_\alpha$. This transformation is widely used in thermodynamics to convert a function defined in one set of variables into another, effectively representing the function in terms of its tangents. It is defined as

$$\mathcal{G}_\alpha^\star(\boldsymbol{\mu}^\alpha) = \sup_{\mathbf{n}^\alpha \in \mathbb{R}^{N^\alpha}} \sum_{i \in \sigma^{-1}(\alpha)} n_i \mu_i - \mathcal{G}_\alpha(\mathbf{n}^\alpha), \quad \boldsymbol{\mu}^\alpha \in \mathbb{R}^{N^\alpha}. \tag{17}$$

For $\mathbf{n}^\alpha \notin \mathbb{R}^{N^\alpha}_{\geq 0}$, we have $\mathcal{G}_\alpha(\mathbf{n}^\alpha) = +\infty$, we can then restrict the supremum in (17) to $\mathbf{n}^\alpha \geq 0$ and consider the function G_α instead of $\mathcal{G}_\alpha$. We thus obtain

$$\mathcal{G}_\alpha^\star(\boldsymbol{\mu}^\alpha) = \sup_{\mathbf{n}^\alpha \geq 0} \sum_{i \in \sigma^{-1}(\alpha)} n_i \mu_i - G_\alpha(\mathbf{n}^\alpha) = \sup_{\mathbf{n}^\alpha \geq 0} \sum_{i \in \sigma^{-1}(\alpha)} n_i [\mu_i - \mu_i^\circ - \mathrm{RT} \ln x_i(\mathbf{n}^\alpha)].$$

Shifting from the supremum to the infimum, we get

$$\mathcal{G}_\alpha^\star(\boldsymbol{\mu}^\alpha) = - \inf_{\mathbf{n}^\alpha \geq 0} \sum_{i \in \sigma^{-1}(\alpha)} n_i [\mu_i^\circ + \mathrm{RT} \ln x_i(\mathbf{n}^\alpha) - \mu_i]. \tag{18}$$

We then introduce mole fractions by means of the following change of variables:

$$x_i s_\alpha = n_i \quad \text{with} \quad s_\alpha = \langle \mathbf{n}^\alpha,\, \mathbf{1} \rangle.$$

The Legendre transform (18) is then written as

$$\mathcal{G}_\alpha^\star(\boldsymbol{\mu}^\alpha) = - \inf_{\substack{s_\alpha,\, \mathbf{x} \geq 0 \\ \langle \mathbf{x}^\alpha,\, \mathbf{1} \rangle = 1}} s_\alpha \sum_{i \in \sigma^{-1}(\alpha)} x_i [\mu_i^\circ + \mathrm{RT} \ln x_i - \mu_i]. \tag{19}$$

From this relation, we can extract a minimization problem by defining the convex function $g_\alpha(\cdot\,;\, \boldsymbol{\mu}^\alpha) : \mathbb{R}^{N^\alpha}_{>0} \to \mathbb{R}$ defined as

$$g_\alpha(\mathbf{x}^\alpha;\, \boldsymbol{\mu}^\alpha) = \sum_{i \in \sigma^{-1}(\alpha)} x_i [\mu_i^\circ + \mathrm{RT} \ln x_i - \mu_i], \tag{20}$$

where $\boldsymbol{\mu}^\alpha$ is considered as a parameter. Then, introducing the convex sets

$$X_\alpha = \left\{ \mathbf{x}^\alpha > 0 \mid \langle \mathbf{x}^\alpha,\, \mathbf{1} \rangle = 1 \right\}, \quad \text{and} \quad \overline{X_\alpha} = \left\{ \mathbf{x}^\alpha \geq 0 \mid \langle \mathbf{x}^\alpha,\, \mathbf{1} \rangle = 1 \right\},$$

the relation (19) becomes

$$\mathcal{G}_\alpha^\star(\boldsymbol{\mu}^\alpha) = - \inf_{s_\alpha \geq 0} s_\alpha \min_{\mathbf{x}^\alpha \in \overline{X_\alpha}} g_\alpha(\mathbf{x}^\alpha;\, \boldsymbol{\mu}^\alpha). \tag{21}$$

We can then study the minimization problem:

$$\min_{\mathbf{x}^\alpha \in \overline{X_\alpha}} g_\alpha(\mathbf{x}^\alpha;\, \boldsymbol{\mu}^\alpha). \tag{22}$$

The following lemma, to be proven at the end of the subsection, states that the inequality constraint on $\mathbf{x}^\alpha$ is never saturated at the solution of (22).

Lemma 3.2. *If $\mathbf{x}^\star \in \overline{X_\alpha}$ minimizes $g_\alpha(\cdot\,;\boldsymbol{\mu}^\alpha)$ on $\overline{X_\alpha}$, then $\mathbf{x}^\star \in X_\alpha$.*

Based on Lemma 3.2, the problem (22) can be simplified to

$$\min_{\langle \mathbf{x}^\alpha,\,\mathbf{1}\rangle - 1 = 0} \sum_{i \in \sigma^{-1}(\alpha)} x_i[\mu_i^\circ + \mathrm{RT}\ln x_i - \mu_i]. \tag{23}$$

The Euler-Lagrange equations associated to the problem (23) are

$$\mu_i^\circ + \mathrm{RT}\ln x_i - \mu_i + \mathrm{RT} - \lambda = 0, \quad i \in \sigma^{-1}(\alpha), \tag{24}$$

$$\langle \mathbf{x}^\alpha,\,\mathbf{1}\rangle - 1 = 0, \tag{25}$$

where λ is the Lagrange multiplier for the constraint (25). Applying equation (24) and then equation (25) to the problem (23) yields the following minimum:

$$\min_{\langle \mathbf{x}^\alpha,\,\mathbf{1}\rangle - 1 = 0} \sum_{i \in \sigma^{-1}(\alpha)} x_i[\mu_i^\circ + \mathrm{RT}\ln x_i - \mu_i] = \min_{\langle \mathbf{x}^\alpha,\,\mathbf{1}\rangle - 1 = 0} \sum_{i \in \sigma^{-1}(\alpha)} x_i[\lambda - \mathrm{RT}] = \lambda - \mathrm{RT}.$$

In order to introduce this minimum into the Legendre transform (21), we will rewrite it in terms of $\boldsymbol{\mu}^\alpha$. To do so, we define the functions $\xi_i : \mathbb{R} \to \mathbb{R}_+^\star$, $i \in \sigma^{-1}(\alpha)$, by

$$\xi_i(\mu_i) := \exp\left(\frac{\mu_i - \mu_i^\circ}{\mathrm{RT}}\right). \tag{26}$$

Then using (24) and (25), we get that:

$$\xi_i(\mu_i) = x_i \exp\left(\frac{\mathrm{RT} - \lambda}{\mathrm{RT}}\right) \quad \text{and} \quad \langle \boldsymbol{\xi}^\alpha(\boldsymbol{\mu}^\alpha),\,\mathbf{1}\rangle = \sum_{i \in \sigma^{-1}(\alpha)} x_i \exp\left(\frac{\mathrm{RT} - \lambda}{\mathrm{RT}}\right) = \exp\left(\frac{\mathrm{RT} - \lambda}{\mathrm{RT}}\right),$$

where $\boldsymbol{\xi}^\alpha(\boldsymbol{\mu}^\alpha) = (\xi_i(\mu_i))_{i \in \sigma^{-1}(\alpha)}$. The solution of problem (22) is then given by:

$$\min_{\mathbf{x}^\alpha \in \overline{X_\alpha}} g_\alpha(\mathbf{x}^\alpha) = \lambda - \mathrm{RT} = -\mathrm{RT}\ln\langle \boldsymbol{\xi}^\alpha(\boldsymbol{\mu}^\alpha),\,\mathbf{1}\rangle. \tag{27}$$

Using the minimum (27) into (21), we obtain:

$$\mathcal{G}_\alpha^\star(\boldsymbol{\mu}^\alpha) = -\inf_{s_\alpha \geq 0} -s_\alpha \mathrm{RT}\ln\langle \boldsymbol{\xi}^\alpha(\boldsymbol{\mu}^\alpha),\,\mathbf{1}\rangle = \sup_{s_\alpha \geq 0} s_\alpha \mathrm{RT}\ln\langle \boldsymbol{\xi}^\alpha(\boldsymbol{\mu}^\alpha),\,\mathbf{1}\rangle.$$

Therefore, since $s_\alpha \mathrm{RT} \geq 0$, we can consider two cases based on the sign of $\ln\langle \boldsymbol{\xi}^\alpha(\boldsymbol{\mu}^\alpha),\,\mathbf{1}\rangle$:

- $\langle \boldsymbol{\xi}^\alpha(\boldsymbol{\mu}^\alpha),\,\mathbf{1}\rangle \leq 1 \quad \Rightarrow \quad \ln\langle \boldsymbol{\xi}^\alpha(\boldsymbol{\mu}^\alpha),\,\mathbf{1}\rangle \leq 0 \quad \Rightarrow \quad \mathcal{G}_\alpha^\star(\boldsymbol{\mu}^\alpha) = 0;$
- $\langle \boldsymbol{\xi}^\alpha(\boldsymbol{\mu}^\alpha),\,\mathbf{1}\rangle > 1 \quad \Rightarrow \quad \ln\langle \boldsymbol{\xi}^\alpha(\boldsymbol{\mu}^\alpha),\,\mathbf{1}\rangle > 0 \quad \Rightarrow \quad \mathcal{G}_\alpha^\star(\boldsymbol{\mu}^\alpha) = +\infty.$

We define the set K_α as follows

$$K_\alpha := \{\boldsymbol{\mu}^\alpha \mid \langle \boldsymbol{\xi}^\alpha(\boldsymbol{\mu}^\alpha),\,\mathbf{1}\rangle \leq 1\}.$$

This set is a convex set as it is the sublevel set of the convex function $\Psi : \boldsymbol{\mu}^\alpha \mapsto \langle \boldsymbol{\xi}^\alpha(\boldsymbol{\mu}^\alpha),\,\mathbf{1}\rangle$. Alongside K_α, we introduce the associated characteristic function:

$$\chi_{K_\alpha}(\boldsymbol{\mu}^\alpha) = \begin{cases} 0 & \text{if } \boldsymbol{\mu}^\alpha \in K_\alpha, \\ +\infty & \text{if } \boldsymbol{\mu}^\alpha \notin K_\alpha. \end{cases}$$

With these definitions, we can express the Legendre transform as follows:

$$\mathcal{G}_\alpha^\star(\boldsymbol{\mu}^\alpha) = \chi_{K_\alpha}(\boldsymbol{\mu}^\alpha).$$

We can now easily calculate the subdifferential of $\mathcal{G}_\alpha^\star$ at a point $\boldsymbol{\mu}^\alpha \in K_\alpha$. If $\boldsymbol{\mu}^\alpha$ lies in the interior of K_α, the subdifferential is reduced to the gradient of $\mathcal{G}_\alpha^\star$, which is equal to zero. Otherwise, if $\boldsymbol{\mu}^\alpha$ is on the boundary of K_α, the definition of the subdifferential leads to the condition:

$$\mathbf{n}^\alpha \in \partial\mathcal{G}_\alpha^\star(\boldsymbol{\mu}^\alpha) \quad \Leftrightarrow \quad 0 \geq \langle \mathbf{n}^\alpha,\,\boldsymbol{\eta}^\alpha - \boldsymbol{\mu}^\alpha\rangle,\ \forall \boldsymbol{\eta}^\alpha \in \mathbb{R}^{N^\alpha},$$

which is known as the normal cone of the convex set K_α. The boundary of K_α is smooth because it is defined by $\Psi(\boldsymbol{\mu}^\alpha) = 1$. Consequently, this cone can be simplified to the normal direction, which is determined by the gradient of Ψ. Therefore, the subdifferential of $\mathcal{G}_\alpha^\star$ is:

$$\partial \mathcal{G}_\alpha^\star(\boldsymbol{\mu}^\alpha) = \begin{cases} 0 & \text{if } \langle \boldsymbol{\xi}^\alpha(\boldsymbol{\mu}^\alpha),\, \mathbf{1}\rangle < 1, \\ \gamma_\alpha \boldsymbol{\xi}^\alpha(\boldsymbol{\mu}^\alpha), \gamma_\alpha \geq 0 & \text{if } \langle \boldsymbol{\xi}^\alpha(\boldsymbol{\mu}^\alpha),\, \mathbf{1}\rangle = 1. \end{cases}$$

We can now conclude the proof of Lemma 3.1.

Let $\mathbf{n}^\alpha \geq 0$, then

$$\begin{aligned} \boldsymbol{\mu}^\alpha \in \partial \mathcal{G}_\alpha(\mathbf{n}^\alpha) \Leftrightarrow \mathbf{n}^\alpha \in \partial \mathcal{G}_\alpha^\star(\boldsymbol{\mu}^\alpha) &\Leftrightarrow \begin{cases} n_i = \gamma_\alpha \xi_i(\mu_i),\, \gamma_\alpha \geq 0,\, i \in \sigma^{-1}(\alpha), \\ \langle \boldsymbol{\xi}^\alpha(\boldsymbol{\mu}^\alpha),\, \mathbf{1}\rangle \leq 1, \end{cases} \\ &\Leftrightarrow \begin{cases} n_i = \gamma_\alpha \xi_i,\, \gamma_\alpha \geq 0,\, i \in \sigma^{-1}(\alpha), \\ \mu_i = \mu_i^\circ + \mathrm{RT} \ln \xi_i,\, \xi_i > 0, \\ \langle \boldsymbol{\xi}^\alpha,\, \mathbf{1}\rangle \leq 1. \end{cases} \end{aligned}$$

We already know that if $\mathbf{n}^\alpha > 0$, the derivative of g_α is given by $\mu_i(\mathbf{n}^\alpha) = \mu_i^\circ + \mathrm{RT} \ln n_i/s_\alpha$, thus $\xi_i = n_i/s_\alpha$, $\gamma_\alpha = s_\alpha > 0$ and $\langle \boldsymbol{\xi}^\alpha,\, \mathbf{1}\rangle = 1$. In the case where $\mathbf{n}^\alpha = 0$, it is clear that $\gamma_\alpha = 0 = s_\alpha$ and that $\langle \boldsymbol{\xi}^\alpha,\, \mathbf{1}\rangle \leq 1$. Therefore, in each of these cases, one has

$$s_\alpha \geq 0,\, \boldsymbol{\xi}^\alpha > 0, \quad \langle \boldsymbol{\xi}^\alpha,\, \mathbf{1}\rangle \leq 1, \quad s_\alpha \left(1 - \langle \boldsymbol{\xi}^\alpha,\, \mathbf{1}\rangle\right) = 0,\, n_i = s_\alpha \xi_i,\, i \in \sigma^{-1}(\alpha),$$

concluding the proof of Lemma 3.1. □

Using this result, we can rewrite the KKT optimality conditions (15) as

$$\begin{aligned} \sum_{\alpha=1}^{N_{Ph}} s_\alpha \mathbf{A}^\alpha \boldsymbol{\xi}^\alpha - \mathbf{b} &= \mathbf{0}, \\ \mathbf{S}^T [\mu_i^\circ + \mathrm{RT} \ln \xi_i]_{i=1,\ldots,N} &= \mathbf{0}, \\ s_\alpha \left(1 - \langle \boldsymbol{\xi}^\alpha,\, \mathbf{1}\rangle\right) &= 0, && (\alpha = 1, \ldots, N_{Ph}), \\ s_\alpha \geq 0,\, 1 - \langle \boldsymbol{\xi}^\alpha,\, \mathbf{1}\rangle &\geq 0, && (\alpha = 1, \ldots, N_{Ph}). \end{aligned}$$

where $\boldsymbol{\mu}(\boldsymbol{\xi}) := (\mu_i^\circ + \mathrm{RT} \ln \xi_i)_{i=1,\ldots,N}$. Therefore, based on the relation $\mathbf{S}^T \mathbf{A}^T = \mathbf{0}$ from (4), and defining

$$r_\alpha := 1 - \langle \boldsymbol{\xi}^\alpha,\, \mathbf{1}\rangle,$$

we finally obtain:

$$\sum_{\alpha=1}^{N_{Ph}} s_\alpha \mathbf{A}^\alpha \boldsymbol{\xi}^\alpha - \mathbf{b} = \mathbf{0}, \tag{12a}$$

$$\mathbf{S}^T [\mu_i^\circ + \mathrm{RT} \ln \xi_i]_{i=1,\ldots,N} = \mathbf{0}, \tag{12b}$$

$$\langle \boldsymbol{\xi}^\alpha,\, \mathbf{1}\rangle + r_\alpha - 1 = 0, \qquad (\alpha = 1, \ldots, N_{Ph}), \tag{12c}$$

$$s_\alpha r_\alpha = 0, \qquad (\alpha = 1, \ldots, N_{Ph}), \tag{12d}$$

$$s_\alpha \geq 0,\, r_\alpha \geq 0, \qquad (\alpha = 1, \ldots, N_{Ph}). \tag{12e}$$

Given the convexity of the Gibbs function and constraints, the equivalence between the minimization problem and the solution of system (12) follows from assumption (H1), which guarantees that the Slater condition holds – namely, the existence of a strictly feasible point. This classical constraint qualification ensures strong duality and the equivalence between the constrained minimization problem and its KKT system. This concludes the proof of Proposition 3.1. □

It remains to proof Lemma 3.2 to complete the proof of the main result of the section, namely Proposition 3.1.

Proof of Lemma 3.2. The function g_α is continuous on the compact set $\overline{X_\alpha}$, then from the Weierstrass theorem, there exists $\mathbf{x}^\star \in \{\arg\min g_\alpha(\mathbf{x}^\alpha;\, \boldsymbol{\mu}^\alpha) \mid \mathbf{x}^\alpha \in \overline{X_\alpha}\}$. Let us assume that $\mathbf{x}^\star \in \overline{X_\alpha} \setminus X_\alpha$, meaning that there exists a subset $\mathcal{J} \subsetneq \sigma^{-1}(\alpha)$ such that $x_j^\star = 0$, for all $j \in \mathcal{J}$. Note that the case $\mathcal{J} = \sigma^{-1}(\alpha)$ is excluded since $\langle \mathbf{x}^\alpha,\, \mathbf{1}\rangle = 1$. Let $\mathbf{x} \in X_\alpha$ and $\varepsilon \in (0, 1)$, then one defines $\mathbf{x}^0 := \mathbf{x} - \mathbf{x}^\star$ and $\mathbf{x}^\varepsilon := \mathbf{x}^\star + \varepsilon \mathbf{x}^0 = \mathbf{x}^\varepsilon = \varepsilon \mathbf{x} + (1 - \varepsilon)\mathbf{x}^\star$. The vector $\mathbf{x}^\varepsilon$ is a convex linear combination of vectors of $\overline{X_\alpha}$

which is a convex set, hence $\mathbf{x}^\varepsilon \in \overline{X_\alpha}$. Furthermore, $\mathbf{x}^\varepsilon \in X_\alpha$ since $\mathbf{x}^\varepsilon = \varepsilon\mathbf{x} + (1-\varepsilon)\mathbf{x}^\star \geq \varepsilon\mathbf{x} > 0$. By convexity of g_α on $\mathbb{R}^{N^\alpha}_{\geq 0}$,

$$g_\alpha(\mathbf{x}^\star;\, \boldsymbol{\mu}^\alpha) \geq g_\alpha(\mathbf{x}^\varepsilon;\, \boldsymbol{\mu}^\alpha) + \langle \boldsymbol{\nabla} g_\alpha(\mathbf{x}^\varepsilon),\, \mathbf{x}^\star - \mathbf{x}^\varepsilon \rangle \tag{28}$$

$$\Leftrightarrow \frac{g_\alpha(\mathbf{x}^\star;\, \boldsymbol{\mu}^\alpha) - g_\alpha(\mathbf{x}^\varepsilon;\, \boldsymbol{\mu}^\alpha)}{\varepsilon} \geq -\langle \boldsymbol{\nabla} g_\alpha(\mathbf{x}^\varepsilon),\, \mathbf{x}^0 \rangle, \tag{29}$$

where $\boldsymbol{\nabla} g_\alpha(\mathbf{x}^\varepsilon) := (\mu_i^\circ + \mathrm{RT} \ln x_i^\varepsilon + \mathrm{RT})_{i \in \sigma^{-1}(\alpha)}$.

