A First-Order Decomposition Algorithm for Generalized Nash Equilibrium Problems

Jong Gwang Kim
Coles College of Business, Kennesaw State University, Kennesaw, GA 30144, jkim311@kennesaw.edu

This paper presents a new primal-dual method for computing an equilibrium of a generalized Nash equilibrium problem (GNEP), where each player's feasible strategy set depends on the other players' strategies. The method is based on a new form of Lagrangian with a quadratic approximation. First, we reformulate a GNEP as a saddle point computation problem using the new Lagrangian and establish equivalence between a saddle point of the Lagrangian and an equilibrium of the GNEP. We propose a simple first-order algorithm that is convergent to the saddle point. Furthermore, we establish global convergence under the assumption of *Kurdyka-Łojasiewicz* property. Our method has two novel features over existing approaches: (i) it requires neither boundedness assumptions on the strategy set and the set of multipliers of each player, nor boundedness assumptions on the iterates generated by the algorithm; (ii) to the best of our knowledge, it is the first development of a *first-order distributed* method to solve a *general class of GNEPs.* Numerical experiments are performed on test problems, and the results demonstrate the effectiveness of the proposed method.

1. Introduction

We consider generalized Nash equilibrium problems (GNEPs) that describe a broad class of non-cooperative games, in which each player seeks to optimize her/his own objective function while subject to certain constraints that are affected by the other players' strategies. The standard Nash game (Nash 1950) is a subclass of GNEPs, as the strategic interactions among players in a Nash game are only reflected in their objective functions, not in the constraints. Specifically, the game features a set of N players denoted by $\mathcal{N} = \{1,\ldots,N\}$ where each player ν has its own strategy $x^\nu \in \mathbb{R}^{n_\nu}$. Each player ν has an objective function $\theta_\nu(x^\nu, x^{-\nu})$ and a finite set of *coupling constraints* $g_i^\nu(x^\nu, x^{-\nu}) \le 0$ $(i = 1,\ldots,m_\nu)$, both of which depend on player ν's own strategy x^ν as well as other players' strategies $x^{-\nu} := (x^{\nu'})_{\nu' \neq \nu}$. Denote all players' strategies by a vector $\mathbf{x} = (x^\nu, x^{-\nu}) := (x^1,\ldots,x^\nu,\ldots,x^N)$ with dimension $n = \sum_{\nu=1}^N n_\nu$. The GNEP can be formally defined as a problem of finding a solution for each of the following problems. Given other players' strategies $x^{-\nu}$, each player ν seeks to find a strategy x^ν that solves the optimization problem:

$$\begin{aligned} \underset{x^\nu \in \mathcal{X}_\nu}{\text{minimize}} \quad & \theta_\nu(x^\nu, x^{-\nu}) \\ \text{subject to} \quad & g_i^\nu(x^\nu, x^{-\nu}) \le 0, \quad i = 1,\ldots,m_\nu, \end{aligned} \tag{1}$$

where $\mathcal{X}_\nu \subseteq \mathbb{R}^{n_\nu}$ represents the *private* strategy set of player ν that is nonempty, closed, and convex. The feasible strategy set of each player ν can be represented by the parametric inequalities:

$$\mathcal{F}_\nu(x^{-\nu}) := \left\{x^\nu \in \mathcal{X}_\nu : g_i^\nu(x^\nu, x^{-\nu}) \leq 0,\ i = 1, \dots, m_\nu\right\} \subseteq \mathbb{R}^{n_\nu}.$$

Note that for simplicity, private functional constraints $c_j^\nu(x^\nu) \leq 0$ for $j = 1, \dots, p_\nu$ are not explicitly highlighted in the paper. They can be easily handled in the same way to deal with $g_i^\nu(x^\nu, x^{-\nu}) \leq 0$. Here, n_ν, m_ν, and p_ν are positive integers. The set $\mathcal{X}_\nu$ is defined as $\mathcal{X}_\nu := \{x^\nu \in \mathbb{R}^{n_\nu} \mid l_\nu \leq x^\nu \leq u_\nu\}$, where l_ν or u_ν may be unbounded; that is, $l_\nu = -\infty$ or $u_\nu = +\infty$ or both.