We will now take the limit when ε tends to 0 in inequality (29). In the right-hand side one has

$$\lim_{\varepsilon \to 0} -\langle \boldsymbol{\nabla} g_\alpha(\mathbf{x}^\varepsilon),\, \mathbf{x}^0 \rangle = -\sum_{j \in \mathcal{J}} x_j^0 \lim_{\varepsilon \to 0} \partial_{x_j} g_\alpha(\mathbf{x}^\varepsilon) - \sum_{i \in \sigma^{-1}(\alpha) \setminus \mathcal{J}} x_i^0 \lim_{\varepsilon \to 0} \partial_{x_i} g_\alpha(\mathbf{x}^\varepsilon). \tag{30}$$

Noting that $\lim_{\epsilon \to 0} \mathbf{x}^\varepsilon = \mathbf{x}^\star$ and in particular that $\lim_{\varepsilon \to 0} x_j^\varepsilon = x_j^\star = 0$, for all $j \in \mathcal{J}$, it follows from the continuity of $\boldsymbol{\nabla} g_\alpha$ that

$$\lim_{\varepsilon \to 0} \partial_{x_j} g_\alpha(\mathbf{x}^\varepsilon) = -\infty,\ \forall j \in \mathcal{J} \quad \text{and} \quad \lim_{\varepsilon \to 0} \partial_{x_i} g_\alpha(\mathbf{x}^\varepsilon) = \partial_{x_i} g_\alpha(\mathbf{x}^\star) \in \mathbb{R},\ \forall i \in \sigma^{-1}(\alpha) \setminus \mathcal{J}. \tag{31}$$

By combining (30) and (31) with $x_j^0 = x_j > 0$, for all $j \in \mathcal{J}$, one finds that the right-hand side of (29) tends to $+\infty$. However if $\mathbf{x}^\star$ minimizes g_α on $\overline{X_\alpha}$, then the left-hand side of (29) is non-positive which is a contradiction. Therefore $\mathcal{J} = \emptyset$ and $\mathbf{x}^\star \in X_\alpha$. □

3.3 On the uniqueness for the unified formulation

The existence of a solution is guaranteed by the assumption that $\mathcal{M}_{\mathbf{A},\mathbf{b}} \cap \mathbb{R}^N_{>0}$ is nonempty (H1) and by the compactness of $\mathcal{M}_{\mathbf{A},\mathbf{b}} \cap \mathbb{R}^N_{\geq 0}$. We have seen in Theorem 3.1 that the minimizer $\mathbf{n}$ for (8), defined in (9) and (10), is unique if and only if the set of vectors $\{\mathbf{A}^\alpha \mathbf{n}^\alpha\}_{\alpha \in \mathit{\Gamma}_\mathbf{n}}$, where $\mathit{\Gamma}_\mathbf{n} = \{\alpha \in \{1, \ldots, N_{Ph}\} \mid \mathbf{n}^\alpha \neq \mathbf{0}\}$, is linearly independent.

To better understand this condition, let us assume that condition (11) holds true for some minimizer $\mathbf{n} = (s_{\sigma(i)}\xi_i)_{i \in \mathcal{C}}$ of the Gibbs energy, with s_α and $\boldsymbol{\xi}^\alpha$, $\alpha \in \mathcal{P}$, as in Proposition 3.1. Then setting $\lambda_\alpha = 0$ for all $\alpha \in \mathcal{P} \setminus \mathit{\Gamma}_\mathbf{n}$, then $(\lambda_\alpha)_{\alpha \in \mathcal{P}} \neq \mathbf{0}$ is such that

$$\sum_{\alpha \in \mathcal{P}} \lambda_\alpha \mathbf{A}^\alpha \boldsymbol{\xi}^\alpha = \mathbf{0},$$

or, equivalently, such that

$$\boldsymbol{w} = \big(\lambda_{\sigma(i)} \xi_i\big)_{i \in \mathcal{C}} \in \ker \mathbf{A} \setminus \mathbf{0}. \tag{32}$$

Since $\lambda_\alpha = 0$ if $s_\alpha = 0$, there exists $\bar{\epsilon} > 0$ such that $s_\alpha + \epsilon\lambda_\alpha \geq 0$ for all $\epsilon \in (0, \bar{\epsilon})$. Denote by

$$\mathbf{n}_\epsilon = \big((s_{\sigma(i)} + \epsilon\lambda_{\sigma(i)})\xi_i\big)_{i \in \mathcal{C}} \geq \mathbf{0},$$

then $\mathbf{A}\mathbf{n}_\epsilon = \mathbf{b}$, and, setting $\boldsymbol{\mu} = (\mu_i^\circ + RT \log(\xi_i))_{i \in \mathcal{C}} = (\mu_i)_{i \in \mathcal{C}}$ as in the proof of Proposition 3.1, we check that

$$G(\mathbf{n}_\epsilon) = \sum_{\alpha \in \mathcal{P}} (s_\alpha + \epsilon\lambda_\alpha) \sum_{i \in \sigma^{-1}(\alpha)} \xi_i\, \mu_i = G(\mathbf{n}) + \epsilon\, \boldsymbol{w}^T \boldsymbol{\mu} = G(\mathbf{n})$$

since $\boldsymbol{\mu} \in \ker \mathbf{S}^T$ owing to (12b), since $\boldsymbol{w} \in \ker \mathbf{A}$ owing to (32), and since $\ker \mathbf{S}^T = (\ker \mathbf{A})^\perp$ thanks to (4). Therefore, $\mathbf{n}_\epsilon$ is also minimizer of problem (8), thus a solution to the unified formulation of Proposition 3.1, for all $\epsilon \in (0, \bar{\epsilon})$. In particular, condition (11) yields an infinite number of solutions to the minimization problem, and thus to the unified formulation.

While condition

$$(\mathbf{A}^\alpha \mathbf{n}^\alpha)_{\alpha \in \mathit{\Gamma}_\mathbf{n}} \text{ is linearly independent}$$

is intrinsic to the physical problem to avoid azeotropy, and does not stem from our unified formulation, uniqueness can also fail in some very specific configurations where the extended mole fraction may become ambiguous for some absent species. To illustrate such configurations, let us analyze the following example system:

$$\begin{aligned}
\mathcal{C} &= \{\mathrm{H_2O(aq)},\, \mathrm{H^+},\, \mathrm{OH^-},\, \mathrm{H_2O(g)}\},\\
\mathcal{E} &= \{\mathrm{H},\, \mathrm{O}\},\\
\mathcal{R} &= \{\quad \mathrm{OH^-} = \mathrm{H_2O(aq)} - \mathrm{H^+},\\
&\qquad\ \mathrm{H_2O(g)} = \mathrm{H_2O(aq)}\},\\
\mathcal{P} &= \{\text{aqueous},\, \text{gaseous}\},
\end{aligned} \tag{33}$$

with the right-hand side

$$\mathbf{b} = \lambda(2,1)^T,\ \lambda > 0,$$

and with given pressure and temperature such that $\mu^\circ_{\mathrm{H_2O(g)}} < \mu^\circ_{\mathrm{H_2O(aq)}}$. The associated formula and stoichiometry matrices are

$$\mathbf{A} = \begin{pmatrix} 2 & 1 & 1 & 2 \\ 1 & 0 & 1 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{S}^T = \begin{pmatrix} 1 & -1 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{pmatrix}.$$

Let

$$\boldsymbol{\xi} = (\xi_{\mathrm{H_2O(aq)}},\, \xi_{\mathrm{H^+}},\, \xi_{\mathrm{OH^-}},\, \xi_{\mathrm{H_2O(g)}}), \quad \mathbf{s} = (s_{\mathrm{aq}},\, s_{\mathrm{g}}) \quad \text{and} \quad \mathbf{r} = (r_{\mathrm{aq}},\, r_{\mathrm{g}})$$

be such that

$$\xi_{\mathrm{H_2O(g)}} = 1 \quad \text{and} \quad s_{\mathrm{aq}} = 0.$$

Let us establish under which conditions the system (12) is fulfilled. Equation (12a) yields

$$s_{\mathrm{aq}} \begin{pmatrix} 2 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \xi_{\mathrm{H_2O(aq)}} \\ \xi_{\mathrm{H^+}} \\ \xi_{\mathrm{OH^-}} \end{pmatrix} + s_{\mathrm{g}} \begin{pmatrix} 2 \\ 1 \end{pmatrix} \xi_{\mathrm{H_2O(g)}} = \lambda \begin{pmatrix} 2 \\ 1 \end{pmatrix},$$

and is satisfied if $s_{\mathrm{g}} = \lambda$.

It follows from equations (12c) and (12d) that $(\xi_{\mathrm{H_2O(aq)}},\, \xi_{\mathrm{H^+}},\, \xi_{\mathrm{OH^-}})$ must satisfy

$$\xi_{\mathrm{H_2O(aq)}} + \xi_{\mathrm{H^+}} + \xi_{\mathrm{OH^-}} \leq 1 \tag{34}$$

and that $r_{\mathrm{g}} = 0$, $r_{\mathrm{aq}} \geq 0$. Finally, equations (12b) yields

$$\frac{\mu^\circ_{\mathrm{H_2O(aq)}} - \mu^\circ_{\mathrm{H^+}} - \mu^\circ_{\mathrm{OH^-}}}{\mathrm{RT}} + \ln \xi_{\mathrm{H_2O(aq)}} - \ln \xi_{\mathrm{H^+}} - \ln \xi_{\mathrm{OH^-}} = 0, \tag{35a}$$

$$\frac{\mu^\circ_{\mathrm{H_2O(aq)}} - \mu^\circ_{\mathrm{H_2O(g)}}}{\mathrm{RT}} + \ln \xi_{\mathrm{H_2O(aq)}} - \underbrace{\ln \xi_{\mathrm{H_2O(g)}}}_{=0} = 0. \tag{35b}$$

We can rewrite (35b) as

$$\xi_{\mathrm{H_2O(aq)}} = \underbrace{\mathrm{e}^{\frac{\mu^\circ_{\mathrm{H_2O(g)}} - \mu^\circ_{\mathrm{H_2O(aq)}}}{\mathrm{RT}}}}_{:=\kappa_1}. \tag{36}$$

Therefore, using (36) into (35a), we get that

$$\xi_{\mathrm{OH^-}} = \frac{1}{\xi_{\mathrm{H^+}}} \underbrace{\mathrm{e}^{\frac{\mu^\circ_{\mathrm{H_2O(g)}} - \mu^\circ_{\mathrm{H^+}} - \mu^\circ_{\mathrm{OH^-}}}{\mathrm{RT}}}}_{:=\kappa_2}. \tag{37}$$

It follows from (34), (36) and (37) that:

$$\xi_{\mathrm{H_2O(aq)}} + \xi_{\mathrm{H^+}} + \xi_{\mathrm{OH^-}} \leq 1 \ \Leftrightarrow\ \xi^2_{\mathrm{H^+}} + (\kappa_1 - 1)\xi_{\mathrm{H^+}} + \kappa_2 \leq 0.$$

The discriminant of this quadratic inequality is given by:

$$\Delta = (\kappa_1 - 1)^2 - 4\kappa_2.$$

Using (34), (36) and (37) we get that:

$$1 - \kappa_1 \geq \xi_{\mathrm{H^+}} + \xi_{\mathrm{OH^-}} \ \Leftrightarrow\ \Delta \geq (\xi_{\mathrm{H^+}} + \xi_{\mathrm{OH^-}})^2 - 4\xi_{\mathrm{H^+}}\xi_{\mathrm{OH^-}} \ \Leftrightarrow\ \Delta \geq (\xi_{\mathrm{H^+}} - \xi_{\mathrm{OH^-}})^2 \geq 0.$$

It follows that:

$$\xi_{\mathrm{H^+}} \in \left]\frac{1 - \kappa_1 - \sqrt{(1-\kappa_1)^2 - 4\kappa_2}}{2},\ \frac{1 - \kappa_1 + \sqrt{(1-\kappa_1)^2 - 4\kappa_2}}{2}\right[.$$

Therefore, the chemical system (33) leads to a nonunique solution of equations (12). Roughly speaking, in a configuration with only pure water vapor, our unified formulation cannot give a precise value for the pH of the absent liquid phase.

The previous example indicates that a condition on the vector $\mathbf{b}$ is necessary to ensure the uniqueness of the vector $\boldsymbol{\xi}$. This condition basically requires that the vector $\mathbf{b}$ cannot be recovered with less that M species. Recalling that $\mathbf{a}_j$ is the formula vector for the species C_j as defined in Section 2, let us make the following assumptions about the vector $\mathbf{b}$:

(H2) for all subset $\Gamma \subset \{1, \ldots, N_{Ph}\}$ such that $\{\mathbf{a}_j\}_{j\in\sigma^{-1}(\Gamma)}$ is a non-spanning set of vectors in $\mathbb{R}^M$, one has $\mathbf{b} \notin \operatorname{span}\Big(\{\mathbf{a}_j\}_{j\in\sigma^{-1}(\Gamma)}\Big)$.

We can then write the following uniqueness proposition.

Proposition 3.2. *Let $\boldsymbol{\mathcal{X}} = (\boldsymbol{\xi}^\alpha, s_\alpha, r_\alpha)_{\alpha=1,\ldots,N_{Ph}}$ be a solution of system (12) where $\mathbf{b}$ satisfies assumption (H2), and let $\Gamma_{\boldsymbol{\mathcal{X}}}$ be the corresponding set of present phases. Then $\boldsymbol{\mathcal{X}}$ is unique if and only if the set of vectors $\{\mathbf{A}^\alpha\boldsymbol{\xi}^\alpha\}_{\alpha\in\Gamma_{\boldsymbol{\mathcal{X}}}}$ is linearly independent.*

To prove this proposition, we first need to establish the following two intermediate results.

Lemma 3.3. *Let $\boldsymbol{\mathcal{X}} = (\boldsymbol{\xi}^\alpha, s_\alpha, r_\alpha)_{\alpha=1,\ldots,N_{Ph}}$ be a solution of (12). If $\mathbf{b}$ satisfies assumption (H2), then there are at least M species present in $\boldsymbol{\mathcal{X}}$. Furthermore, system (12) can be rewritten with these M species as the primary species.*

Proof. Let $\Gamma_{\boldsymbol{\mathcal{X}}}$ be the set of present phases of $\boldsymbol{\mathcal{X}}$, in this way $\sigma^{-1}(\Gamma_{\boldsymbol{\mathcal{X}}})$ is the set of all present species. From the assumption (H2), $\{\mathbf{a}_j\}_{j\in\sigma^{-1}(\Gamma_{\boldsymbol{\mathcal{X}}})}$ is a spanning set in $\mathbb{R}^M$. It is then possible to extract a basis of $\mathbb{R}^M$ from the set $\{\mathbf{a}_j\}_{j\in\sigma^{-1}(\Gamma_{\boldsymbol{\mathcal{X}}})}$ and to construct a permutation matrix $\mathbf{P} \in \mathbb{R}^{N\times N}$ such that this basis forms a primary block in $\tilde{\mathbf{A}} := \mathbf{AP}$. In this way, $\tilde{\mathbf{A}} = [\tilde{\mathbf{A}}_{Pr}, \tilde{\mathbf{A}}_{Sd}]$ where $\tilde{\mathbf{A}}_{Pr}$ is an invertible matrix. Let $\boldsymbol{\mu} := [\mu_i^\circ + \mathrm{RT}\ln\xi_i]_{i=1,\ldots,N}$, since $\mathbf{S}^T\boldsymbol{\mu} = \mathbf{0}$ and $\ker\mathbf{S}^T = \operatorname{Im}\mathbf{A}^T$ from (4), there exists $\mathbf{y} \in \mathbb{R}^M$ such that

$$\boldsymbol{\mu} = \mathbf{A}^T\mathbf{y} \quad\Leftrightarrow\quad \mathbf{P}^T\boldsymbol{\mu} = \tilde{\mathbf{A}}^T\mathbf{y} \quad\Leftrightarrow\quad \tilde{\mathbf{S}}^T\mathbf{P}^T\boldsymbol{\mu} = \mathbf{0},$$

where

$$\tilde{\mathbf{S}} := \begin{bmatrix} \tilde{\mathbf{A}}_{Pr}^{-1}\tilde{\mathbf{A}}_{Sd} \\ -\mathbf{I}_{Sd} \end{bmatrix}. \tag{38}$$

Then one readily checks that $(\mathbf{n}, \boldsymbol{\mu})$ solves

$$\begin{cases} \mathbf{An} = \mathbf{b}, \\ \mathbf{S}^T\boldsymbol{\mu} = \mathbf{0} \end{cases}$$

if and only if $(\tilde{\mathbf{n}}, \tilde{\boldsymbol{\mu}}) = (\mathbf{P}^T\mathbf{n}, \mathbf{P}^T\boldsymbol{\mu})$ solves

$$\begin{cases} \tilde{\mathbf{A}}\tilde{\mathbf{n}} = \mathbf{b}, \\ \tilde{\mathbf{S}}^T\tilde{\boldsymbol{\mu}} = \mathbf{0}, \end{cases}$$

from which a counterpart to (12) can be derived. □

Lemma 3.4. *Under assumption (H2), the vector $\boldsymbol{\xi} = (\boldsymbol{\xi}^\alpha)_{\alpha=1,\ldots,N_{Ph}}$, satisfying the system (12), is unique.*

To establish this lemma, we require the following result from information theory.

Lemma 3.5 (Gibb's inequality)**.** *Suppose that $\boldsymbol{\xi} = (\xi_i)_{i=1,\ldots,n}$ and $\overline{\boldsymbol{\xi}} = (\overline{\xi})_{i=1,\ldots,n}$ are two discrete probability distributions then*

$$\sum_{i=1}^{n} \xi_i(\ln\xi_i - \ln\overline{\xi}_i) \geq 0,$$

with equality if and only if $\xi_i = \overline{\xi}_i$, $i = 1, \ldots, n$.

Proof of Lemma 3.4. Let $\boldsymbol{\mathcal{X}} = (\boldsymbol{\xi}^\alpha, s_\alpha, r_\alpha)_{\alpha=1,\ldots,N_{Ph}}$ and $\overline{\boldsymbol{\mathcal{X}}} = (\overline{\boldsymbol{\xi}}^\alpha, \overline{s_\alpha}, \overline{r_\alpha})_{\alpha=1,\ldots,N_{Ph}}$ be two solutions of (12), then

$$\mathbf{A}[s_\alpha\boldsymbol{\xi}^\alpha - \overline{s_\alpha}\overline{\boldsymbol{\xi}}^\alpha]_{\alpha=1,\ldots,N_{Ph}} = 0, \tag{39}$$

$$\mathbf{S}^T[\ln\boldsymbol{\xi}^\alpha - \ln\overline{\boldsymbol{\xi}}^\alpha]_{\alpha=1,\ldots,N_{Ph}} = 0. \tag{40}$$

Due to (4), we have the relation $\ker\mathbf{S}^T = (\ker\mathbf{A})^\perp$, therefore

$$\sum_{\alpha=1}^{N_{Ph}} \langle s_\alpha\boldsymbol{\xi}^\alpha - \overline{s_\alpha}\overline{\boldsymbol{\xi}}^\alpha, \ln\boldsymbol{\xi}^\alpha - \ln\overline{\boldsymbol{\xi}}^\alpha\rangle = 0. \tag{41}$$

We will prove that the sum (41) is composed of non-negative terms. For each phase α, there are three cases:

- if $s_\alpha = \overline{s_\alpha} = 0$, then $\langle s_\alpha \boldsymbol{\xi}^\alpha - \overline{s_\alpha}\overline{\boldsymbol{\xi}}^\alpha, \ln \boldsymbol{\xi}^\alpha - \ln \overline{\boldsymbol{\xi}}^\alpha \rangle = 0$;
- if $s_\alpha = \overline{s_\alpha} > 0$, then $\langle s_\alpha \boldsymbol{\xi}^\alpha - \overline{s_\alpha}\overline{\boldsymbol{\xi}}^\alpha, \ln \boldsymbol{\xi}^\alpha - \ln \overline{\boldsymbol{\xi}}^\alpha \rangle \geq 0$ since the logarithm is and increasing function;
- if $s_\alpha \neq \overline{s_\alpha}$, we can assume, without loss of generality, that $s_\alpha > \overline{s_\alpha}$. This implies that $s_\alpha > 0$ and then $r_\alpha = 0$. We can rewrite the term in the sum (41) as

$$\overline{s_\alpha}\langle \boldsymbol{\xi}^\alpha - \overline{\boldsymbol{\xi}}^\alpha, \ln \boldsymbol{\xi}^\alpha - \ln \overline{\boldsymbol{\xi}}^\alpha \rangle + (s_\alpha - \overline{s_\alpha})\langle \boldsymbol{\xi}^\alpha, \ln \boldsymbol{\xi}^\alpha - \ln \overline{\boldsymbol{\xi}}^\alpha \rangle =: \underbrace{\overline{s_\alpha}\mathcal{A}_\alpha}_{\geq 0} + \underbrace{(s_\alpha - \overline{s_\alpha})}_{>0}\mathcal{B}_\alpha. \tag{42}$$

  Applying the convexity inequality:

$$x(\ln x - \ln y) \geq x - y,$$

  to $\mathcal{B}_\alpha$ yields

$$\mathcal{B}_\alpha = \langle \boldsymbol{\xi}^\alpha, \ln \boldsymbol{\xi}^\alpha - \ln \overline{\boldsymbol{\xi}}^\alpha \rangle \geq \sum_{i \in \sigma^{-1}(\alpha)} \xi_i - \overline{\xi}_i = \overline{r_\alpha} - r_\alpha = \overline{r_\alpha} \geq 0 \tag{43}$$

  since $r_\alpha = 0$. Consequently, (42) is non-negative.

Therefore, (41) is a vanishing sum of non-negative terms, it follows that for each $\alpha \in \mathcal{P}$:

$$\langle s_\alpha \boldsymbol{\xi}^\alpha - \overline{s_\alpha}\overline{\boldsymbol{\xi}}^\alpha, \ln \boldsymbol{\xi}^\alpha - \ln \overline{\boldsymbol{\xi}}^\alpha \rangle = \overline{s_\alpha}\mathcal{A}_\alpha + (s_\alpha - \overline{s_\alpha})\mathcal{B}_\alpha = 0. \tag{44}$$

In order to prove the uniqueness of ξ^α, there are the same three cases to deal with:

- if $s_\alpha = \overline{s_\alpha} > 0$, then (44) becomes $\mathcal{A}_\alpha = 0$. We introduce the function

$$\mathbf{g}_\alpha(\boldsymbol{\xi}^\alpha) := \sum_{i \in \sigma^{-1}(\alpha)} \xi_i \ln \xi_i,$$

  which is strictly convex on $\mathbb{R}^{N^\alpha}_{>0}$. We can then rewrite $\mathcal{A}_\alpha$ as

$$\mathcal{A}_\alpha = \langle \boldsymbol{\xi}^\alpha - \overline{\boldsymbol{\xi}}^\alpha, \boldsymbol{\nabla}\mathbf{g}_\alpha(\boldsymbol{\xi}^\alpha) - \boldsymbol{\nabla}\mathbf{g}_\alpha(\overline{\boldsymbol{\xi}}^\alpha) \rangle.$$

  The strict convexity of the function $\mathbf{g}_\alpha$ on the set $\mathbb{R}^{N^\alpha}_{>0}$ implies the strict monotonicity of its gradient $\boldsymbol{\nabla}\mathbf{g}_\alpha$, consequently

$$\mathcal{A}_\alpha = 0 \quad \Rightarrow \quad \boldsymbol{\xi}^\alpha = \overline{\boldsymbol{\xi}}^\alpha;$$

- if $s_\alpha > \overline{s_\alpha}$, then $s_\alpha > 0$ we have seen that $r_\alpha = 0$ and

$$\underbrace{\overline{s_\alpha}\mathcal{A}_\alpha}_{\geq 0} + \underbrace{(s_\alpha - \overline{s_\alpha})\mathcal{B}_\alpha}_{\geq 0} = 0. \tag{45}$$

  As a result, both terms in the sum (45) are zero, and therefore, $\mathcal{B}_\alpha = 0$. From (43), we deduce that $\overline{r_\alpha} = 0$. It follows that $\langle \boldsymbol{\xi}^\alpha, \mathbf{1} \rangle = 1$ and $\langle \overline{\boldsymbol{\xi}}^\alpha, \mathbf{1} \rangle = 1$. This result allows us to apply Gibb's inequality (Lemma 3.5), which states:

$$\mathcal{B}_\alpha = \langle \boldsymbol{\xi}^\alpha, \ln \boldsymbol{\xi}^\alpha - \ln \overline{\boldsymbol{\xi}}^\alpha \rangle = 0 \quad \Leftrightarrow \quad \boldsymbol{\xi}^\alpha = \overline{\boldsymbol{\xi}}^\alpha;$$

- if $s_\alpha = \overline{s_\alpha} = 0$, we cannot conclude about the uniqueness of $\boldsymbol{\xi}^\alpha$ from (44) but Lemma 3.3 guarantees that there are at least M primary species present in $\boldsymbol{\mathcal{X}}$ and that the formula matrix $\mathbf{A}$ can be reformulated with these M species as the first M columns. Let $\boldsymbol{\xi}_{Pr}$ and $\overline{\boldsymbol{\xi}}_{Pr}$ the corresponding vectors in $\boldsymbol{\mathcal{X}}$ and $\overline{\boldsymbol{\mathcal{X}}}$ respectively. We have established in Proposition 3.1 that $\boldsymbol{\xi}^\alpha > 0$ for each phase α. This implies that if a species is present in a phase α, then $s_\alpha > 0$, indicating that the entire phase is present. It follows that the phases where $\boldsymbol{\xi}_{Pr}$ belongs are all present, and from the previous cases we can conclude that $\boldsymbol{\xi}_{Pr} = \overline{\boldsymbol{\xi}}_{Pr}$. Let $\mathbf{d} := \mathbf{S}^T\boldsymbol{\mu}^\circ/(\mathrm{RT})$, it follows from (12) that for each $i \in \sigma^{-1}(\alpha)$,

$$\left(\mathbf{d} + \tilde{\mathbf{A}}_{Sd}^T \tilde{\mathbf{A}}_{Pr}^{-T}[\ln \boldsymbol{\xi}_{Pr}]\right)_i = \ln \xi_i^\alpha,$$
$$\left(\mathbf{d} + \tilde{\mathbf{A}}_{Sd}^T \tilde{\mathbf{A}}_{Pr}^{-T}[\ln \overline{\boldsymbol{\xi}}_{Pr}]\right)_i = \ln \overline{\xi}_i^\alpha,$$

  and since $\boldsymbol{\xi}_{Pr} = \overline{\boldsymbol{\xi}}_{Pr}$,

$$\left(\mathbf{d} + \tilde{\mathbf{A}}_{Sd}^T \tilde{\mathbf{A}}_{Pr}^{-T}[\ln \boldsymbol{\xi}_{Pr}]\right)_i = \ln \overline{\xi}_i^\alpha,$$

  which implies that $\boldsymbol{\xi}^\alpha = \overline{\boldsymbol{\xi}}^\alpha$.

□

With these two lemmas, we can now demonstrate Proposition 3.2.

Proof of Proposition 3.2. In Lemma 3.4, we have proved that under assumption (H2), the vector $\boldsymbol{\xi}$ is unique. The uniqueness of $(r_\alpha)_{\alpha=1,\,...,\,N_{Ph}}$ follows from (12c). To demonstrate the uniqueness of $(s_\alpha)_{\alpha=1,\,...,\,N_{Ph}}$, let $(\overline{s}_\alpha)_{\alpha=1,\,...,\,N_{Ph}}$ be another solution. According to equation (12a), we can derive the following relation:

$$\sum_{\alpha\in\mathit{\Gamma}_{\boldsymbol{\mathcal{X}}}} (s_\alpha - \overline{s}_\alpha)\mathbf{A}^\alpha\boldsymbol{\xi}^\alpha = \mathbf{0}.$$

Therefore the uniqueness is satisfied since $\{\mathbf{A}^\alpha\boldsymbol{\xi}^\alpha\}_{\alpha\in\mathit{\Gamma}_{\boldsymbol{\mathcal{X}}}}$ is linearly independent. □

4 Efficient resolution with parametrized Newton's method

This section focuses on the development of numerical algorithms designed to address the complexities of multiphase chemical equilibrium problems. The core objective is to determine the variables $(\boldsymbol{\xi}^\alpha,\, s_\alpha,\, r_\alpha)_{\alpha=1,\,...,\,N_{Ph}}$ that satisfy the system of equations (12) using Newton's method on a new parametrized system which is computationally easier to solve. This system presents two primary challenges:

- Nonlinearities of the logarithm: the presence of logarithmic terms introduces stiff nonlinearity, complicating the convergence of iterative methods.
- Complementarity problem: the constraints inherent in the system create a complementarity problem (12d)–(12e), necessitating adapted techniques for resolution.

In the first part of this section, we recall the parametrization technique introduced in [18] which was successfully applied to single-phase chemical equilibrium in order to address the issues caused by the logarithmic terms.

In the second part, we propose an innovative approach, referred to as complementarity parametrization, that integrates the complementarity equations directly into the system through parametrization techniques. This approach offers a new framework for solving these problems while maintaining essential conditions along the Newton iterations.

In the third and final part, we establish the invertibility of the Jacobian matrix associated with the combined log and complementarity parametrizations at equilibrium, focusing on a simple two-phase system. This result provides a theoretical foundation for the robustness of the proposed numerical approach.

4.1 Addressing log nonlinearity with parametrization

The parametrization technique consists in introducing a fictitious variable τ_i for each species and two functions $X : \mathbb{R} \to \mathbb{R}$ and $Y : \mathbb{R} \to \mathbb{R}$ such that

$$Y(\tau_i) = \ln(X(\tau_i)).$$

These functions allow us to parameterize the graph

$$\left\{(\xi,\, y) \in \mathbb{R}^2 \;\middle|\; y = \ln\xi\right\},$$

which, up a rescaling and a shift, encodes the relation between the chemical potentials and the (extended) mole fractions. There are various possibilities to define these functions, but the choice made here is to use the *switch* function:

$$(X(\tau), Y(\tau)) = \begin{cases} (\exp(\tau),\, \tau), & \text{if} \quad \tau < 0, \\ (\tau + 1,\, \ln(\tau + 1)), & \text{if} \quad \tau \geq 0. \end{cases} \tag{46}$$

Applying this parametrization to (12a)–(12c) gives:

$$\sum_{\alpha=1}^{N_{Ph}} s_\alpha\mathbf{A}^\alpha\mathbf{X}(\boldsymbol{\tau}^\alpha) - b = 0, \tag{47a}$$

$$\mathbf{S}^T[\mu_i^\circ/(\mathrm{RT}) + Y(\tau_i)]_{i=1,\,...,\,N} = \mathbf{0}, \tag{47b}$$

$$\langle \mathbf{X}(\boldsymbol{\tau}^{\alpha}),\, \mathbf{1}\rangle + r_{\alpha} - 1 = 0, \qquad (\alpha = 1,\, \ldots,\, N_{Ph}). \tag{47c}$$

The associated block in the Jacobian is written as:

$$\begin{bmatrix} s_1\mathbf{A}^1\mathrm{diag}\{\mathbf{X}'(\boldsymbol{\tau}^1)\} & \cdots & s_{N_{Ph}}\mathbf{A}^{N_{Ph}}\mathrm{diag}\{\mathbf{X}'(\boldsymbol{\tau}^{N_{Ph}})\} & \mathbf{AX}(\boldsymbol{\tau}) & \mathbf{0} \\ & & \mathbf{S}^T\mathrm{diag}\{Y'(\boldsymbol{\tau})\} & \mathbf{0} & \mathbf{0} \\ X'(\boldsymbol{\tau}^1)^T & & & & \\ & \ddots & & \mathbf{0} & \mathbf{I}_{N_{Ph}} \\ & & X'(\boldsymbol{\tau}^{N_{Ph}})^T & & \end{bmatrix}.$$

More details about applying parametrization techniques to chemical equilibria can be found in [18].

4.2 Solving the complementarity problem with parametrization

The complementarity problem associated with system (12) is defined by the following conditions:

$$s_{\alpha} r_{\alpha} = 0 \quad \text{and} \quad s_{\alpha},\, r_{\alpha} \geq 0. \tag{48}$$

This formulation indicates that for each pair of non-negative variables s_{α} and r_{α}, at least one of them must be zero, which is a characteristic of complementarity conditions. Complementarity problems encompass a broad class of mathematical optimization challenges and have generated extensive literature, as noted in works such as those by Acary and Brogliato [1] and Facchinei and Pang [12, 13]. Graphically, the complementarity problem can be visualized as the intersection of the non-negative axes in a Cartesian coordinate system, specifically represented by the two demi-axes $\{s_{\alpha} \in \mathbb{R} \mid s_{\alpha} \geq 0\}$ and $\{r_{\alpha} \in \mathbb{R} \mid r_{\alpha} \geq 0\}$.

The resolution of the complementarity problem using Newton's method presents two significant challenges:

- Non-differentiability at the origin: the case where $s_{\alpha} = r_{\alpha}$ is not differentiable, which complicates the definition of the Jacobian for the system.
- Nonnegativity constraint violation: the nonnegativity constraint is not inherently preserved by the Newton iterates, which can lead to solutions that fall outside the feasible region.

In this paper, we propose a novel approach to the complementarity problem based on parametrization techniques. In this method, referred to as complementarity parametrization method, the complementarity equations are integrated into the system, and the relationships between the complementarity variables are maintained through a parameter and two parametric functions. This approach provides a new framework for solving complementarity problems while ensuring that the essential conditions are met throughout the iterative process.

The complementarity problem (48) can be described using the min function as

$$\min(s_{\alpha},\, r_{\alpha}) = 0.$$

Similarly to our approach of parametrization for the relationship $y = \ln \xi$, the graph

$$\mathcal{T} = \{(s,\, r) \in \mathbb{R}^2 \mid \min(s,\, r) = 0\}$$

is parameterized by two monotonic piecewise continuously differentiable functions, $S : \mathbb{R} \to \mathbb{R}$ and $R : \mathbb{R} \to \mathbb{R}$, such that

$$s_{\alpha} = S(\eta_{\alpha}) \quad \text{and} \quad r_{\alpha} = R(\eta_{\alpha}).$$

These functions are defined to satisfy:

$$\min(S(\eta),\, R(\eta)) = 0,$$

which implies that $\mathcal{T} = (S,\, R)(\mathbb{R})$. To ensure proper parametrization and avoid singularities, the functions S and R must satisfy the following conditions, there exists $M > 0$, $\varepsilon > 0$ such that for each $\eta \in \mathbb{R}$:

(PC1) $S(\eta)R(\eta) = 0$ and $S(\eta),\, R(\eta) \geq 0$ with S and R monotonic and M-Lipschitz continuous;

(PC2) the derivatives S' and R' are bounded and piecewise continuous with $0 \leq S'(\eta) \leq M$ and $-M \leq R'(\eta) \leq 0$;

(PC3) $S'(\eta) - R'(\eta) > \varepsilon$.

We define a parametrization as admissible if it fulfills conditions (PC1)–(PC3). To define these specific functions, we require that S and R are both monotonic and M-Lipschitz continuous, which implies that they are differentiable almost everywhere, as established by Rademacher's theorem. For the points at which S and R are not differentiable, we assign specific values to S' and R'. In practice, S' and R' exhibit piecewise continuity. Furthermore, we ensure that $|S'|$ and $|R'|$ are upper semi-continuous by choosing the largest possible value (in magnitude) for both S' and R'. We also impose a normalization condition:

$$\max(|S'(\eta)|, |R'(\eta)|) = 1, \quad \forall \eta \in \mathbb{R}.$$

Condition (PC1)–(PC3) implies the existence of a point $\eta^0 \in \mathbb{R}$ such that $S(\eta^0) = R(\eta^0) = 0$. By selecting $\eta^0 = 0$ and enforcing $S' \geq 0$, $R' \leq 0$, we derive the following implications:

$$\begin{aligned} \eta > 0 \quad &\Rightarrow \quad S(\eta) > 0 \text{ and } R(\eta) = 0 \quad \Rightarrow \quad R'(\eta) = 0, \\ \eta < 0 \quad &\Rightarrow \quad S(\eta) = 0 \text{ and } R(\eta) > 0 \quad \Rightarrow \quad S'(\eta) = 0. \end{aligned}$$

From the normalization condition, we have:

$$\max(|S'(\eta)|, |R'(\eta)|) = 1 \quad \Leftrightarrow \quad \max(S'(\eta), -R'(\eta)) = 1$$

This leads to the following conclusions:

$$\begin{aligned} \eta > 0 \quad &\Rightarrow \quad S'(\eta) = 1 \quad \Rightarrow \quad S(\eta) = \eta, \\ \eta < 0 \quad &\Rightarrow \quad R'(\eta) = -1 \quad \Rightarrow \quad R(\eta) = -\eta. \end{aligned}$$

Thus, we define the complementarity parametrization method as follows:

$$S(\eta) = \begin{cases} 0 & \text{if } \eta < 0, \\ \eta & \text{if } \eta \geq 0, \end{cases} \quad \text{and} \quad R(\eta) = \begin{cases} -\eta & \text{if } \eta \leq 0, \\ 0 & \text{if } \eta > 0. \end{cases}$$

These functions are illustrated in Figure 1. Furthermore, the derivatives of these functions are defined as:

$$S'(\eta) = \begin{cases} 0 & \text{if } \eta < 0, \\ 1 & \text{if } \eta \geq 0, \end{cases} \quad \text{and} \quad R'(\eta) = \begin{cases} -1 & \text{if } \eta \leq 0, \\ 0 & \text{if } \eta > 0. \end{cases}$$

At the point $\eta = 0$, we have chosen to set both $|S'(0)|$ and $|R'(0)|$ equal to 1 to ensure the invertibility of the Jacobian matrix (see Proposition 4.1).

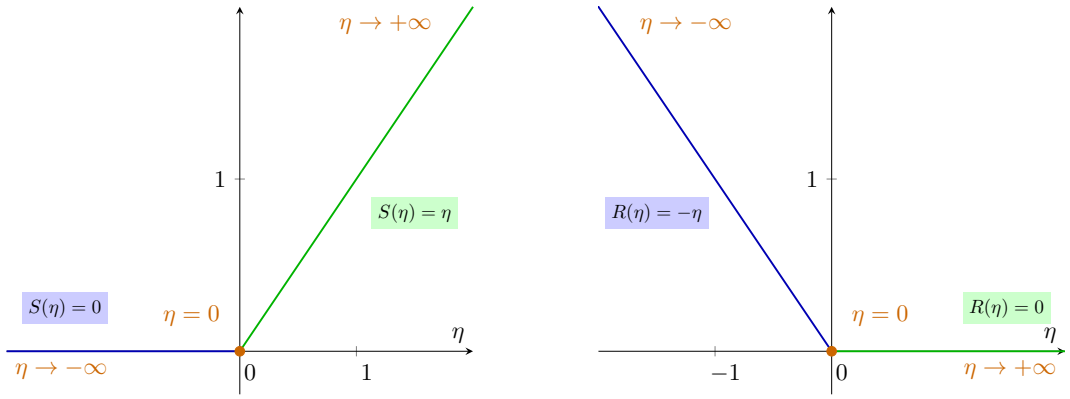

Figure 1: Functions S and R for the complementarity parametrization method.

These functions lead to a reformulation of system (12) which becomes: find $(\boldsymbol{\xi}^\alpha, \eta_\alpha)_{\alpha=1,\ldots,N_{Ph}}$ such that:

$$\begin{aligned} \sum_{\alpha=1}^{N_{Ph}} S(\eta_\alpha)\mathbf{A}^\alpha\boldsymbol{\xi}^\alpha - b = 0, \\ \mathbf{S}^T[\mu_i^\circ/(\mathrm{RT}) + \ln \xi_i]_{i=1,\ldots,N} = \mathbf{0}, \\ \langle \boldsymbol{\xi}^\alpha, \mathbf{1}\rangle + R(\eta_\alpha) - 1 = 0, \quad (\alpha = 1, \ldots, N_{Ph}). \end{aligned}$$

Using the parametrization of the log, the system to solve is: find $(\boldsymbol{\tau}^\alpha,\, \eta_\alpha)_{\alpha=1,\,\ldots,\,N_{Ph}}$ such that:

$$\begin{aligned}
\sum_{\alpha=1}^{N_{Ph}} S(\eta_\alpha)\mathbf{A}^\alpha\mathbf{X}(\boldsymbol{\tau}^\alpha) - b = 0,\\
\mathbf{S}^T[\mu_i^\circ/(\mathrm{RT}) + Y(\tau_i)]_{i=1,\,\ldots,\,N} = \mathbf{0},\\
\langle \mathbf{X}(\boldsymbol{\tau}^\alpha),\, \mathbf{1}\rangle + R(\eta_\alpha) - 1 = 0, \quad (\alpha = 1,\, \ldots,\, N_{Ph}).
\end{aligned}$$

The resulting Jacobian is given by:

$$\begin{bmatrix}
S(\eta_1)\mathbf{A}^1\mathrm{diag}\{\mathbf{X}'(\boldsymbol{\tau}^1)\} & \cdots & S(\eta_{N_{Ph}})\mathbf{A}^{N_{Ph}}\mathrm{diag}\{\mathbf{X}'(\boldsymbol{\tau}^{N_{Ph}})\} & S'(\eta_1)\mathbf{A}^1\mathbf{X}(\boldsymbol{\tau}^1) & \cdots & S'(\eta_{N_{Ph}})\mathbf{A}^{N_{Ph}}\mathbf{X}(\boldsymbol{\tau}^{N_{Ph}})\\
\multicolumn{3}{c}{\mathbf{S}^T\mathrm{diag}\{Y'(\boldsymbol{\tau})\}} & \multicolumn{3}{c}{\mathbf{0}}\\
\mathbf{X}'(\boldsymbol{\tau}^1)^T & & & R'(\eta_1) & & \\
& \ddots & & & \ddots & \\
& & \mathbf{X}'(\boldsymbol{\tau}^{N_{Ph}})^T & & & R'(\eta_{N_{Ph}})
\end{bmatrix}.$$

4.3 Invertibility of the complementarity parametrization system

We consider the system combining the logarithmic parametrization with the complementarity parametrization method. To study the invertibility of the associated Jacobian, we focus on the simpler two-phase case: find $\boldsymbol{\tau} = (\boldsymbol{\tau}^\alpha,\, \boldsymbol{\tau}^\beta)$ and $\boldsymbol{\eta} = (\eta_\alpha,\, \eta_\beta)$ such that:

$$\begin{aligned}
\mathbf{A}\begin{bmatrix} S(\eta_\alpha)\mathbf{X}(\boldsymbol{\tau}^\alpha)\\ S(\eta_\beta)\mathbf{X}(\boldsymbol{\tau}^\beta)\end{bmatrix} - \mathbf{b} = \mathbf{0}\\
\mathbf{S}^T\begin{bmatrix}\mathbf{Y}(\boldsymbol{\tau}^\alpha)\\ \mathbf{Y}(\boldsymbol{\tau}^\beta)\end{bmatrix} + \mathbf{S}^T\boldsymbol{\mu}^\circ/(\mathrm{RT}) = \mathbf{0}\\
\mathbf{1}^T\mathbf{X}(\boldsymbol{\tau}^\alpha) + R(\eta_\alpha) - 1 = 0\\
\mathbf{1}^T\mathbf{X}(\boldsymbol{\tau}^\beta) + R(\eta_\beta) - 1 = 0
\end{aligned} \tag{49}$$

We now prove the invertibility of the Jacobian $\mathbf{J}(\boldsymbol{\tau},\, \boldsymbol{\eta})$ under the assumptions stated earlier.