A Nash equilibrium of the GNEP can be defined as follows.

Definition 1. A collection of strategies $\mathbf{x}^* = (x^{1,*}, \dots, x^{N,*})$ is a (pure-strategy) generalized Nash equilibrium (GNE) if for every $\nu = 1, \dots, N$,

$$\theta_\nu\left(x^{\nu,*}, x^{-\nu,*}\right) \leq \theta_\nu\left(x^\nu, x^{-\nu,*}\right), \quad \forall x^\nu \in \mathcal{F}_\nu(x^{-\nu,*}),$$

i.e., $\mathbf{x}^* = (x^{1,*}, \dots, x^{N,*})$ is a GNE, if and only if no player has an incentive to unilaterally deviate from $x^{\nu,*}$ when other players choose $x^{-\nu,*}$.

We make the following assumption on the functions throughout the paper.

Assumption 1. *For every $\nu \in \mathcal{N}$ and fixed $x^{-\nu}$, objective function $\theta_\nu(x^\nu, x^{-\nu})$ and constraint functions $g_i^\nu(x^\nu, x^{-\nu})$, $i = 1, \dots, m_\nu$, are continuously differentiable and convex with respect to x^ν.*

Note that $\theta_\nu(x^\nu, x^{-\nu})$ and $g_i^\nu(x^\nu, x^{-\nu})$ are possibly nonconvex in other players' strategies $x^{\nu'} \in x^{-\nu}$, and $g_i^\nu(x^\nu, x^{-\nu})$ are not necessarily shared by all players (*non-shared coupling constraints*). Under Assumption 1, problem (1) is known as a very general form of GNEP (Dreves et al. 2011) (We call it *general* GNEP).

In this paper, we aim to provide and analyze a *first-order decomposition* algorithm, based on a novel form of Lagrangian, to compute an equilibrium of the *general* GNEP, provided that equilibria of generalized Nash game exist.

We also make two standard assumptions; Lipschitz gradient continuity of the objective and constraint functions (smoothness) and coercivity of the objective functions.

Assumption 2. *For $\nu = 1, \dots, N$, the gradients of θ_ν and g^ν are $L_{\nabla\theta_\nu}$-Lipschitz continuous and $L_{\nabla g^\nu}$-Lipschitz continuous, respectively. That is,*

$$\|\nabla_{\mathbf{x}}\theta_\nu(\mathbf{x}_1) - \nabla_{\mathbf{x}}\theta_\nu(\mathbf{x}_2)\| \leq L_{\nabla\theta_\nu}\|\mathbf{x}_1 - \mathbf{x}_2\|, \quad \forall \mathbf{x}_1, \mathbf{x}_2 \in \mathbf{X}, \tag{2a}$$

$$\|\nabla_{\mathbf{x}}g^\nu(\mathbf{x}_1) - \nabla_{\mathbf{x}}g^\nu(\mathbf{x}_2)\| \leq L_{\nabla g^\nu}\|\mathbf{x}_1 - \mathbf{x}_2\|, \quad \forall \mathbf{x}_1, \mathbf{x}_2 \in \mathbf{X}, \tag{2b}$$

where $\nabla_{\mathbf{x}}\theta_\nu(\mathbf{x})$ and $\nabla_{\mathbf{x}}g^\nu(\mathbf{x})$ represent $(\nabla_{x^1}\theta_\nu(\mathbf{x}), \dots, \nabla_{x^N}\theta_\nu(\mathbf{x}))$ and $(\nabla_{x^1}g^\nu(\mathbf{x}), \dots, \nabla_{x^N}g^\nu(\mathbf{x}))$, respectively, and $\mathbf{X} := \prod_{\nu=1}\mathcal{X}_\nu$.

ASSUMPTION 3. *For every $\nu = 1,\ldots,N$, the objective function $\theta_\nu(x^\nu, x^{-\nu})$ is coercive with respect to $\mathbf{x} = (x^\nu, x^{-\nu}) \in \mathcal{X}_\nu \times \mathcal{X}_{-\nu}$, i.e., $\lim_{\|\mathbf{x}\|\to\infty} \theta_\nu(\mathbf{x}) = \infty$.*

Note that we do not impose the coercivity assumption on the feasible strategy sets, contrary to the interior-point algorithm (Dreves et al. 2011) for *general* GNEPs. The algorithm in Dreves et al. (2011) relies on the strong assumption that the feasible strategy sets of all players are bounded, i.e., $\lim_{\|\mathbf{x}\|\to\infty} \left\|g_+^\nu(\mathbf{x})\right\| = +\infty$ where $g_+^\nu(\mathbf{x}) := \max\{0, g^\nu(\mathbf{x})\}$ for all $\nu = 1,\ldots,N$.

1.1. Literature Review

The GNEP was originally introduced in seminal works by Debreu (1952) and Arrow and Debreu (1954) in the early 1950s, where the GNEP was referred to as a social equilibrium problem or an abstract economy. One important subclass of GNEPs, known as jointly-convex GNEPs (also called shared-constrained GNEPs), was first investigated by Rosen (1965).This class of GNEPs is characterized by shared constraints across all players, i.e., the convex coupling constraints are identical for all players ($g^1 = \cdots = g^N = g$). While early studies on GNEPs were primarily concerned with economics, recent decades have witnessed a growing interest in GNEPs as a modeling framework and solution concept in various application areas. Some examples include electricity market models (Jing-Yuan and Smeers 1999, Contreras et al. 2004, Hobbs and Pang 2007), power allocation in telecommunications (Pang et al. 2008, Scutari et al. 2014), environmental pollution control (Krawczyk and Uryasev 2000, Breton et al. 2006), transportation systems (Stein and Sudermann-Merx 2018), and cloud computing (Cardellini et al. 2016), to name a few.

Numerous algorithms have been developed for computing a GNE of a GNEP in the literature. One popular approach involves transforming a GNEP into a variational inequality (VI) problem and applying algorithms designed to find a solution of a VI reformulation, i.e., *variational equilibrium* (VE) or also called *normalized Nash equilibrium* (Facchinei and Kanzow 2010a); see e.g., Harker (1991), Pang and Fukushima (2005), Facchinei et al. (2007), Nabetani et al. (2011), Yin et al. (2011), Kulkarni and Shanbhag (2012), Migot and Cojocaru (2020). The VI approach simplifies solving the GNEP to finding a solution for a VI, instead of solving a more complicated quasi-variational inequality (QVI) as required for a GNEP (Facchinei and Kanzow 2010a). Importantly, the set of VEs is known to be a subset of GNEs (Ba and Pang 2022). However, a notable limitation of the VI-based approach is that it is only applicable to jointly-convex GNEPs.

Another widely used method for computing GNE involves reformulating a GNEP into a global optimization problem via the Nikaido-Isoda (NI) function (Nikaidô and Isoda 1955). The resulting optimization problem is then solved using the so-called relaxation algorithms (Uryas'ev and Rubinstein 1994, Krawczyk and Uryasev 2000, Contreras et al. 2004, Von Heusinger and Kanzow 2009a,b). However, these methods are also restricted to jointly-convex GNEPs and are known to be computationally expensive. Other algorithms designed for this class of GNEPs include Newton-type methods (Facchinei et al. 2009, von Heusinger et al. 2012, Izmailov and Solodov 2014) and Lemeke's method (Schiro et al. 2013) for specifically affine GNEPs.