Proposition 4.1. *Assuming (H2) holds and the uniqueness condition from Proposition* 3.2 *is satisfied, let* $(\boldsymbol{\tau},\, \boldsymbol{\eta})$ *denote the unique solution to* (49)*. Then, the Jacobian matrix* $\mathbf{J}(\boldsymbol{\tau},\, \boldsymbol{\eta})$ *is nonsingular.*

Proof. Let $\delta\boldsymbol{\tau} = (\delta\boldsymbol{\tau}^\alpha,\, \delta\boldsymbol{\tau}^\beta) \in \mathbb{R}^N$ and $\delta\boldsymbol{\eta} = (\delta\eta_\alpha,\, \delta\eta_\beta) \in \mathbb{R}^2$ such that $\mathbf{J}(\boldsymbol{\tau},\, \boldsymbol{\eta})(\delta\boldsymbol{\tau},\, \delta\boldsymbol{\eta})^T = \mathbf{0}$, then:

$$\mathbf{A}\begin{bmatrix} S(\eta_\alpha)\mathrm{diag}\{\mathbf{X}'(\boldsymbol{\tau}^\alpha)\}\delta\boldsymbol{\tau}^\alpha + S'(\eta_\alpha)\mathbf{X}(\boldsymbol{\tau}^\alpha)\delta\eta_\alpha\\ S(\eta_\beta)\mathrm{diag}\{\mathbf{X}'(\boldsymbol{\tau}^\beta)\}\delta\boldsymbol{\tau}^\beta + S'(\eta_\beta)\mathbf{X}(\boldsymbol{\tau}^\beta)\delta\eta_\beta\end{bmatrix} = \mathbf{0}, \tag{50a}$$

$$\mathbf{S}^T\begin{bmatrix}\mathrm{diag}\{\mathbf{Y}'(\boldsymbol{\tau}^\alpha)\}\delta\boldsymbol{\tau}^\alpha\\ \mathrm{diag}\{\mathbf{Y}'(\boldsymbol{\tau}^\beta)\}\delta\boldsymbol{\tau}^\beta\end{bmatrix} = \mathbf{0}, \tag{50b}$$

$$\mathbf{X}'(\boldsymbol{\tau}^\alpha)^T\delta\boldsymbol{\tau}^\alpha + R'(\eta_\alpha)\delta\eta_\alpha = 0, \tag{50c}$$

$$\mathbf{X}'(\boldsymbol{\tau}^\beta)^T\delta\boldsymbol{\tau}^\beta + R'(\eta_\beta)\delta\eta_\beta = 0. \tag{50d}$$

From (4), we have $\ker\mathbf{S}^T = (\ker\mathbf{A})^\perp$. Taking the scalar product of the vector in (50a) with the vector in (50b), we get:

$$\begin{aligned}
&\left\langle \begin{pmatrix} S(\eta_\alpha)\mathrm{diag}\{\mathbf{X}'(\boldsymbol{\tau}^\alpha)\}\delta\boldsymbol{\tau}^\alpha\\ S(\eta_\beta)\mathrm{diag}\{\mathbf{X}'(\boldsymbol{\tau}^\beta)\}\delta\boldsymbol{\tau}^\beta\end{pmatrix},\, \begin{pmatrix}\mathrm{diag}\{\mathbf{Y}'(\boldsymbol{\tau}^\alpha)\}\delta\boldsymbol{\tau}^\alpha\\ \mathrm{diag}\{\mathbf{Y}'(\boldsymbol{\tau}^\beta)\}\delta\boldsymbol{\tau}^\beta\end{pmatrix}\right\rangle\\
&= -\left\langle \begin{pmatrix} S'(\eta_\alpha)\mathbf{X}(\boldsymbol{\tau}^\alpha)\delta\eta_\alpha\\ S'(\eta_\beta)\mathbf{Y}(\boldsymbol{\tau}^\beta)\delta\eta_\beta\end{pmatrix},\, \begin{pmatrix}\mathrm{diag}\{\mathbf{Y}'(\boldsymbol{\tau}^\alpha)\}\delta\boldsymbol{\tau}^\alpha\\ \mathrm{diag}\{\mathbf{Y}'(\boldsymbol{\tau}^\beta)\}\delta\boldsymbol{\tau}^\beta\end{pmatrix}\right\rangle.
\end{aligned} \tag{51}$$

Using the identity $\mathbf{X}'(\boldsymbol{\tau}) = \mathrm{diag}\{\mathbf{Y}'(\boldsymbol{\tau})\}\mathbf{X}(\boldsymbol{\tau})$, the right-hand side of (51) becomes:

$$-\left\langle \begin{pmatrix} S'(\eta_\alpha)\mathbf{X}'(\boldsymbol{\tau}^\alpha)\delta\eta_\alpha\\ S'(\eta_\beta)\mathbf{X}'(\boldsymbol{\tau}^\beta)\delta\eta_\beta\end{pmatrix},\, \delta\boldsymbol{\tau}\right\rangle. \tag{52}$$

We now distinguish three cases:

Case 1: Both α and β are present. Here, η_α, $\eta_\beta > 0$ and we have:

$$S(\eta_\alpha),\, S(\eta_\beta) > 0, \quad S'(\eta_\alpha) = S'(\eta_\beta) = 1, \quad R(\eta_\alpha) = R(\eta_\beta) = 0, \quad R'(\eta_\alpha) = R'(\eta_\beta) = 0.$$

From (50c) and (50d), we get:

$$\mathbf{X}'(\boldsymbol{\tau}^\alpha)^T \delta\boldsymbol{\tau}^\alpha = 0 \quad \text{and} \quad \mathbf{X}'(\boldsymbol{\tau}^\beta)^T \delta\boldsymbol{\tau}^\beta = 0.$$

Plugging into (52) and then (51), we obtain:

$$\Bigg\langle \underbrace{\begin{pmatrix} S(\eta_\alpha)\mathrm{diag}\{\mathbf{X}'(\boldsymbol{\tau}^\alpha)\}\mathrm{diag}\{\mathbf{Y}'(\boldsymbol{\tau}^\alpha)\}\delta\boldsymbol{\tau}^\alpha & \mathbf{0} \\ \mathbf{0} & S(\eta_\beta)\mathrm{diag}\{\mathbf{X}'(\boldsymbol{\tau}^\beta)\}\mathrm{diag}\{\mathbf{Y}'(\boldsymbol{\tau}^\beta)\} \end{pmatrix}}_{:=\mathbf{D}} \delta\boldsymbol{\tau},\, \delta\boldsymbol{\tau} \Bigg\rangle = 0,$$

where $\mathbf{D}$ is positive definite since $\mathbf{X}'(\boldsymbol{\tau})$, $\mathbf{Y}'(\boldsymbol{\tau})$, $\mathbf{S}(\boldsymbol{\eta}) > 0$. Thus, $\delta\boldsymbol{\tau} = \mathbf{0}$. Equation (50a) then gives:

$$\mathbf{A} \begin{bmatrix} \mathbf{X}(\boldsymbol{\tau}^\alpha)\delta\eta_\alpha \\ \mathbf{X}(\boldsymbol{\tau}^\beta)\delta\eta_\beta \end{bmatrix} = \mathbf{0} \quad \Leftrightarrow \quad \mathbf{A}^\alpha \mathbf{X}(\boldsymbol{\tau}^\alpha)\delta\eta_\alpha + \mathbf{A}^\beta \mathbf{X}(\boldsymbol{\tau}^\beta)\delta\eta_\beta = \mathbf{0}.$$

According to Proposition 3.2, the vectors $\{\mathbf{A}^\alpha \mathbf{X}(\boldsymbol{\tau}^\alpha),\, \mathbf{A}^\beta \mathbf{X}(\boldsymbol{\tau}^\beta)\}$ form a linearly independent set. Consequently, we must have $\delta\eta_\alpha = \delta\eta_\beta = 0$.

Case 2: Phase α is present, β is absent with $R(\eta_\beta) > 0$. We have $\eta_\alpha > 0$, $\eta_\beta < 0$, then

$$S(\eta_\alpha) > 0,\, S(\eta_\beta) = 0, \quad S'(\eta_\alpha) = 1,\, S'(\eta_\beta) = 0, \quad R(\eta_\alpha) = 0,\, R(\eta_\beta) > 0, \quad R'(\eta_\alpha) = 0,\, R'(\eta_\beta) = -1.$$

From (50c) we again have: $\mathbf{X}'(\boldsymbol{\tau}^\alpha)^T \delta\boldsymbol{\tau}^\alpha = 0$ and (52) vanishes. Then (51) yields:

$$\Bigg\langle \underbrace{S(\eta_\alpha)\mathrm{diag}\{\mathbf{X}'(\boldsymbol{\tau}^\alpha)\}\mathrm{diag}\{\mathbf{Y}'(\boldsymbol{\tau}^\alpha)\}}_{:=\mathbf{D}_\alpha} \delta\boldsymbol{\tau}^\alpha,\, \delta\boldsymbol{\tau}^\alpha \Bigg\rangle = 0,$$

where $\mathbf{D}_\alpha$ is positive definite since $\mathbf{X}'(\boldsymbol{\tau}^\alpha)$, $\mathbf{Y}'(\boldsymbol{\tau}^\alpha)$, $\mathbf{S}(\eta_\alpha) > 0$. It follows that $\delta\boldsymbol{\tau}^\alpha = \mathbf{0}$. Under assumption (H2), the set of vectors $\mathbf{A}_\alpha = [\mathbf{a}_i]_{i\in\sigma^{-1}(\alpha)}$ is a spanning set in $\mathbb{R}^M$. In particular, $\mathbf{A}_{Pr}$ is included into $\mathbf{A}_\alpha$ and from the structure

$$\mathbf{S}^T = \left[(\mathbf{A}_{Pr}^{-1}\mathbf{A}_{Sd})^T,\, -\mathbf{I}_{Sd}\right],$$

it follows that (50b) becomes:

$$-\mathrm{diag}\{\mathbf{Y}'(\boldsymbol{\tau}^\beta)\}\delta\boldsymbol{\tau}^\beta = \mathbf{0}.$$

Since $\mathbf{Y}'(\boldsymbol{\tau}^\beta) > 0$, we have $\delta\boldsymbol{\tau}^\beta = \mathbf{0}$. Then, from (50d), it follows that $\delta\eta_\beta = 0$. Finally, equation (50a) reduces to:

$$\mathbf{A} \begin{bmatrix} \mathbf{X}(\boldsymbol{\tau}^\alpha)\delta\eta_\alpha \\ \mathbf{0} \end{bmatrix} = \mathbf{0},$$

implying $\delta\eta_\alpha = 0$ by Proposition 3.2.

Case 3: Phase α is present, β is absent with $R(\eta_\beta) = 0$. Here, $\eta_\alpha > 0$, $\eta_\beta = 0$ and

$$S(\eta_\alpha) > 0,\, S(\eta_\beta) = 0, \quad S'(\eta_\alpha) = 1,\, S'(\eta_\beta) = 0, \quad R(\eta_\alpha) = 0,\, R(\eta_\beta) = 0, \quad R'(\eta_\alpha) = 0, = R'(\eta_\beta) = -1.$$

Equations (50c) and (50d) become:

$$\mathbf{X}'(\boldsymbol{\tau}^\alpha)^T \delta\boldsymbol{\tau}^\alpha = 0 \quad \text{and} \quad \mathbf{X}'(\boldsymbol{\tau}^\beta)^T \delta\boldsymbol{\tau}^\beta = \delta\eta_\beta.$$

Then from (51) and (52), we obtain:

$$\langle \mathbf{D}_\alpha \delta\boldsymbol{\tau}^\alpha,\, \delta\boldsymbol{\tau}^\alpha \rangle = -\left(\delta\eta_\beta\right)^2,$$

which implies $\delta\boldsymbol{\tau}^\alpha = \mathbf{0}$, and $\delta\eta_\beta = 0$ since $\mathbf{D}_\alpha$ is positive definite. From (50a), we get:

$$\mathbf{A} \begin{bmatrix} \mathbf{X}(\boldsymbol{\tau}^\alpha)\delta\eta_\alpha \\ \mathbf{0} \end{bmatrix} = \mathbf{0} \quad \Rightarrow \quad \delta\eta_\alpha = 0$$

thanks to Proposition 3.2. Reasoning analogously to Case 2, we have:

$$-\mathrm{diag}\{\mathbf{Y}'(\boldsymbol{\tau}^\beta)\}\delta\boldsymbol{\tau}^\beta = \mathbf{0} \quad \Rightarrow \quad \delta\boldsymbol{\tau}^\beta = \mathbf{0}.$$

In all cases, the only solution is the trivial one. Hence, the Jacobian is nonsingular.

□

5 Numerical experiments

The following section presents a comprehensive study of numerical experiments designed to evaluate computational methods for solving chemical equilibrium problems in multiphase systems. We examine three test cases of increasing complexity: (1) the simple silica (SiO_2) system already introduced as an illustration in Section 2.1, (2) a carbon dioxide (CO_2) system involving aqueous, mineral, and gaseous phases, and (3) a complex seawater system with numerous species and minerals.

The primary objective of this study is to compare our proposed methods—the parametrized approach for the log formulation and the complementarity parametrization strategy—with established methods from the literature, namely the log-trick for the log formulation and the minimum function (Newton-min), Fischer-Burmeister function (FB), and Interior Point Method (IPM) for the complementarity problem. Details of the literature methods are provided in Appendix C.

Our analysis focuses on the performance and robustness of these methods for chemical equilibrium calculations. We pay particular attention to convergence behavior, iteration counts, and the challenges posed by complementarity conditions in equilibrium modeling. The results provide critical insights into the efficiency and reliability of these methods across diverse chemical scenarios.

Numerical parameters

The functions X and Y used in the complementarity parametrization are those defined by the switching mechanism in Eq. (46). For the IPM, we use the following parameter values: $\kappa_\nu = 0.5$, $\theta_\nu = 2$, and $\gamma_{\min} = 0.1$.

Let $\mathcal{F}(\boldsymbol{\mathcal{X}})$ denote the residual function whose root we seek. The convergence criterion for Newton's method is defined as:

$$\|\mathcal{F}(\boldsymbol{\mathcal{X}}^{(k+1)})\|_\infty \leq 10^{-10} \quad \text{and} \quad \|\boldsymbol{\mathcal{X}}^{(k+1)} - \boldsymbol{\mathcal{X}}^{(k)}\|_\infty \leq 10^{-10},$$

where $k+1$ denotes the current Newton iteration.

In the IPM framework, the complementarity variables s_α and r_α remain strictly positive and never reach zero. This poses challenges when modeling equilibrium conditions that require both variables to be simultaneously zero. To address this, we apply a stricter convergence criterion to the complementarity residual:

$$\|\mathcal{F}(\boldsymbol{\mathcal{X}}^{(k)})\|_\infty \leq 10^{-10} \quad \text{and} \quad \|\mathcal{F}_{\text{compl}}(\boldsymbol{\mathcal{X}}^{(k)})\|_\infty \leq \epsilon_{64}^2,$$

where $\epsilon_{64} \approx 2.22 \times 10^{-16}$ is the machine epsilon in double-precision floating-point arithmetic.

Initialization of Newton's Method.

We initialize the Newton solver using the following procedure. The initial vectors of quantities are defined as:

$$n_{\text{aq}} = (\mathbf{b}_{\text{O}}, 1, \ldots, 1), \quad n_{\min} = 1 \text{ for all minerals}, \quad n_{\text{gas}} = (1, \ldots, 1),$$

where $\mathbf{b}_{\text{O}}$ is the fixed oxygen content, serving as an approximation for the aqueous H_2O quantity. The total amount in each phase is computed as:

$$s_{\text{aq}} = \langle n_{\text{aq}}, \mathbf{1} \rangle, \quad s_{\min} = n_{\min} \text{ for all minerals}, \quad s_{\text{gas}} = \langle n_{\text{gas}}, \mathbf{1} \rangle.$$

This leads to the initial values for the phase composition vector $\boldsymbol{\xi}$:

$$\boldsymbol{\xi}_{\text{aq}} = n_{\text{aq}}/s_{\text{aq}}, \quad \boldsymbol{\xi}_{\min} = 1 \text{ for all minerals}, \quad \boldsymbol{\xi}_{\text{gas}} = n_{\text{gas}}/s_{\text{gas}}.$$

The auxiliary variable $\mathbf{r}$ is initialized uniformly:

$$r_{\text{aq}} = 1, \quad r_{\min} = 1 \text{ for all minerals}, \quad r_{\text{gas}} = 1.$$

For the complementarity parametrization method, the variable $\boldsymbol{\eta}$ is initialized as:

$$\eta_{\text{aq}} = s_{\text{aq}}, \quad \eta_{\min} = -s_{\min} \text{ for all minerals}, \quad \eta_{\text{gas}} = -s_{\text{gas}}.$$

Finally, the parameter ν for IPM is initialized as:

$$\nu = \max\left\{ \min_{\alpha \in \{1, \ldots, N_{Ph}\}} (s_\alpha r_\alpha) - 0.1,\ 0.5 \right\}.$$

Description of test cases

We consider three chemical equilibrium test cases of increasing complexity:

1. SiO_2 system: this simple system consists of four aqueous species:

$$H_2O,\ H^+,\ SiO_2(aq),\ OH^-,$$

and one mineral phase: quartz (SiO_2(quartz)).

2. CO_2 system: this system comprises ten species, divided among three phases:
   - Aqueous phase:
   $$H_2O,\ Ca^{2+},\ CO_2(aq),\ H^+,\ OH^-,\ HCO_3^-,\ CO_3^{2-};$$
   - Mineral phase: calcite ($CaCO_3$(calcite));
   - Gaseous phase: carbon dioxide (CO_2(g)) and water vapor ($H_2O(g)$).

3. Multiphase Seawater: this complex system includes:
   - 50 aqueous species;
   - 20 minerals, each in its own pure phase;
   - 2 gaseous species.

Detailed compositions and thermodynamic data for all three test cases are provided in Appendix A.

Numerical evaluation

In the two following subsections, we assess the proposed algorithms on the three representative test cases of increasing complexity presented above.

The SiO_2 case, a small-scale chemical system, allows detailed examination of convergence behavior and local phenomena, providing insights into algorithmic mechanisms impractical to obtain in larger systems.

The CO_2 and multiphase Seawater cases follow the same robustness assessment methodology, differing primarily in system complexity. The CO_2 system, relatively simple, serves as a reference for evaluating algorithmic reliability under varied initial guesses and perturbed data. The multiphase Seawater system, highly complex and multiphase, extends this evaluation to large-scale scenarios.

This progressive approach enables both a precise understanding of algorithmic behavior and a rigorous evaluation of practical performance in challenging settings.

5.1 SiO_2: small-scale convergence analysis

For this test case, we have run our algorithms with three different vectors $\mathbf{b}$ in order to obtain the following subcases:

- **A**: the mineral is present in a large amount;
- **B**: the mineral is absent with condition that are close to precipitation, indicating that both variables of the complementarity conditions vanish at equilibrium;
- **C**: the mineral is absent with conditions that are far from precipitation.

The results obtained from the various combinations of methods are presented in Table 1. The evolution of residuals for setting **A** is illustrated in Figure 2 and the evolution of IPM in setting **B** in Figure 7. Notably, for the methods that converge, local quadratic convergence is observed. These results enable us to draw the following conclusions:

1. In **B**, the IPM is very slow, due to the needed precision on the complementarity equations and to a linear convergence.
2. In **C**, the min function diverges with the parametrization formulation and the FB function is slow with the log-trick and parametrization formulations.
3. The complementarity parametrization method and the IPM approaches are robust. However, the convergence of the IPM is slower.

In the remainder of this section, we will conduct a comprehensive analysis of the diverse behaviors exhibited by algorithms.

Log formulation	Complementarity	Setting		
		A	**B**	**C**
Log-trick	**param**	11	9	13
	min	9	7	10
	FB	8	7	15
	IPM	9	52	12
Param	**param**	7	**6**	14
	min	**6**	**6**	×
	FB	7	**6**	14
	IPM	9	52	12

Table 1: Number of iterations for Newton's method.

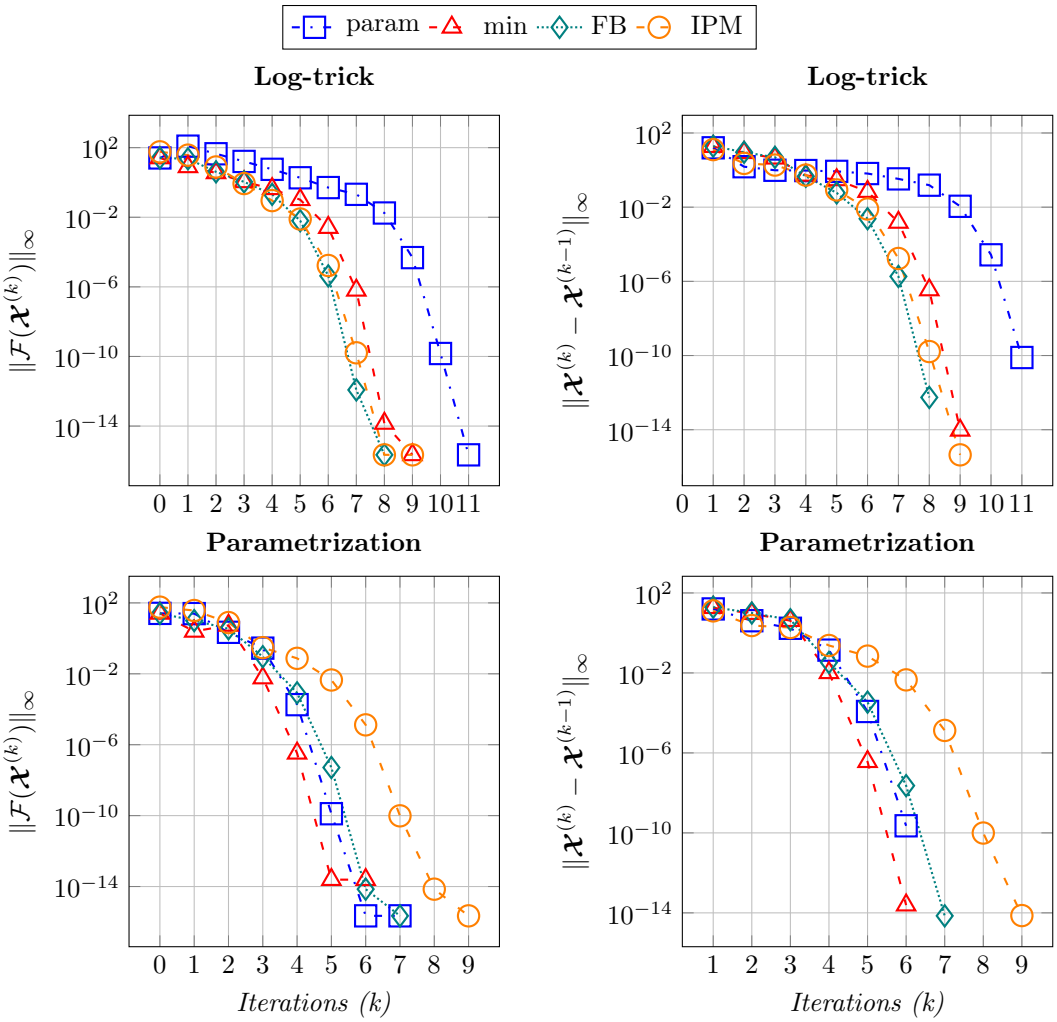

Figure 2: Evolution of residuals for the SiO_2 text case in configuration **A**.

Comparison between the log-trick and the parametrization with complementarity parametrization

In this part, we compare the iterates produced by the two log-based formulations—the log-trick and the parametrization method—in the context of complementarity-based parametrization. To this end, Figure 3 illustrates the evolution of the norm of the function $\mathcal{F}$. The left panel shows the residuals associated with the conservation and mole fraction sum equations, while the right panel focuses on the residuals of the equilibrium equations. Figures 4–5 display the evolution of the iterates for the species H_2O and *Quartz*, along with their mole fractions. Figure 6 presents the corresponding evolution for H^+, OH^-, and SiO_2(aq). Table 2 indicates the parametrization function used for H_2O at each iteration.

As shown in Figure 3, the decrease in the residual for the parametrization method can be attributed

to the complete resolution of the chemical equilibrium equation from the fourth iteration onward. Specifically, starting at the third iteration, the equilibrium equation becomes linear. It is well known that Newton's method solves linear equations exactly in a single step. According to Figure 6, the τ_i variables—excluding those for H_2O and $Quartz$—are consistently negative, which ensures that the $Y(\tau_i)$ values remain linear. Consequently, the linearity or non-linearity of the equilibrium equations depends solely on the values of τ_{H_2O} and τ_{Quartz}. Figure 4 and Table 2 show that τ_{H_2O} becomes negative starting from the third iteration, and Figure 5 shows that τ_{Quartz} becomes exactly zero at the same point. These conditions together yield a linear equilibrium equation from the third iteration onward.

Furthermore, Figure 5 shows that during the first iteration, the value of τ_{Quartz} from the parametrization method coincides with $\ln(x_{Quartz})$ obtained using the log-trick approach. However, the corresponding mole fractions, $X(\tau_{Quartz})$ and x_{Quartz}, differ because $X(\tau) = \tau+1$ for positive values of τ. This property prevents the mole fraction from becoming excessively large, resulting in a τ_{Quartz} value that is significantly closer to the solution by the second iteration compared to the log-trick. Consequently, in this test case, the parametrization method leads to faster convergence toward the solution.

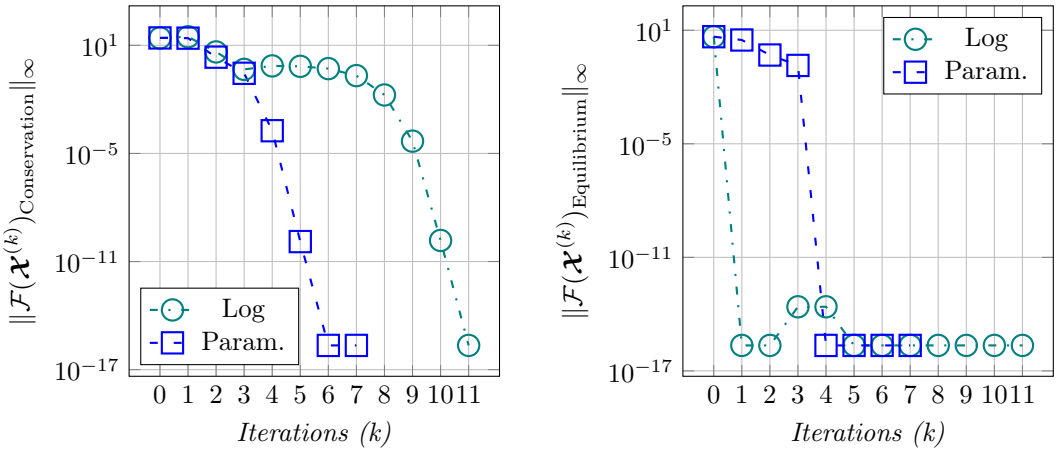

Figure 3: Evolution of the residuals for the log-trick and the parametrization methods in the SiO_2 test case. The left panel shows the residuals of the conservation equations, while the right panel shows those of the equilibrium equations.

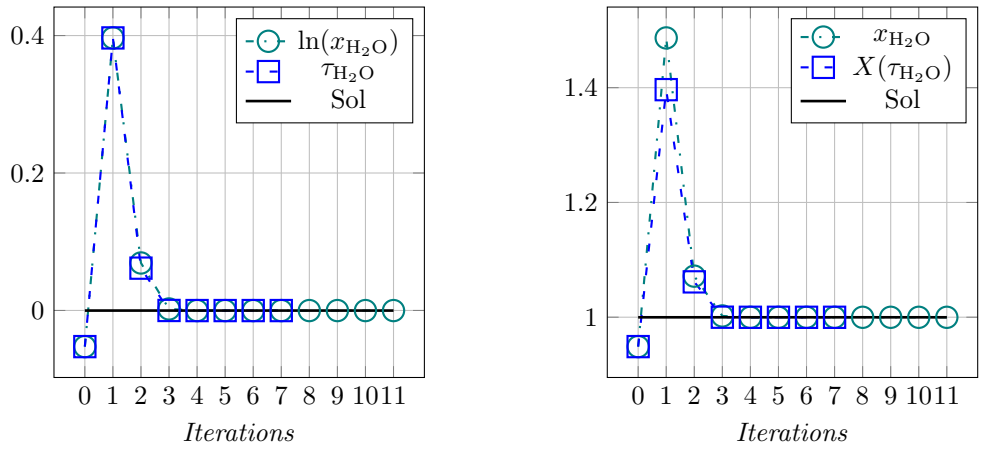

Figure 4: Evolution of the species H_2O with the log-trick and parametrization methods.

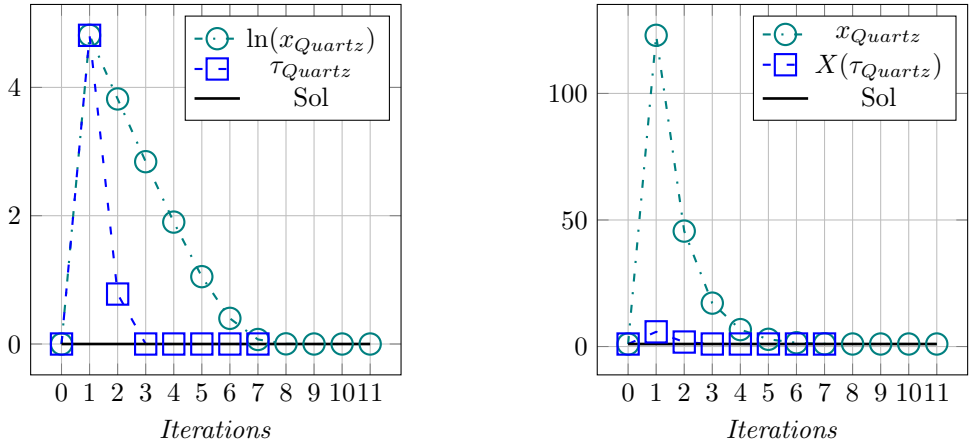

Figure 5: Evolution of the *Quartz* with the log-trick and parametrization methods.

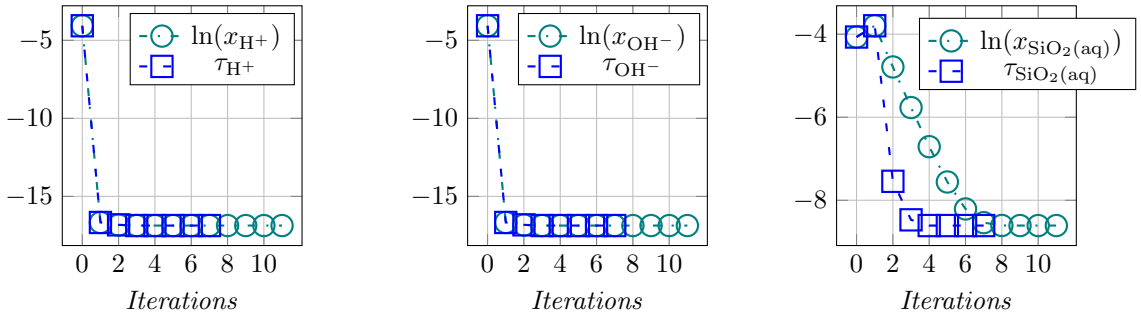

Figure 6: Evolution of the species H^+, OH^- and $SiO_2(aq)$ for the log trick and parametrization method.

Iteration	0	1	2	3	$\cdots$	7
$\text{sign}(\tau_{H_2O})$	$-$	$+$	$+$	$-$	$\cdots$	$-$
$X(\tau_{H_2O})$	$\exp \tau_{H_2O}$	$\tau_{H_2O}+1$	$\tau_{H_2O}+1$	$\exp \tau_{H_2O}$	$\cdots$	$\exp \tau_{H_2O}$
$Y(\tau_{H_2O})$	τ_{H_2O}	$\ln(\tau_{H_2O}+1)$	$\ln(\tau_{H_2O}+1)$	τ_{H_2O}	$\cdots$	τ_{H_2O}

Table 2: Evolution of $X(\tau_{H_2O})$ and $Y(\tau_{H_2O})$.

IPM slowdown in setting B

Test case **B** is characterized by a uniquely doubly active complementarity constraint associated with the mineral phase. As shown in Figure 7, the residual norm $|\mathcal{F}(\boldsymbol{\mathcal{X}}^{(k)})|\infty$ reaches the prescribed tolerance at iteration 21. Beyond this point, the residual exhibits a plateau near machine epsilon, persisting until the convergence criterion for the complementarity residual $|\mathcal{F}\text{compl}(\boldsymbol{\mathcal{X}}^{(k)})|_\infty$ is eventually satisfied. Notably, the convergence behavior transitions from quadratic to linear, indicating a degradation in the rate of convergence.

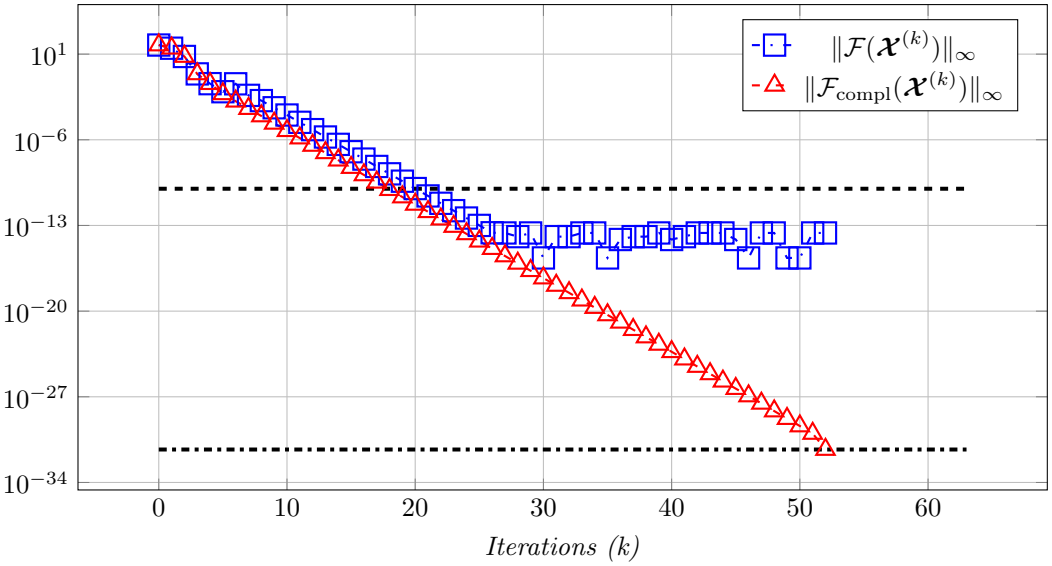

Figure 7: Evolution of residuals of IPM for setting **B** with the parametrization formulation.

Slowdown in setting C

The slowdown observed in setting **C**, as shown in Figure 8, can be attributed to a sharp decline in the concentrations of SiO_2 and Quartz, depicted in Figure 10 in the case of FB function. This abrupt variation fails to effectively reduce the residual, causing the iterates to undershoot the solution before undergoing a large compensatory correction that overshoots it. A comparable behavior occurs with the parametrization based on the min function (Figure 9); however, in this case, the correction fails to recover convergence. Specifically, Newton's method applied to the min-function parametrization terminates prematurely due to a singular Jacobian. At iteration 23, the variable n_{aq} is set to zero, leading to a loss of structural information in the Jacobian matrix (53). As a result, the matrix no longer maintains linear independence between rows 3 and 8, and the entries in row 3 approach zero, exacerbating the singularity.

$$\mathcal{J}(\boldsymbol{\mathcal{X}}^{(23)}) \approx \begin{pmatrix} 0.0 & 0.0 & 0.0 & 0.0 & -0.037 & 0.47 & 2.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & -0.018 & 0.00018 & 1.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 3\times10^{-22} & 0.0 & 0.0 & 0.0 \\ 1.0 & -1.0 & 0.0 & -1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 1.0 & 0.0 & -1.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.47 & 4\times10^{-8} & 0.00018 & 4\times10^{-8} & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 1.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 \end{pmatrix} \tag{53}$$

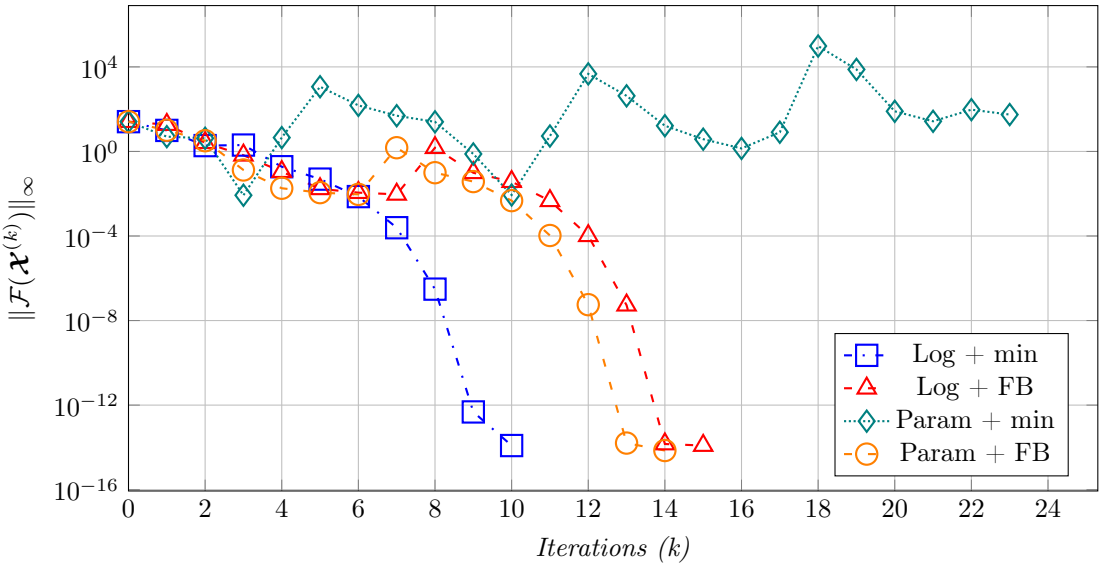

Figure 8: Evolution of residuals in setting **C**.

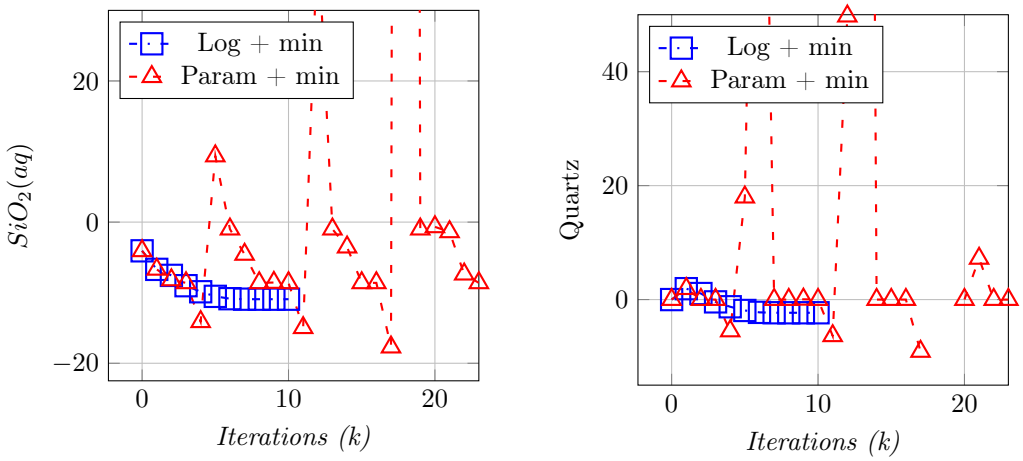

Figure 9: Evolution of the concentrations of SiO_2(aq) and Quartz in setting **C** with the min function.

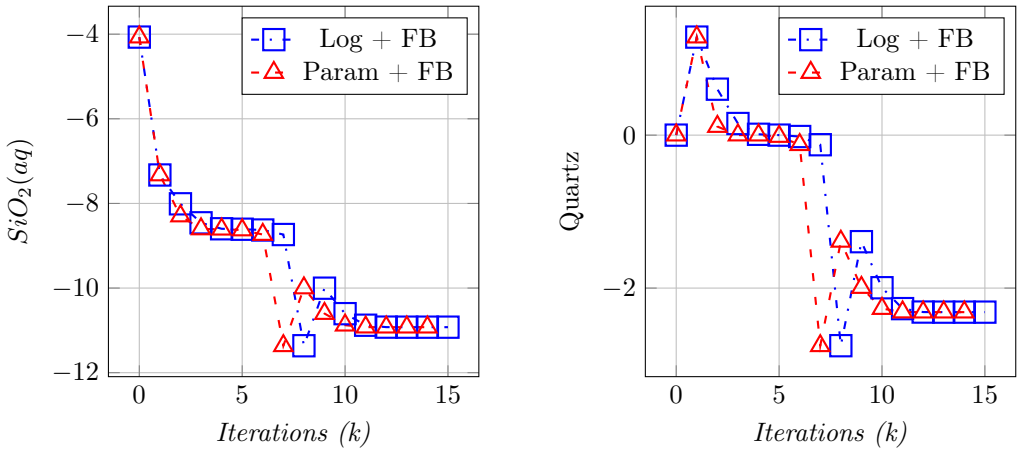

Figure 10: Evolution of the concentrations of SiO_2(aq) and Quartz in setting **C** with the FB function.

5.2 Robustness analysis across system complexity

In this subsection, we examine the robustness of our proposed methods by analyzing various $\mathbf{b}$ vectors and different initializations. We will compare the success rates and the number of iterations required for each approach. Through this methodology, we aim to rigorously assess the performance and reliability of our methods under varying conditions.

Generation of random initialization

To construct a set of random initial guess, we start with a vector $\mathbf{n}$ such that

$$n_{aq} = (\mathbf{b}_{\mathrm{O}}, 1, \ldots, 1), \quad n_{min} = 1, \text{ for all mineral}, \quad n_{gas} = (1, \ldots, 1),$$

and we generate a perturbed vector $\tilde{\mathbf{n}}_{\boldsymbol{\theta}}$ for all species except H_2O(aq) using a log normal distribution characterized by parameters (μ, σ):

$$\tilde{\mathbf{n}}_{\boldsymbol{\theta}} = \mathbf{n} \odot \boldsymbol{\theta} \quad \text{with} \quad \boldsymbol{\theta} \sim \mathcal{LN}(\mu, \sigma^2).$$

where $\odot$ denotes the Hadamard product. The lognormal distribution ensures that all elements of $\tilde{\mathbf{n}}_{\boldsymbol{\theta}}$ remain positive. From the vector $\mathbf{n}_{\boldsymbol{\theta}} = [\mathbf{b}_{\mathrm{O}}, \tilde{\mathbf{n}}_{\boldsymbol{\theta}}]$, we then define an initial guess $\boldsymbol{\mathcal{X}}_{\boldsymbol{\theta}}^{(0)}$ as detailed in the begin of Section 5. Consequently, we consider a collection of K random initial guess defined as:

$$\left\{ (\boldsymbol{\mathcal{X}}_{\boldsymbol{\theta}_1}^{(0)}, \ldots, \boldsymbol{\mathcal{X}}_{\boldsymbol{\theta}_K}^{(0)}) \mid \boldsymbol{\theta}_i \sim \mathcal{LN}(\mu, \sigma^2) \right\}$$

In our study, we will consider $K = 1000$ random initial guess characterized by $\mu = -0.5$ and $\sigma = 1$ in order to get a median of 1. The corresponding probability density function for $\boldsymbol{\theta}$ is depicted in Figure 11

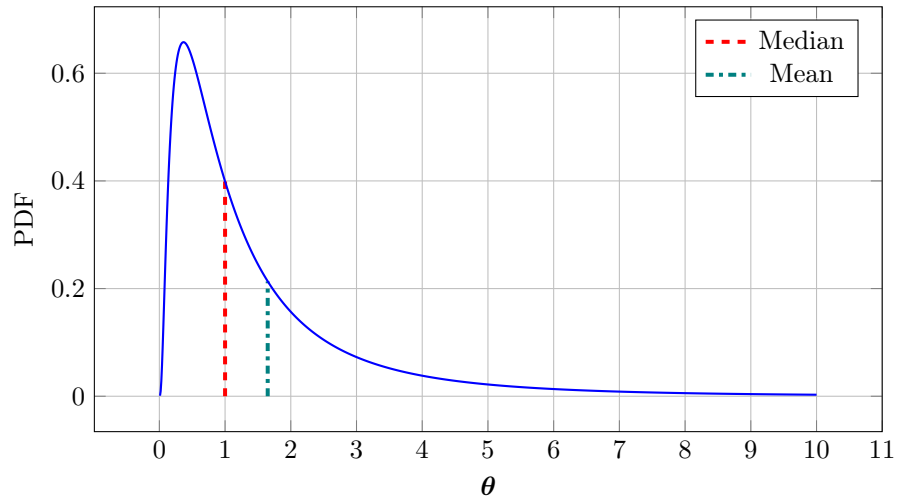

Figure 11: Probability density function of $\boldsymbol{\theta} \sim \mathcal{LN}(0, 1)$.