Another line of relevant work is concerned with distributed algorithms for solving the GNEPs. In the context of primal-dual schemes for computing GNE, there has been a surge of interest in developing distributed primal-dual schemes for computing GNE for shared-constrained GNEPs (Zhu and Frazzoli 2016, Grammatico 2017, Paccagnan et al. 2018, Yi and Pavel 2018, 2019, Deng 2021, Cenedese et al. 2021, Migot and Cojocaru 2021, Belgioioso et al. 2022). These methods, however, are applicable only to jointly-convex GNEPs (GNEPs with shared constraints or affine coupling constraints). Distributed algorithms to date have been relying on monotonicity properties, which do not generally hold in *general* GNEPs we're focusing on in this paper.

The equilibrium computation of GNEPs beyond the class of jointly-convex GNEPs remains a very challenging task. This is mainly due to interdependence between each player's strategy and some other players' strategies through both objective and coupling constraints, along with the potential nonconvexity of each player's problem in the other players' strategies. A few algorithms have indeed been proposed, including penalty-type methods (Pang and Fukushima 2005, Facchinei and Kanzow 2010b, Kanzow and Steck 2018, Ba and Pang 2022), interior point algorithm (Dreves et al. 2011), and augmented Lagrangian method (Kanzow and Steck 2016, 2018).

In all such methods, it is assumed that the Extended Mangasarian-Fromovitz Constraint Qualification (EMFCQ), an extension of the MFCQ for infeasible points, holds for every player[1]. However, this EMFCQ is a restrictive assumption because it is equivalent to the set of the multipliers of each player being bounded (Nocedal and Wright 2006). This assumption is often violated in the context of GNEPs, as illustrated by simple examples in Dorsch et al. (2013). This violation occurs due to the interdependency between x^ν and $x^{-\nu}$ through coupling constraints $g_i^\nu(x^\nu, x^{-\nu}) \le 0,\ i = 1,\dots,m_\nu$, where the gradients of constraints can be (positively) linear dependent. In such cases, algorithms can generate unbounded function values, which can lead to failures of convergence to GNEs or even feasible points.

Penalty-based algorithms reduce the GNEP to a standard Nash equilibrium problem (NEP) by penalizing coupling constraints and focus on updating the penalty parameter. In particular, the exact penalty method in Facchinei and Kanzow (2010b) results in nonsmooth subproblems, so it obtains a GNE under various differentiability assumptions on the objectives and constraints. This lack of differentiability is a serious problem for designing efficient algorithms. To address the

[1] For all $\nu = 1,\dots,N$ and for $\mathbf{x} = (x^\nu, x^{-\nu})$, there exists a vector $d^\nu \in \mathbb{R}^{n_\nu}$ such that $\nabla_{\mathbf{x}} g_i^\nu(\mathbf{x})^T d^\nu < 0 \quad \forall i \in I_{\geq}^\nu(\mathbf{x})$, where $I_{\geq}^\nu(\mathbf{x}) = \{i \in \{1,\dots,m_\nu\} : g_i^\nu(\mathbf{x}) \geq 0\}$ denotes the set of active or violated constraints for player ν (Facchinei and Kanzow 2010b, Kanzow and Steck 2016).

drawbacks of penalty-based methods, Kanzow and Steck (2016) proposed an augmented Lagrangian method. This approach requires an assumption that there exists a limit point of the sequence $\{\mathbf{x}^k\}$. However, this assumption is not clear without compactness of each player's private set.

It is noteworthy that even with coercivity assumption on the objective function, the augmented Lagrangian (AL) method (Kanzow and Steck 2016) does not guarantee boundedness of primal/dual sequences. To ensure the boundedness of the sequences, bounded level sets of AL functions are needed, but they are typically unbounded. This is mainly related to the behavior of the multiplier sequence $\{\lambda^{\nu,k}\}$. Specifically, the AL method (Kanzow and Steck 2016) is of min-max dynamics (due to the increase in the dual variables), and by nature, the AL function alternatively increases and decreases, and the dual sequence $\{\lambda^{\nu,k}\}$ might be unbounded. Hence, the coercivity of the objective function does not imply the boundedness of primal and dual sequences in the AL framework.

1.2. Our Contributions

This paper presents a novel algorithmic framework for computing an equilibrium of a *general* GNEP without imposing boundedness assumptions on primal-dual sequences and (feasible) strategy sets.

- We introduce a new Lagrangian combined with artificial variables to reduce the GNEP to a standard Nash game, where the artificial variables are used to get rid of the coupling constraints while regularization terms lead to strong concavity of the Lagrangian in the multipliers. This allows for the design of an algorithm that generates a bounded primal-dual sequence without imposing EMFCQ assumption and removes computational effort in updating the penalty parameter, as in Facchinei and Kanzow (2010b) and Kanzow and Steck (2016).
- The proposed algorithm can effectively handle the potential nonconvexity of each player's functions with respect to other players' strategies by utilizing a simple quadratic approximation of P-Lagrangian. This quadratic approximation also provides a first-order decomposition scheme, enabling distributed updates of primal variables. As a result, this algorithm represents the first *distributed* approach to solving *general GNEPs*.
- We prove that our algorithm is convergent to a saddle point of P-Lagrangian under standard assumptions. Unlike existing methods for general GNEPs, our analysis does not require boundedness assumption on the iterates generated by the algorithm. We also do not use safeguarding technique (Andreani et al. 2007, 2008) to bound multiplier iterates as in Kanzow and Steck (2016). We establish the global convergence under an additional assumption that the objective and constraint functions satisfy the *Kurdyka-Lojasiewicz* property.

Outline of the paper. This paper is organized as follows. In section 2, we introduce the P-Lagrangian function, describe its characteristics, and reformulate the GNEP as a saddle point computation problem using the P-Lagrangian. Section 3 presents a distributed first-order primal-dual algorithm based on a quadratic approximation. In Section 4, we establish the convergence properties of the proposed algorithm. Numerical results are presented in Section 5.

Notation. We use $\mathbb{R}^{n_\nu}$ and $\mathbb{R}^{m_\nu}$ to denote the n_ν-dimensional Euclidean vector space and m_ν-dimensional Euclidean vector space, respectively. For two vectors $x, y \in \mathbb{R}^{n_\nu}$, the inner product is denoted by $x^T y$, and the standard Euclidean norm is denoted by $\|x\| = \sqrt{x^T x}$. For a real scalar $z \in \mathbb{R}$, we define $[z]^+ = \max\{z, 0\}$. We use $\mathbb{R}_+^{m_\nu}$ to denote the nonnegative orthant of $\mathbb{R}^{m_\nu}$, and the notation $x \geq 0$ denotes that the vector x belongs to the nonnegative orthant.

2. Proximal-Perturbed Lagrangian Formulation

Before introducing Proximal-Perturbed Lagrangian (P-Lagrangian), we recall that under Assumption 1 and suitable constraint qualifications, a GNE $\mathbf{x}^* = (x^{1,*}, \ldots, x^{N,*})$ can be characterized by the Karush-Kuhn-Tucker (KKT) conditions (Facchinei and Kanzow 2010a, Kanzow and Steck 2016):

The KKT conditions. Assume that a suitable constraint qualification holds. If there exists a point $\mathbf{x}^* = (x^{1,*}, \ldots, x^{N,*})$ together with some Lagrange multipliers $\eta^{\nu,*}$ satisfying the KKT conditions:

$$\begin{cases} 0 \in \nabla_{x^\nu} L_0^\nu(x^{\nu,*}, x^{-\nu,*}, \eta^{\nu,*}) + \mathcal{N}_{\mathcal{X}_\nu}(x^{\nu,*}), \quad x^{\nu,*} \in \mathcal{X}_\nu, \\ \eta_i^{\nu,*} \geq 0, \quad g_i^\nu(x^{\nu,*}, x^{-\nu,*}) \leq 0, \quad \eta_i^{\nu,*} g_i^\nu(x^{\nu,*}, x^{-\nu,*}) = 0, \quad \forall i = 1, \ldots, m_\nu, \end{cases} \tag{3}$$