Generation of random right-hand side

To construct a set of random vectors for the right-hand side, we start with the solution $\mathbf{n}^\star$ to the minimization problem:

$$\mathbf{n}^\star = \arg\min\{G(\mathbf{n}) \mid \mathbf{An} = \mathbf{b}\}.$$

To maintain the concept of a diluted solution, we conserve the amount of water from the optimal solution while randomly perturbing the concentrations of the diluted species. We generate a perturbed vector $\tilde{\mathbf{n}}_{\boldsymbol{\theta}}$ for all diluted species using a lognormal distribution characterized by parameters (μ, σ):

$$\tilde{\mathbf{n}}_{\boldsymbol{\theta}} = \tilde{\mathbf{n}}^\star \odot \boldsymbol{\theta} \quad \text{with} \quad \boldsymbol{\theta} \sim \mathcal{LN}(\mu, \sigma^2).$$

where $\odot$ denotes the Hadamard product. The lognormal distribution ensures that all elements of $\tilde{\mathbf{n}}_{\boldsymbol{\theta}}$ remain positive. From the vector $\mathbf{n}_{\boldsymbol{\theta}} = [n_1^\star, \tilde{\mathbf{n}}_{\boldsymbol{\theta}}]$, we then define a new right-hand side vector as $\mathbf{b}_{\boldsymbol{\theta}} = \mathbf{A}\mathbf{n}_{\boldsymbol{\theta}}$.

This construction satisfies the non-empty assumption for the set $\mathcal{M}_{\mathbf{A},\mathbf{b}_{\boldsymbol{\theta}}}$. Consequently, we consider a collection of K random right-hand sides defined as:

$$\{(\mathbf{b}_{\boldsymbol{\theta}_1}, \ldots, \mathbf{b}_{\boldsymbol{\theta}_K}) \mid \boldsymbol{\theta}_i \sim \mathcal{LN}(\mu, \sigma^2)\}$$

5.2.1 CO_2: reference system for robustness evaluation

In this test case, we analyze the sensitivity of the methods to both initialization and variations in the right-hand side vector $\mathbf{b}$. Three configurations are considered:

- **A**: only the aqueous phase is present;
- **B**: the aqueous and mineral phases are present;
- **C**: all phases are present.

Sensitivity to initialization. The results, shown in Figure 12, include two key metrics: the top panel reports the success rate of each method over 1000 chemical equilibrium computations, while the bottom panel presents the average number of iterations for cases where Newton's algorithm converges. These results demonstrate the strong robustness of IPM. Notably, the combination of the log-trick with complementarity-based parametrization achieves a comparable success rate, although it requires approximately twice as many iterations to converge compared to the IPM approach.

The relatively poorer performance of the other approaches should be interpreted with care, since the IPM benefits from the line search strategy used in IPOPT, as discussed in Section C.4. It would be worthwhile to test this same line search strategy for the min and FB methods, as they employ the same set of variables as IPM. In contrast, the complementarity parametrization uses only one variable per phase, which introduces different challenges.

One such challenge is related to invertibility: when a significant portion of the Jacobian becomes null, convergence can be compromised. To address this, we introduce a strategy that "pauses" the update at zero when the sign of a component η_α changes. This approach helps the iterate remains in its current zone at the next iteration, leveraging the particular choice of the derivatives $S'(0)$ and $R'(0)$ described in Section 4.2. Although this strategy is applied selectively, we keep using the Newton direction for $\boldsymbol{\eta}$. The line search is then modified as follows:

$$\begin{aligned}
\boldsymbol{\eta}^{(k+1)} &= \boldsymbol{\eta}^{(k)} + \min_{j \in \{1, \ldots, N_{Ph}\}} \left\{ \beta_j^{(k)} \right\} \delta\boldsymbol{\eta}^{(k)}, \\
\beta_j^{(k)} &= \begin{cases} -\eta_j^{(k)} / \delta\eta_j^{(k)} & \text{if sign}\left(\eta_j^{(k)} + \delta\eta_j^{(k)}\right) \neq \text{sign}\left(\eta_j^{(k)}\right), \\ 1 & \text{otherwise.} \end{cases}
\end{aligned}$$

The performance of this approach is illustrated in Figure 13. We observe a significant improvement in success rates of previously unstable methods. Regarding convergence speed, the log-trick combined with complementarity parametrization is now competitive with the fastest methods. However, both the min and FB approaches can exhibit slow convergence, particularly in configuration **C**, where all phases are present—the min method being especially affected.

Overall, for this small chemical system, the combination of the log-trick with complementarity parametrization appears to offer the most favorable basin of attraction.

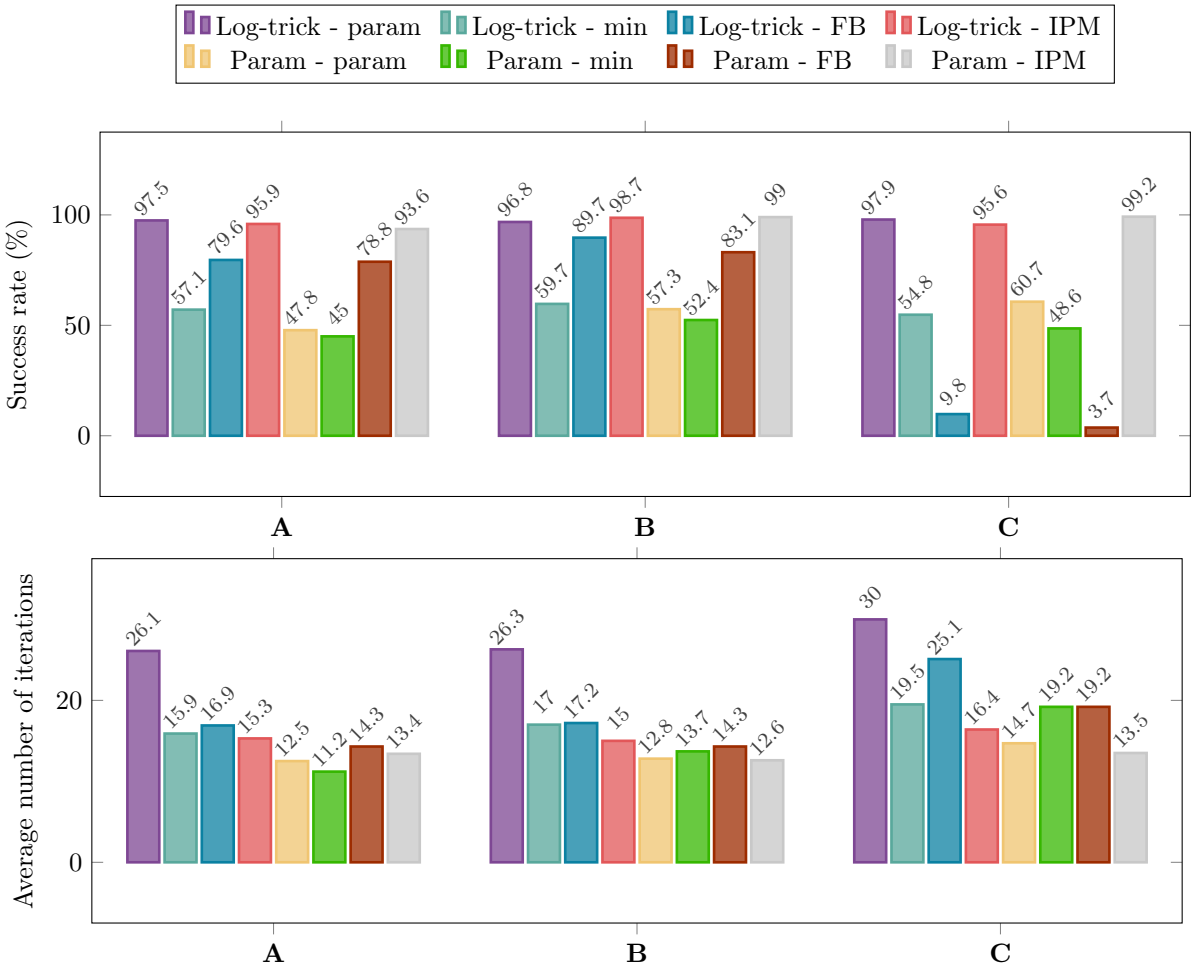

Figure 12: Performance metrics of CO_2 test case for a set of random initial guess.

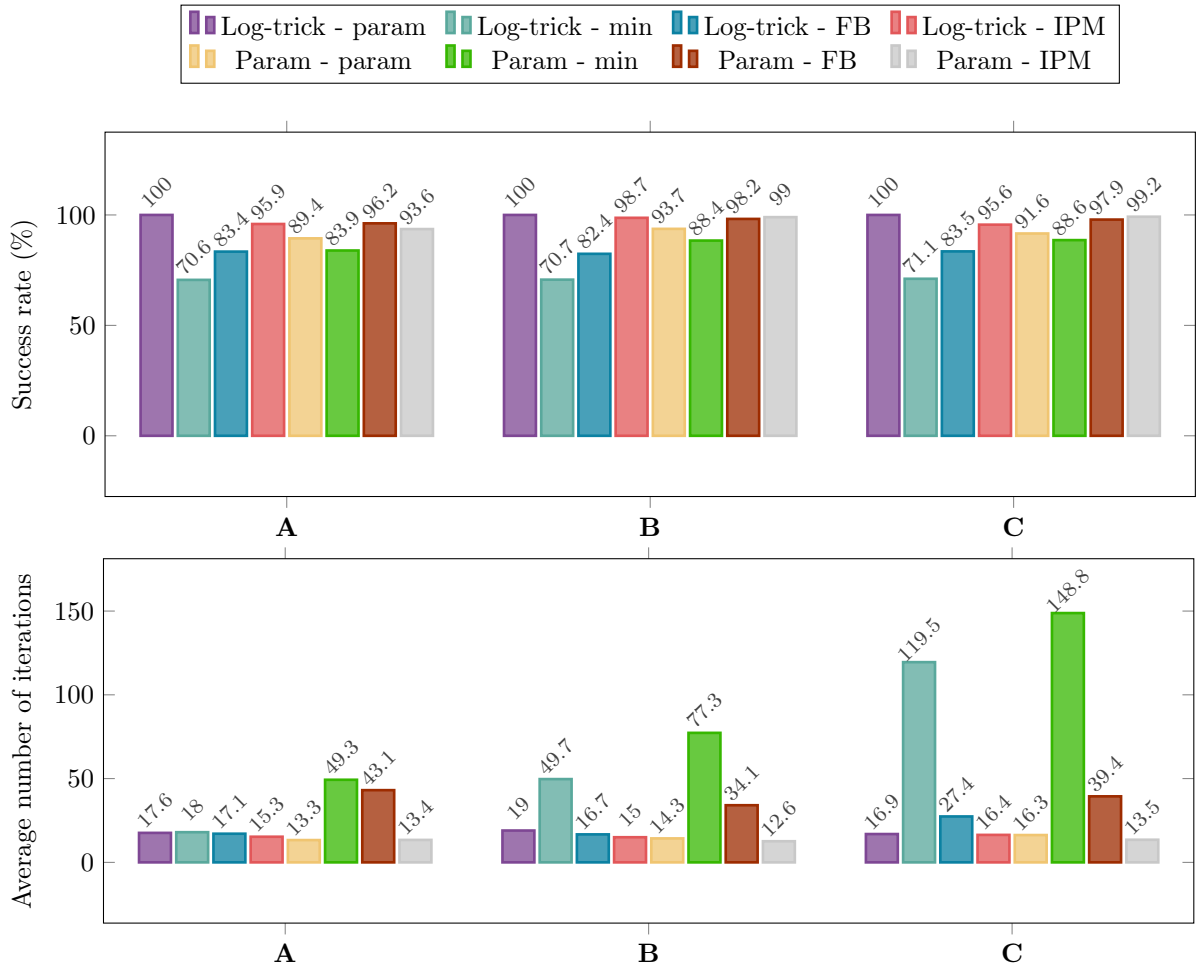

Figure 13: Performance metrics of CO_2 test case for a set of random initial guess with a line search strategy.

Sensitivity to right-hand side. The results obtained for all methods using line search strategies are presented in Figure 14. All methods exhibit strong robustness with respect to variations in the right-hand side: the success rate is consistently close to or equal to 100%. In terms of convergence speed, the min function shows some difficulties, particularly in configuration C.

From these two sensitivity analyses, it is evident that incorporating a line search strategy is essential for ensuring robustness. Furthermore, the proposed complementarity-based parametrization proves to be competitive with, and in some cases superior to, classical methods.

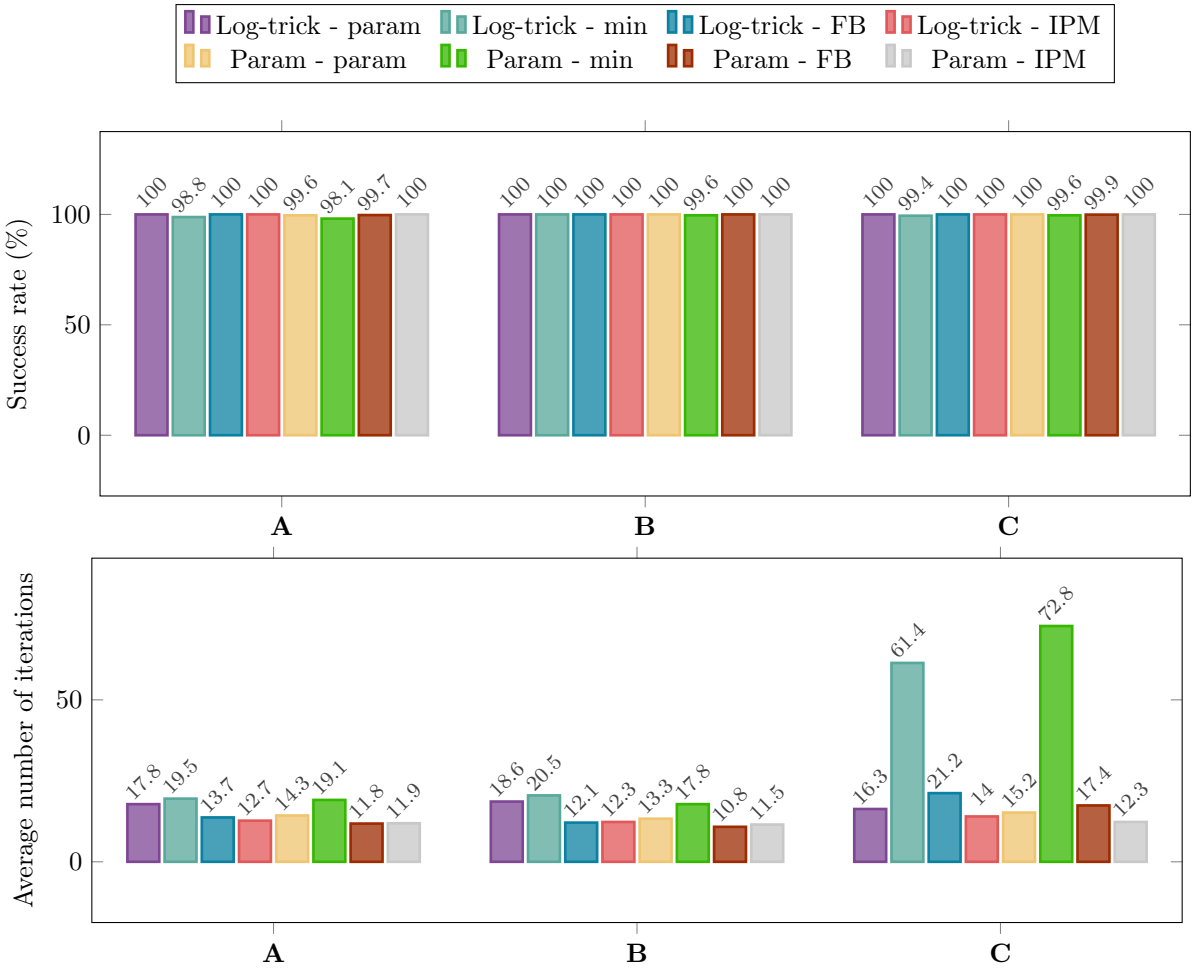

Figure 14: Performance metrics of CO_2 test case for a set of random right-hand side with a line search strategy.

Conclusion on CO_2 system and outlook. This simplified CO_2 test case provides a controlled environment to validate the behavior of the different approaches. It highlights the significant role of the line search mechanism in enhancing robustness.

The combination of the log-trick with complementarity parametrization exhibits a wide basin of attraction and competitive convergence rates. In contrast, the FB and min functions may suffer from slower convergence in multi-phase contexts, especially without tailored damping strategies.

Overall, these preliminary results justify the use of advanced line search techniques and dedicated regularization mechanisms when tackling more challenging chemical systems. We now turn our attention to larger and more realistic systems, where phase behavior, nonlinearity, and degeneracy pose additional difficulties for robust convergence.

5.2.2 *Multiphase Seawater*: large-scale system for robustness evaluation

This second test case is composed of 72 species divided into 22 phases:

- 50 aqueous species in 1 phase;

Setting present phases	aqueous	mineral	gas
A	1	0	0
B	1	2	0
C	1	4	0
D	1	5	0
E	1	4	1

Table 3: Number of present phases at equilibrium for each setting.

- 20 minerals in 20 pure phases;
- 2 gaseous species in 1 phase.

It is a challenging test case that allows us to evaluate the limits of each method. There are no methods that can converge without a line search strategy. As for the previous test case, we have performed a study of the sensitivity to the initial guess and right-hand side. Tests were conducted on five chemical configurations (A to E), representing systems of increasing complexity:

Sensitivity to initialization. Figure 15 presents the performance for solving the *Seawater* case. Each configuration was tested over 1000 random initial guesses.

The results show that the parameterized formulation of the log significantly outperforms the *log-trick* in terms of success rate, especially when multiple mineral phases are present. While the *log-trick* performs well on simpler settings (A, B) with the complementarity parametrization method (>90%), it drastically fails in more complex cases (C to E), with success rates dropping below 40%. The FB and IPM strategies are almost entirely ineffective in this formulation. In contrast, the parameterized formulation of the log maintains high and stable success rates across all configurations when combined with either the complementarity parametrization ("param") method (around 88%) or the FB method (approximately 80% across most settings, with a slight drop in E). The min method, however, is practically unusable: it fails to converge in the vast majority of simulations, resulting in a 0% success rate.

In terms of efficiency, measured by the average number of iterations, both "param" and FB methods in the parameterized formulation of the log are competitive. The "param" method is the fastest (30 to 50 iterations), while the FB method remains efficient with iteration counts between 40 and 80.

In summary, the combination of the parameterized formulation with the complementarity parametrization exhibits the highest overall performance.

Convergence analysis with random perturbations on the right-hand side vector. Figure 16 presents the performance when the right-hand side vector $\mathbf{b}$ is randomly perturbed across 1000 simulations. Due to the nature of the perturbations, it is difficult to strictly control the number of phases present at equilibrium. However, the majority of simulations still fall within the five main chemical equilibrium configurations (A to E).

Overall, the success rates remain very high for the parameterized formulation of the log combined with the "param" method, consistently exceeding 90% even for the most complex configurations (C, D, and E). This robustness contrasts with the *log-trick* formulation, which exhibits a marked drop in success rates for complex systems, falling below 50% for configurations C and D.

The FB approach in the parameterized formulation also achieves very good convergence rates (ranging from 55.6% to 89.1%), showing notable improvement compared to the *log-trick* + FB results. The IPM consistently underperforms across all configurations, with very low success rates. The min method again fails to converge in all cases, confirming its unreliability under random perturbations of b.

Regarding iteration counts, the parameterized formulation combined with the "param" method shows stable and moderate iteration numbers (around 33 to 47), which are comparable to those observed with the *log-trick* + param method. The FB method requires a higher number of iterations on average but remains within reasonable bounds (approximately 42 to 110 iterations).