for every $\nu = 1, \ldots, N$, then $\mathbf{x}^* = (x^{1,*}, \ldots, x^{N,*})$ is a *generalized Nash equilibrium* (GNE). Here, $L_0^\nu(x^\nu, x^{-\nu}, \eta^\nu) := \theta_\nu(x^\nu, x^{-\nu}) + \sum_{i=1}^{m_\nu} (\eta_i^\nu) g_i^\nu(x^\nu, x^{-\nu})$ is each player ν's Lagrangian, and $\mathcal{N}_{\mathcal{X}_\nu}(x^{\nu,*}) := \{\mathrm{d}_\nu \in \mathcal{X}_\nu : \mathrm{d}_\nu^T(x^\nu - x^{\nu,*}) \leq 0, \forall x^\nu \in \mathcal{X}_\nu\}$ is the normal cone to $\mathcal{X}_\nu$ at $\mathbf{x}^*$.

Assuming a suitable constraint qualification (CQ) holds and under the convexity assumption of the functions $\theta_\nu(\cdot, x^{-\nu})$ and $g_i^\nu(\cdot, x^{-\nu})$ (see Assumption 1), the KKT conditions in (3) become necessary and sufficient optimality conditions for problem (1) (Facchinei and Kanzow 2010a, Theorem 4.6). In addition, problem (1) is equivalent to solving the dual formulation, i.e.,

$$\theta_\nu(\mathbf{x}^*) = \max_{\eta^\nu \geq 0} \left(D_0^\nu(\eta^\nu) := \min_{x^\nu \in \mathcal{X}_\nu} L_0^\nu\left(x^\nu, x^{-\nu,*}, \eta^\nu\right) \right). \tag{4}$$

In the general GNEP model, the multiplier set of each player can be unbounded, even when satisfying the KKT conditions in (3). As previously mentioned, this unboundedness results from the inherent characteristics of general GNEPs, where the gradients of the constraints at the point $\mathbf{x}^*$ can be (positively) linear dependent, leading to an unbounded multiplier set. This aspect complicates the computation of a GNE, thus making the boundedness of multipliers a key issue when solving GNEPs. Our motivation for introducing a new Lagrangian is to address this challenge.

This section introduces a new form of Lagrangian that has a desirable structure for equilibrium computation. We then show that computing a saddle point of the P-Lagrangian is equivalent to finding an equilibrium of the GNEP (1).

2.1. The Proximal-Perturbed Lagrangian

Motivated by the reformulation techniques in Bertsekas and Tsitsiklis (1989, Chapter 3.4) and Bertsekas (2014, Chapter 3.2), we start by transforming problem (1) into an equivalent extended formulation by introducing perturbation variables $z^\nu = (z_1^\nu, \ldots, z_{m_\nu}^\nu) = 0$ as additional constraints and letting $g^\nu(x^\nu, x^{-\nu}) \le z^\nu$ given $x^{-\nu}$:

$$\begin{aligned} \underset{x^\nu \in \mathcal{X}_\nu,\, z^\nu \in \mathbb{R}^{m_\nu}}{\text{minimize}} \quad & \theta_\nu(x^\nu, x^{-\nu}) \\ \text{subject to} \quad & g^\nu(x^\nu, x^{-\nu}) \le z^\nu, \quad z^\nu = 0. \end{aligned} \tag{5}$$

Obviously, for $z^\nu = 0$, the extended formulation is equal to problem (1). Noting that the reformulation (5) allows the use of $\frac{\alpha_\nu}{2}\|z^\nu\|^2$ as a penalty term, let us first consider the following partially augmented Lagrangian for every $\nu = 1, \ldots, N$:

$$L_\alpha^\nu(x^\nu, x^{-\nu}, z^\nu, \lambda^\nu, \mu^\nu) = \theta_\nu(x^\nu, x^{-\nu}) + (\lambda^\nu)^T \left(g^\nu\left(x^\nu, x^{-\nu}\right) - z^\nu\right) + (\mu^\nu)^T z^\nu + \frac{\alpha_\nu}{2}\|z^\nu\|^2,$$