In conclusion, the parameterized formulation with the "param" approach once again demonstrates superior robustness and efficiency when facing perturbations in the right-hand side vector, confirming its suitability for solving challenging chemical equilibrium problems involving complementarity constraints.

Conclusion on multiphase Seawater system. Among all the tested strategies, the combination of the parameterized formulation of the log with the parameterized complementarity approach exhibits the

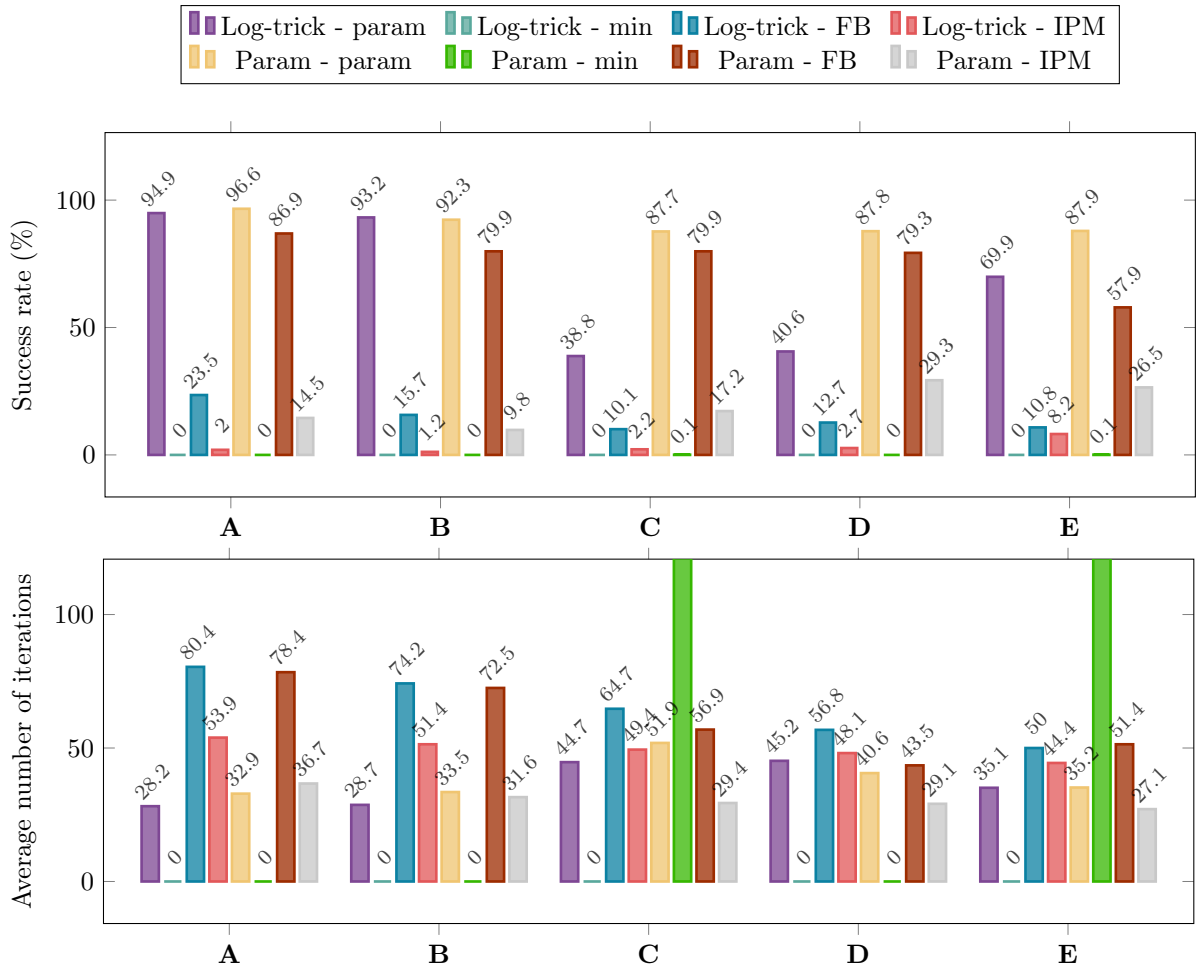

Figure 15: Performance metrics of *Seawater* test case for a set of random initial guess with a line search strategy.

highest overall performance. It achieves the best convergence rates across all chemical configurations (above 87% in every case), while also requiring fewer Newton iterations on average compared to the other methods. This clearly indicates that the proposed parametrized approach is both more robust and more efficient than traditional methods used in the literature. These results validate the relevance of the parameterized complementarity formulation as a promising alternative for solving equilibrium problems involving complementarity constraints.

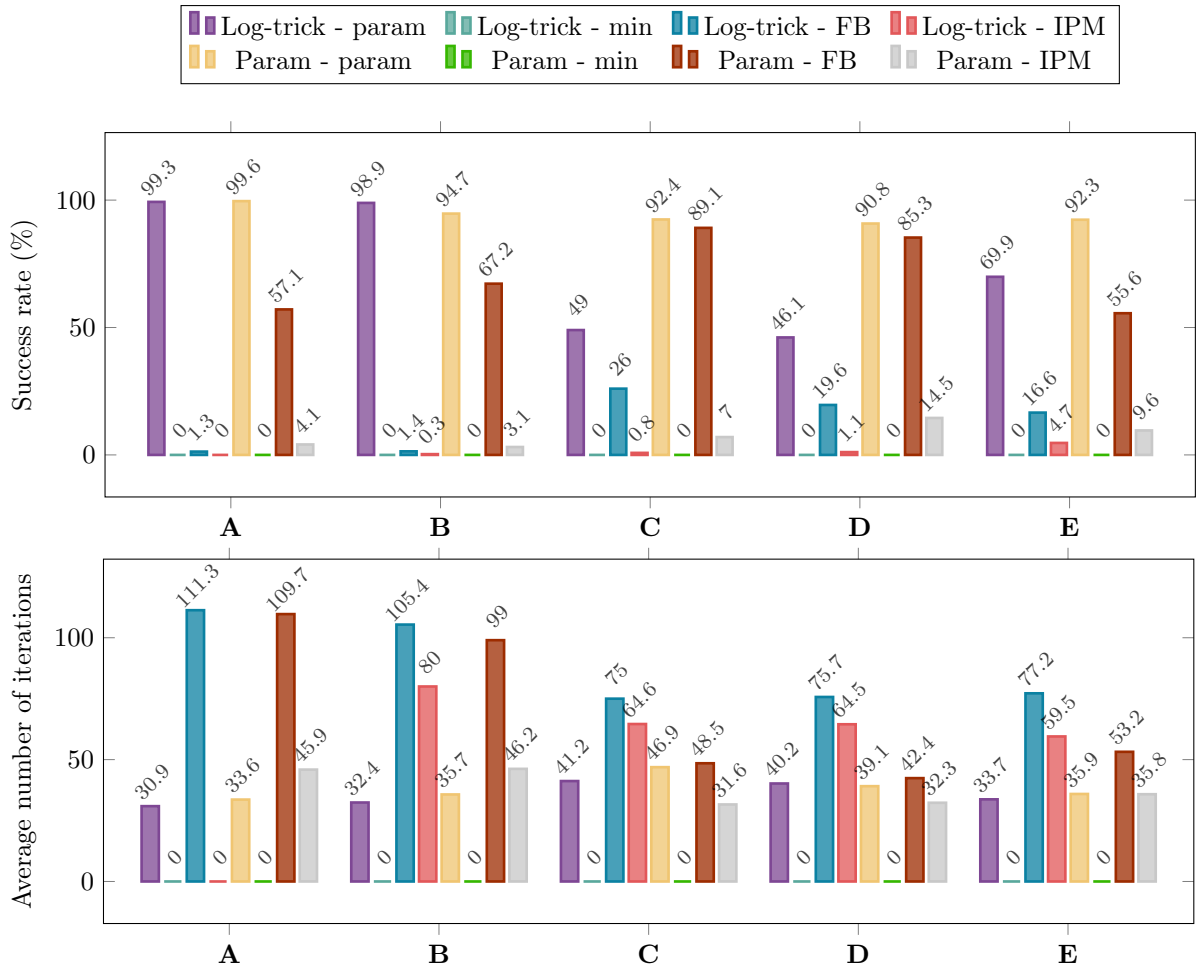

Figure 16: Performance metrics of *Seawater* test case for a set of random right-hand side with a line search strategy.

6 Conclusion

In this paper, we addressed the complexities of multiphase equilibrium by introducing new conditions that guarantee uniqueness even when certain phases may vanish. Additionally, we established an equivalence between the minimization problem and a new set of algebraic equations based on extended mole fractions, leveraging concepts from the subdifferential of Gibbs energy.

This work presented a comprehensive numerical evaluation of Newton-based methods for solving multiphase chemical equilibrium problems under complementarity constraints. We proposed two main innovations: (i) a parameterized reformulation of the link between quantities and chemical potentials, and (ii) a novel parametrization of the complementarity condition.

Through two distinct robustness tests — one based on random initializations, and the other on random perturbations of the right-hand side vector — in a challenging chemical system composed of 72 species and 22 phases, we observed a clear performance advantage for the proposed parameterized approaches. This approach consistently achieved the highest success rates (often exceeding 90%) across all equilibrium configurations, while requiring moderate iteration counts. This approach outperformed all classical techniques, including the commonly used Newton-min, Fischer-Burmeister and interior point methods, which demonstrated either poor robustness or high computational cost.

The classical *log-trick* formulation exhibited strong sensitivity to initial conditions and problem configurations, and appears to be significantly less robust in complex multiphase systems. Similarly, the interior point method failed to deliver reliable convergence across all tests, regardless of the formulation.

These results validate the proposed parametrization strategy as a powerful and reliable alternative to existing methods for solving constrained nonlinear systems arising in chemical thermodynamics. By improving both the convergence rate and robustness, especially in challenging configurations, our methods represent a promising direction for future developments in numerical geochemistry.

CRediT authorship contribution statement

Maxime Jonval: Methodology, Software, Validation, Formal analysis, Investigation, Writing - Original Draft, Writing - Review & Editing, Visualization. **Ibtihel Ben Gharbia:** Conceptualization, Methodology, Validation, Writing - Review & Editing, Supervision. **Clément Cancès:** Conceptualization, Methodology, Formal analysis, Writing - Review & Editing, Supervision. **Thibault Faney**: Conceptualization, Methodology, Validation, Writing - Review & Editing, Supervision. **Quang-Huy Tran:** Conceptualization, Methodology, Writing - Review & Editing, Supervision.

Acknowledgement

This work was jointly supported by IFPEN and Inria. Maxime Jonval and Clément Cancès acknowledge support from the Labex CEMPI (ANR-11-LABX-0007-01), whereas Ibtihel Ben Gharbia, Clément Cancès and Quang Huy Tran acknowledge support from the project MATHSOUT of the PEPR Mathematics in Interaction (ANR-23-EXMA-0010) funded by the French National Research Agency.

A Data for chemical systems

A.1 The SiO_2 test case

The SiO_2 test case is composed of

$\mathcal{C} = (H_2O,\ H^+,\ SiO_2(aq),\ OH^-),\ \mathcal{E} = (H,\ O,\ Si),\ \mathcal{R} = (OH^- = H_2O - H^+,\ SiO_2(Quartz) = SiO_2(aq))$.

A.2 The CO_2 test case

The CO_2 system is composed of

$$
\begin{aligned}
&\mathcal{C} = (H_2O,\ Ca^{2+},\ CO_2(aq),\ H^+,\ OH^-,\ HCO_3^-,\ CO_3^{2-},\ CaCO_3(calcite),\ CO_2(g),\ H_2O(g)),\\
&\mathcal{E} = (O,\ Ca,\ C,\ H),\\
&\mathcal{R} = (\quad OH^- = H_2O - H^+; \qquad\qquad HCO_3^- = H_2O + CO2(aq) - H^+\\
&\qquad\quad CO_3^{2-} = H_2O + CO2(aq) - 2H^+; \quad CaCO_3(calcite) = H_2O + Ca^{2+} + CO2(aq) - 2H^+\\
&\qquad\quad CO_2(g) = CO2(aq); \qquad\qquad H_2O(g) = H_2O)
\end{aligned}
$$

A.3 The *Seawater* test case

The *Seawater* test case is composed of

$$
\begin{aligned}
\mathcal{C} = (&H_2O(l),\ H^+,\ Cl^-,\ Na^+,\ Mg^{2+},\ SO_4^{2-},\ Ca^{2+},\ K^+,\ HCO_3^-,\ Br^-,\\
&B(OH)_3(aq),\ Sr^{2+},\ F^-,\ BF_4^-,\ BO_2^-,\ CaOH^+,\ CO_2(aq),\ CO_3^{2-},\ CaHCO_3^+,\\
&CaCl^+,\ CaCl_2(aq),\ CaF^+,\ CaSO_4(aq),\ HF(aq),\ HF_2^-,\ HSO_4^-,\ KBR(aq),\ KCL(aq),\\
&KHSO_4(aq),\ KOH(aq),\ KSO_4^-,\ Mg(CO_3)(aq),\ Mg(HCO3)^+,\ MgCl^+,\ MgF+,\ MgOH+,\ NaBR(aq),\\
&NaCL(aq),\ NaF(aq),\ NaOH(aq),\ OH^-,\ Sr(HCO_3)^+,\ SrCl^+,\ SrF^+,\ SrOH^+,\ NaSO_4^-,\\
&MgSO_4(aq),\ HCl(aq),\ CaCO_3(aq),\ SrCO_3(aq),\ CaSO_4(anhydrite),\ CaCO3(aragonite),\\
&Mg_2(OH)_2(CO_3)\cdot 3H_2O(artinite),\ Mg(OH)_2(brucite),\ CaCO_3(calcite),\ SrSO_4(celestite),\\
&CaMg(CO_3)_2(dolomite-dis),\ CaMg(CO_3)_2(dolomite-ord),\ CaF_2(fluorite),\ NaCl(halite),\\
&CaMg_3(CO_3)_4(huntite),\ Mg_5(OH)_2(CO_3)_4\cdot 4H_2O(hydromagnesite),\ CaO(lime),\\
&MgCO_3(magnesite),\ MgO(periclase),\ K_2O(potassium-oxide),\ Na_2O(sodium-oxide),\\
&SrCO_3(strontianite),\ KCl(sylvite),\ MgCO_3\cdot 3H_2O(nesquehonite),\ CO_2(g),\ H_2O(g))\\
\mathcal{E} = (&O,\ H,\ Cl,\ Na,\ Mg,\ S,\ Ca,\ K,\ C,\ Br,\ B,\ Sr,\ F),
\end{aligned}
$$

and the set $\mathcal{R}$ composed of the reactions:

- aqueous species:

$$
\begin{aligned}
BF_4^- &= -3H_2O + 3H^+ + B(OH)_3(aq) + 4F^-; \qquad BO_2^- = -H_2O - H^+ + B(OH)_3(aq)\\
CaOH^+ &= H_2O - H^+ + Ca^{2+}; \qquad\qquad CO_2(aq) = -H_2O + H^+ + HCO_3^-
\end{aligned}
$$

$$CO_3^{2-} = -H^+ + HCO_3^-;$$
$$CaCl^+ = Cl^- + Ca^{2+};$$
$$CaF^+ = Ca^{2+} + F^-;$$
$$HF(aq) = H^+ + F^-;$$
$$HSO_4^- = H^+ + SO_4^{2-};$$
$$KCL(aq) = Cl^- + K^+;$$
$$KOH(aq) = H_2O - H^+ + K^+;$$
$$Mg(CO_3)(aq) = -H^+ + Mg^{2+} + HCO_3^-;$$
$$MgCl^+ = Cl^- + Mg^{2+};$$
$$MgOH+ = H_2O - H^+ + Mg^{2+};$$
$$NaCL(aq) = Cl^- + Na^+;$$
$$NaOH(aq) = H_2O - H^+ + Na^+;$$
$$Sr(HCO_3)^+ = HCO_3^- + Sr^{2+};$$
$$SrF^+ = Sr^{2+} + F^-;$$
$$NaSO_4^- = Na^+ + SO_4^{2-};$$
$$HCl(aq) = H^+ + Cl^-;$$
$$SrCO_3(aq) = -H^+ + HCO_3^- + Sr^{2+}$$
$$CaHCO_3^+ = Ca^{2+} + HCO_3^-$$
$$CaCl_2(aq) = 2Cl^- + Ca^{2+}$$
$$CaSO_4(aq) = SO_4^{2-} + Ca^{2+}$$
$$HF_2^- = H^+ 2.0F^-$$
$$KBR(aq) = K^+ + Br^-$$
$$KHSO_4(aq) = H^+ + SO_4^{2-} + K^+$$
$$KSO_4^- = SO_4^{2-} + K^+$$
$$Mg(HCO3)^+ = Mg^{2+} + HCO_3^-$$
$$MgF+ = Mg^{2+} + F^-$$
$$NaBR(aq) = Na^+ + Br^-$$
$$NaF(aq) = Na^+ + F^-$$
$$OH^- = H_2O - H^+$$
$$SrCl^+ = Cl^- + Sr^{2+}$$
$$SrOH^+ = H_2O - H^+ + Sr^{2+}$$
$$MgSO_4(aq) = Mg^{2+} + SO_4^{2-}$$
$$CaCO_3(aq) = -H^+ + Ca^{2+} + HCO_3^-$$

- pure mineral species:

$$CaSO_4(anhydrite) = SO_4^{2-} + Ca^{2+}$$
$$CaCO3(aragonite) = -H^+ + Ca^{2+} + HCO_3^-$$
$$Mg_2(OH)_2(CO_3) \cdot 3H_2O(artinite) = 5H_2O - 3H^+ + 2Mg^{2+} + HCO_3^-$$
$$Mg(OH)_2(brucite) = 2H_2O - 2H^+ + Mg^{2+}$$
$$CaCO_3(calcite) = -H^+ + Ca^{2+} + HCO_3^-$$
$$SrSO_4(celestite) = SO_4^{2-} + Sr^{2+}$$
$$CaMg(CO_3)_2(dolomite - dis) = -2H^+ + Mg^{2+} + Ca^{2+} + 2HCO_3^-$$
$$CaMg(CO_3)_2(dolomite - ord) = -2H^+ + Mg^{2+} + Ca^{2+} + 2HCO_3^-$$
$$CaF_2(fluorite) = Ca^{2+} + 2F^-$$
$$NaCl(halite) = Cl^- + Na^+$$
$$CaMg_3(CO_3)_4(huntite) = -4H^+ + 3Mg^{2+} + Ca^{2+} + 4HCO_3^-$$
$$Mg_5(OH)_2(CO_3)_4 \cdot 4H_2O(hydromagnesite) = 6H_2O - 6H^+ + 5Mg^{2+} + 4HCO_3^-$$
$$CaO(lime) = H_2O - 2H^+ + Ca^{2+}$$
$$MgCO_3(magnesite) = -H^+ + Mg^{2+} + HCO_3^-$$
$$MgO(periclase) = H_2O - 2H^+ + Mg^{2+}$$
$$K_2O(potassium - oxide) = H_2O - 2H^+ + 2K^+$$
$$Na_2O(sodium - oxide) = H_2O - 2H^+ + 2Na^+$$
$$SrCO_3(strontianite) = -H^+ + HCO_3^- + Sr^{2+}$$
$$KCl(sylvite) = Cl^- + K^+$$
$$MgCO_3 \cdot 3H_2O(nesquehonite) = 3H_2O - H^+ + Mg^{2+} + HCO_3^-$$

- gaseous species:

$$CO_2(g) = -H_2O + H^+ + HCO_3^-; \qquad H_2O(g) = H_2O(l)$$

B Standard chemical potentials

The standard chemical potential μ_i°(P, T) of a species C_i for a constant pressure P and temperature T is calculated from the SUPCRT92 database [17].

Table 4: Standard chemical potentials at P = 1 Bar and T = 298.15 K.

Formula	μ_i°(P,T)	Formula	μ_i°(P,T)
H_2O	-237138.9758928523	NaBR(aq)	-358192.19663367077
H^+	0.0	NaCL(aq)	-388735.3787766348
Cl^-	-131289.73255269116	NaF(aq)	-537936.7588824253
Na^+	-261880.68093357567	NaOH(aq)	-417981.5046476963
Mg^{2+}	-453984.7875582697	OH^-	-157297.4420362631
SO_4^{2-}	-744458.9698933411	$Sr(HCO_3)^+$	-1.157796221963139e6
Ca^{2+}	-552789.9276571299	$SrCl^+$	-693707.0438437797
K^+	-282461.785083038	SrF^+	-846381.1427032378
HCO_3^-	-586939.7841306848	$SrOH^+$	-725087.037231101
Br^-	-104056.09221446048	$NaSO_4^-$	-1.0103353876813356e6
$B(OH)_3$(aq)	-968763.1798276871	$MgSO_4$(aq)	-1.2111714801953817e6
Sr^{2+}	-563835.6855778464	HCl(aq)	-127235.40484113985
F^-	-281750.4962683066	$CaCO_3$(aq)	-1.099764116286581e6
BF_4^-	-1.4869933045074963e6	$SrCO_3$(aq)	-1.1081739548331397e6
BO_2^-	-678811.9589547777	$CaSO_4$(Anhydrite)	-1.3218299449455e6
$CaOH^+$	-716719.0378653706	CaCO3(Aragonite)	-1.12835344756575e6
CO_2(aq)	-385973.95119834575	$Mg_2(OH)_2(CO_3) \cdot 3H_2O$(Artinite)	-2.5686198744775e6
CO_3^{2-}	-527983.0213241152	$Mg(OH)_2$(Brucite)	-835318.69698875
$CaHCO_3^+$	-1.145704464880173e6	$CaCO_3$(Calcite)	-1.1291776990575502e6
$CaCl^+$	-682410.2459630222	$SrSO_4$(Celestite)	-1.34069978084625e6
$CaCl_2$(aq)	-811695.8491350175	$CaMg(CO_3)_2$(Dolomite − dis)	-2.1574917216637502e6
CaF^+	-838431.5438473737	$CaMg(CO_3)_2$(Dolomite − ord)	-2.1663074074905002e6
$CaSO_4$(aq)	-1.3092988132949607e6	CaF_2(Fluorite)	-1.17358246402515e6
HF(aq)	-299833.75289883884	NaCl(Halite)	-384120.431987875
HF_2^-	-578061.3350002527	$CaMg_3(CO_3)_4$(Huntite)	-4.2037057981325e6
HSO_4^-	-755755.7895325241	$Mg_5(OH)_2(CO_3)_4 \cdot 4H_2O$(Hydromagnesite)	-5.864657282973e6
KBR(aq)	-376601.8109260084	CaO(Lime)	-604027.2218463
KCL(aq)	-399279.0752045901	$MgCO_3$(Magnesite)	-1.02783286346585e6
$KHSO_4$(aq)	-1.0183854278434662e6	MgO(Periclase)	-569383.7028176
KOH(aq)	-437227.9151528981	K_2O(Potassium − oxide)	-322402.2804475
KSO_4^-	-1.0319415534918394e6	Na_2O(Sodium − oxide)	-376070.415242
$Mg(CO_3)$(aq)	-998971.5788823194	$SrCO_3$(Strontianite)	-1.15256625621825e6
$Mg(HCO3)^+$	-1.0468365642174695e6	KCl(Sylvite)	-408923.19198430004
$MgCl^+$	-584504.6645023846	$MgCO_3 \cdot 3H_2O$(Nesquehonite)	-1.7239541270617498e6
MgF+	-743454.7620717564	CO_2(g)	-394391.27240993205
MgOH+	-624482.7757953086	H_2O(g)	-228164.33933913204

C State of the art methods for the resolution

C.1 The log-trick

As for the single-phase case, it could be interesting to test the robustness of the *log-trick* approach as a reference for comparing the methods. The idea of this approach is to make the following change of variable:

$$y = \ln \xi.$$

This ensures that the mole fractions are always positive and prevents issues with the Jacobian when ξ is small. Applying this change of variable to (12a)–(12c) yields:

$$\sum_{\alpha=1}^{N_{Ph}} s_\alpha \mathbf{A}^\alpha \mathbf{exp}(\mathbf{y}^\alpha) - b = 0, \tag{54a}$$

$$\mathbf{S}^T[\mu_i^\circ/(\mathrm{RT}) + y_i]_{i=1,\ldots,N} = \mathbf{0}, \tag{54b}$$

$$\langle \mathbf{exp}(\mathbf{y}^\alpha), \mathbf{1} \rangle + r_\alpha - 1 = 0, \qquad (\alpha = 1, \ldots, N_{Ph}), \tag{54c}$$

where $\mathbf{exp}(\mathbf{y}) = (\exp(y_i))_{i=1,\dots,N}$. The associated block in the Jacobian is written as:

$$\begin{bmatrix} s_1\mathbf{A}^1\mathrm{diag}\{\mathbf{exp}(\mathbf{y}^1)\} & \cdots & s_{N_{Ph}}\mathbf{A}^{N_{Ph}}\mathrm{diag}\{\mathbf{exp}(\mathbf{y}^{N_{Ph}})\} & \mathbf{A}\mathbf{exp}(\mathbf{y}) & \mathbf{0} \\ & & \mathbf{S}^T & \mathbf{0} & \mathbf{0} \\ \mathbf{exp}(\mathbf{y}^1)^T & & & & \\ & \ddots & & \mathbf{0} & \mathbf{I}_{N_{Ph}} \\ & & \mathbf{exp}(\mathbf{y}^{N_{Ph}})^T & & \end{bmatrix}.$$

C.2 Classical approaches to complementarity problem

Classical approaches to tackle the complementarity problem are typically categorized into two main classes of methods.

The first class consists of semismooth methods, where the complementarity problem is approached using a complementarity function. This function is Lipschitz continuous but may not be differentiable at specific points, particularly when $s_\alpha = r_\alpha$. However, by using the concept of subdifferentials, we can define a subgradient that allows the application of Newton's algorithm effectively. This approach facilitates the handling of non-differentiability while still leveraging the advantages of Newton's method. In Section C.3, we will present the minimum and Fischer-Burmeister complementarity functions, which are commonly used in semismooth approaches to effectively handle the complementarity conditions.

The second class focuses on smoothing techniques for the complementarity equations. In this approach, a regularization parameter is introduced to create a smooth approximation of the complementarity problem. This enables the use of classical Newton's method on the modified problem. As the iterations progress, the regularization parameter is gradually reduced to zero, guiding the solution back to the original complementarity problem as the iterates converge towards the solution. In Section C.4, we will explore the interior point method (IPM), a well-established technique for addressing the smoothing of complementarity equations.

C.3 Semismooth methods

The treatment of the complementarity problem using semismooth methods involves the use of a complementarity function $\Psi : \mathbb{R}^2 \to \mathbb{R}$ that satisfies specific conditions. Different functions lead to distinct smoothness and regularity characteristics, which are crucial for the effectiveness of the numerical techniques applied to solve nonlinear complementarity problems. To be a complementarity function, Ψ must satisfy the following condition:

$$\Psi(s_\alpha, r_\alpha) = 0 \quad \Leftrightarrow \quad s_\alpha r_\alpha = 0 \text{ and } s_\alpha \geq 0,\, r_\alpha \geq 0.$$

There are several options available for the function Ψ, but our study will concentrate on the following two functions:

- **The min function** [23]:
$$\Psi_{\min}(s_\alpha, r_\alpha) = \min(s_\alpha, r_\alpha),$$
which is Lipschitz continuous but lacks differentibility at the points where $s_\alpha = r_\alpha$. This function is depicted in Figure 17.

- **The Fischer-Burmeister function** [14]:
$$\Psi_{\mathrm{FB}}(s_\alpha, r_\alpha) = s_\alpha + r_\alpha - \sqrt{s_\alpha^2 + r_\alpha^2},$$
which is differentiable everywhere except at the point where $s_\alpha = r_\alpha = 0$. This function is illustrated in Figure 18.

The non-differentiality of the min and Fischer-Burmeister (FB) functions leads to the use of semismooth Newton's method to solve the complementarity problem. In this approach, the Jacobian is replaced by an element of the Clarke subdifferential of the function for which we seek the root.

Let $\mathcal{F} : \boldsymbol{\mathcal{X}} \in \mathcal{D}_\mathcal{F} \to \mathbb{R}$ be a locally Lipschitz continuous function defined in a set $\mathcal{D}_\mathcal{F}$. By Rademacher's theorem [10, Section 3.4.1], such a function is continuously differentiable almost everywhere. Let Ω_F be the set where $\mathcal{F}$ is Fréchet differentiable, *i.e.*, where $\boldsymbol{\nabla}\mathcal{F}$ exists. The Bouligand subdifferential $\partial_\mathrm{B}\mathcal{F}(\boldsymbol{\mathcal{X}})$ of $\mathcal{F}$ at $\boldsymbol{\mathcal{X}} \in \Omega_\mathrm{F}$ is defined as:

$$\partial_\mathrm{B}\mathcal{F}(\boldsymbol{\mathcal{X}}) := \left\{ \mathcal{J} \;\middle|\; \exists (\boldsymbol{\mathcal{X}}^{(k)})_{k\in\mathbb{N}} \subset \Omega_\mathrm{F},\ \boldsymbol{\mathcal{X}}^{(k)} \to \boldsymbol{\mathcal{X}},\ \boldsymbol{\nabla}\mathcal{F}(\boldsymbol{\mathcal{X}}^{(k)}) \to \mathcal{J} \right\}.$$

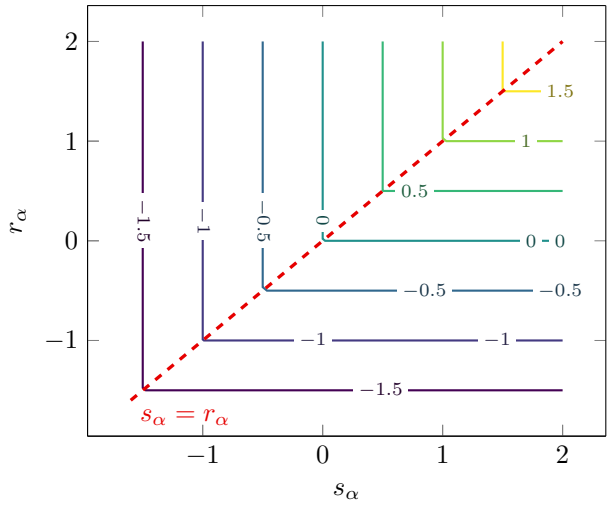

Figure 17: Contour plot of the function $\Psi_{\min}$.

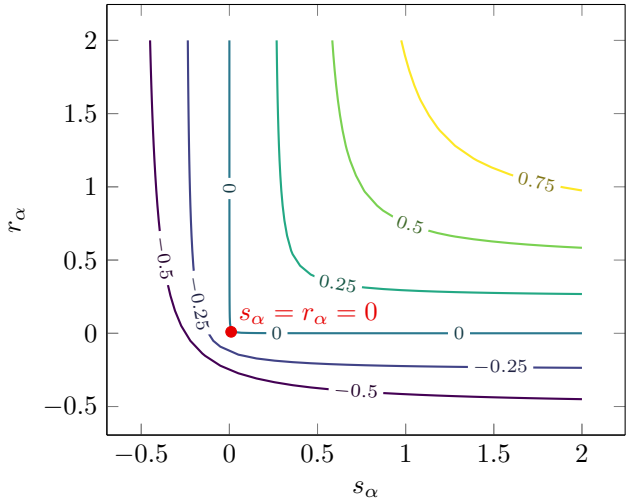

Figure 18: Contour plot of the function Ψ_{FB}.

The Clarke subdifferential [7] is then given by:

$$\partial_{\text{C}}\mathcal{F}(\boldsymbol{\mathcal{X}}) = \text{conv}\left(\partial_{\text{B}}\mathcal{F}(\boldsymbol{\mathcal{X}})\right),$$

where conv denotes the convex hull.

As a simple example, consider the absolute value function $f(x) = |x|$ at $x = 0$. In this case, $\partial_{\text{B}} f(0) = \{-1,\, 1\}$ and $\partial_{\text{C}} f(0) = [-1,\, 1]$.

The semismooth Newton's algorithm then start from an initial value $\boldsymbol{\mathcal{X}}^{(0)}$, and builds a sequence $(\boldsymbol{\mathcal{X}}^{(k)})_{k>0}$ by solving the linear system [24]:

$$\mathcal{J}\boldsymbol{\delta\mathcal{X}}^{(k)} = -\mathcal{F}\left(\boldsymbol{\mathcal{X}}^{(k)}\right), \quad \mathcal{J} \in \partial_{\text{C}}\mathcal{F}\left(\boldsymbol{\mathcal{X}}^{(k)}\right),$$

and updating the sequence as:

$$\boldsymbol{\mathcal{X}}^{(k+1)} = \boldsymbol{\mathcal{X}}^{(k)} + \boldsymbol{\delta\mathcal{X}}^{(k)}.$$

The selection of $\mathcal{J} \in \partial_{\text{C}}\mathcal{F}\left(\boldsymbol{\mathcal{X}}^{(k)}\right)$ when the latter is multivalued can have an impact on the performance of the method. Some authors [16] suggest choosing an element $\mathcal{J}$ from the Bouligand subdifferential $\partial_{\text{B}}\mathcal{F}\left(\boldsymbol{\mathcal{X}}^{(k)}\right)$ instead. In our case, the corresponding Bouligand subdifferentials for the min and FB functions are:

$$\partial_{\text{B}}\Psi_{\min}(s_\alpha,\, r_\alpha)|_{s_\alpha = r_\alpha} = \{(1,\ 0)^T,\, (0,\ 1)^T\}$$
$$\partial_{\text{B}}\Psi_{\text{FB}}(0,\, 0) = \{(1,\, 1)^T - \mathbf{u} \mid \|\mathbf{u}\|_2 = 1\}.$$

In our algorithms, the chosen subdifferentials are:

$$\partial_{\mathrm{B}}\Psi_{\min}(s_\alpha,\, r_\alpha)|_{s_\alpha=r_\alpha} = (1,\, 0)^T$$
$$\partial_{\mathrm{B}}\Psi_{\mathrm{FB}}(0,\, 0) = (1-\sqrt{2}/2,\, 1-\sqrt{2}/2)^T.$$

However, these specific cases almost never occur in our numerical experiments.

C.4 Smoothing methods

The central concept of smoothing techniques involves the introduction of a new parameter, denoted by $\nu \in \mathbb{R}$, known as the regularization parameter, into the complementarity equation. This parameter enables the "smoothing" of the equation, rendering it differentiable and thus suitable for classical optimization methods, such as the Newton method. In this context, we will focus on a specific class of smoothing methods that are part of interior point techniques [15]. Here, the complementarity equation (12d) is regularized as follows:

$$s_\alpha r_\alpha = \nu.$$

The system is then transformed into a new, smooth equation given by:

$$\mathcal{F}(\boldsymbol{\mathcal{X}};\, \nu) = \mathbf{0}. \tag{55}$$

The smooth equation can be effectively solved using the Newton method. During the Newton iterations, the regularization parameter is gradually decreased, allowing the smooth equation to converge toward the original complementarity equation. As the iterations progress and ν approaches zero, the solution of the smooth equation converges to that of the original complementarity equation. The Newton step is defined by solving:

$$\boldsymbol{\nabla}\mathcal{F}(\boldsymbol{\mathcal{X}}^{(k)};\, \nu^{(k)})\boldsymbol{\delta}\boldsymbol{\mathcal{X}}^{(k)} = -\mathcal{F}(\boldsymbol{\mathcal{X}}^{(k)};\, \nu^{(k)}).$$

Subsequently, ν is updated according to a predefined sequence:

$$\nu^{(k+1)} = \Theta(\nu^{(k)}),$$

which converges toward zero. In our study, we consider the following strategy for reducing $\nu^{(k)}$ [32]:

$$\nu^{(k+1)} = \max\left(10^{-32},\, \min\left(\kappa_\nu \nu^{(k)},\, \left(\nu^{(k)}\right)^{\theta_\nu}\right)\right),$$

with $\kappa_\nu \in (0,\, 1)$ and $\theta_\nu \in (1,\, 2)$. At this stage, the complementarity equation is not fully satisfied, as we must ensure $s_\alpha,\, r_\alpha \geq 0$. While several strategies exist to maintain the positivity of these iterates, we present a method that treats s_α and r_α individually. The updates for $\mathbf{s} = (s_\alpha)_{\alpha=1,\,\ldots,\,N_{Ph}}$ and $\mathbf{r} = (r_\alpha)_{\alpha=1,\,\ldots,\,N_{Ph}}$ are as follows:

$$\mathbf{s}^{(k+1)} = \mathbf{s}^{(k)} + \beta_{\mathbf{s}}^{(k)}\boldsymbol{\delta}\mathbf{s}^{(k)},$$
$$\mathbf{r}^{(k+1)} = \mathbf{r}^{(k)} + \beta_{\mathbf{r}}^{(k)}\boldsymbol{\delta}\mathbf{r}^{(k)}.$$

Following the approach outlined in [32], $\beta_{\mathbf{u}}$ for $\mathbf{u} \in \{\mathbf{s},\, \mathbf{r}\}$ is defined as:

$$\beta_{\mathbf{v}} = \max\{\beta \in (0,\, 1] \mid \mathbf{u} + \beta\boldsymbol{\delta}\mathbf{u} \geq (1-\gamma^{(k)})\mathbf{u}\},$$

where $\gamma^{(k)}$ is referred to as the fraction-to-the-boundary parameter in [32]:

$$\gamma^{(k)} = \max(\gamma_{\min},\, 1-\nu^{(k)}),$$

with $\gamma_{\min} \in (0, 1)$. We will refer to this update as the IPOPT line search.

Let us also mention an alternative method where ν is considered as an unknown. This approach is named nonparametric IPM (NPIPM) and has been developed by Vu *et al.* [29, 30].

References

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

[2] I. Ben Gharbia. *Résolution de problèmes de complémentarité. : Application à un écoulement diphasique dans un milieu poreux*. PhD thesis, Université Paris Dauphine - Paris IX, 2012.

[3] I. Ben Gharbia and E. Flauraud. Study of compositional multiphase flow formulation using complementarity conditions. *Oil Gas Sci. Technol. – Rev. IFP Energies nouvelles*, 74(43), 2019.

[4] 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. *ESAIM: M2AN*, 55(6):38, 2021.

[5] I. Ben Gharbia and J. Jaffré. Gas phase appearance and disappearance as a problem with complementarity constraints. *Mathematics and Computers in Simulation*, 99:28–36, 2014.

[6] K. Brenner and C. Cancès. Improving Newton's method performance by parametrization: The case of the Richards equation. *SIAM Journal on Numerical Analysis*, 55(4):1760–1785, 2017.

[7] F. H. Clarke. Generalized gradients and applications. *Transactions of the American Mathematical Society*, 205:247–262, 1975.

[8] J. Coatléven and A. Michel. A successive substitution approach with embedded phase stability for simultaneous chemical and phase equilibrium calculations. *Computers & Chemical Engineering*, 168:108041, 2022.

[9] K. H. Coats. An equation of state compositional model. *Society of Petroleum Engineers Journal*, 20(05):363–376, 10 1980.

[10] Ş. Cobzaş, R. Miculescu, and A. Nicolae. *Lipschitz Functions*, volume 2241 of *Lecture Notes in Mathematics*. Springer Cham, 2019.

[11] J. P. Dussault, M. Frappier, and J. Ch. Gilbert. Polyhedral Newton-min algorithms for complementarity problems. *Mathematical Programming*, May 2025.

[12] F. Facchinei and J.-S. Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems*, volume Volume I of *Springer Series in Operations Research and Financial Engineering*. Springer New York, 2003.

[13] F. Facchinei and J.-S. Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems*, volume Volume II of *Springer Series in Operations Research and Financial Engineering*. Springer New York, 2003.

[14] A. Fischer. A special newton-type optimization method. *Optimization*, 24(3-4):269–284, 1992.

[15] J. Gondzio. Interior point methods 25 years later. *European Journal of Operational Research*, 218(3):587–601, 2012.

[16] 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, 2014.

[17] J. W. Johnson, E. H. Oelkers, and H. C. Helgeson. SUPCRT92: A software package for calculating the standard molal thermodynamic properties of minerals, gases, aqueous species, and reactions from 1 to 5000 bar and 0 to 1000°C. *Computers & Geosciences*, 18(7):899–947, 1992.

[18] M. Jonval, I. Ben Gharbia, C. Cancès, T. Faney, and Q.-H. Tran. Parametrization and cartesian representation techniques for robust resolution of chemical equilibria. *Journal of Computational Physics*, 522:113596, 2025.

[19] A. Lauser, C. Hager, R. Helmig, and B. Wohlmuth. A new approach for phase transitions in miscible multi-phase flow in porous media. *Advances in Water Resources*, 34(8):957–966, 2011.

[20] A. M. M. Leal, D. A. Kulik, W. R. Smith, and M. O. Saar. An overview of computational methods for chemical equilibrium and kinetic calculations for geochemical and reactive transport modeling. *Pure and Applied Chemistry*, 89(5):597–643, 2017.

[21] M. L. Michelsen. The isothermal flash problem. part ii. phase-split calculation. *Fluid Phase Equilibria*, 9(1):21–40, 1982.

[22] D. K. Nordstrom and K. M. Campbell. Modeling low-temperature geochemical processes. In *Treatise on Geochemistry*, pages 27–68. Elsevier, 2014.

[23] J.-S. Pang. Newton's method for b-differentiable equations. *Mathematics of Operations Research*, 15(2):311–341, 1990.

[24] L. Qi and J. Sun. A nonsmooth version of newton's method. *Mathematical Programming*, 58:353–367, 1993.

[25] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, 1970.

[26] N. Z. Shapiro and L. S. Shapley. Mass action laws and the Gibbs free energy function. *Journal of the Society for Industrial and Applied Mathematics*, 13(2):353–375, 1965.

[27] J. V. Smith, R. W. Missen, and W. R. Smith. General optimality criteria for multiphase multireaction chemical equilibrium. *AIChE Journal*, 39(4):707–710, 1993.

[28] W. R. Smith and R. W. Missen. *Chemical reaction equilibrium analysis: Theory and algorithms*. John Wiley & Sons, New York, 1982.

[29] D. T. S. Vu. Numerical resolution of algebraic systems with complementarity conditions: Application to the thermodynamics of compositional multiphase mixtures. *Université Paris-Saclay*, 2020.

[30] D. T. S. Vu, I. Ben Gharbia, M. Haddou, and Q.-H. Tran. A new approach for solving nonlinear algebraic systems with complementarity conditions. application to compositional multiphase equilibrium problems. *Mathematics and Computers in Simulation*, 190:1243–1274, 2021.

[31] T.J. Wolery. EQ3NR, a computer program for geochemical aqueous speciation-solubility calculations: Theoretical manual, user's guide, and related documentation (Version 7.0); Part 3. Technical Report UCRL-MA–110662-Pt.3, 138643, Lawrence Livermore National Laboratory, 1992.

[32] A. Wächter and L. T. Biegler. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical Programming*, 106(1):25–57, 2006.