where $\lambda^\nu = (\lambda_i^\nu, \ldots, \lambda_{m_\nu}^\nu) \in \mathbb{R}_+^{m_\nu}$ and $\mu^\nu = (\mu_i^\nu, \ldots, \mu_{m_\nu}^\nu) \in \mathbb{R}^{m_\nu}$ are the Lagrange multipliers associated with constraints $g^\nu(x^\nu, x^{-\nu}) - z^\nu \le 0$ and $z^\nu = 0$, respectively. $\alpha_\nu > 0$ is a penalty parameter. Observe that given (λ^ν, μ^ν), minimizing L_α^ν with respect to z^ν gives

$$z^\nu(\lambda^\nu, \mu^\nu) = \frac{1}{\alpha_\nu}(\lambda^\nu - \mu^\nu),$$

which implies that $\lambda^\nu = \mu^\nu$ at the unique (known) solution $z^{\nu,*} = 0$. Based on this relation of λ^ν and μ^ν from the optimality condition for z^ν, we add a proximal term $-\frac{\beta_\nu}{2}\|\lambda^\nu - \mu^\nu\|^2$ to define a Proximal-Perturbed Lagrangian (P-Lagrangian) as

$$\begin{aligned} \mathcal{L}_{\alpha\beta}^\nu(x^\nu, x^{-\nu}, z^\nu, \lambda^\nu, \mu^\nu) := {}& \theta_\nu(x^\nu, x^{-\nu}) + (\lambda^\nu)^T \left(g^\nu\left(x^\nu, x^{-\nu}\right) - z^\nu\right) + (\mu^\nu)^T z^\nu \\ & + \frac{\alpha_\nu}{2}\|z^\nu\|^2 - \frac{\beta_\nu}{2}\|\lambda^\nu - \mu^\nu\|^2, \end{aligned} \tag{6}$$

where $\beta_\nu > 0$ is a proximal regularization parameter.

We observe that the structure of the P-Lagrangian $\mathcal{L}_{\alpha\beta}^\nu$ in (6) differs from the standard augmented Lagrangian and its variants (see Hestenes 1969, Powell 1969, Rockafellar 1974, Bertsekas 2014, Birgin and Martínez 2014). It is characterized by the absence of penalty term for handling the coupling constraint $g^\nu(x^\nu, x^{-\nu}) - z^\nu \le 0$. Only additional constraint $z^\nu = 0$ is penalized with a quadratic penalty term $\frac{\alpha_\nu}{2}\|z^\nu\|^2$, while $g^\nu(x^\nu, x^{-\nu}) - z^\nu \le 0$ is merely relaxed into the objective with the corresponding multiplier. Second, the P-Lagrangian is strongly concave in λ^ν (for fixed μ^ν) and in μ^ν (for fixed λ^ν) due to the presence of the negative quadratic term $-\frac{\beta_\nu}{2}\|\lambda^\nu - \mu^\nu\|^2$. f

2.2. Equivalence between a Saddle Point of P-Lagrangian and a GNE

Now consider the following P-Lagrangian dual problem for given $x^{-\nu}$:

$$\max_{\lambda^\nu\in\mathbb{R}^{m_\nu}_+,\mu^\nu\in\mathbb{R}^{m_\nu}} \left\{\mathcal{D}^\nu_{\alpha\beta}(\lambda^\nu,\mu^\nu):=\min_{x^\nu\in\mathcal{X}_\nu,z^\nu\in\mathbb{R}^{m_\nu}} \mathcal{L}^\nu_{\alpha\beta}(x^\nu,x^{-\nu},z^\nu,\lambda^\nu,\mu^\nu)\right\}. \tag{7}$$

Since $\mathcal{L}^\nu_{\alpha\beta}(\bullet,x^{-\nu},z^\nu,\lambda^\nu,\mu^\nu)$ is convex, the primal-dual solutions of problem (7), $(x^{\nu,*},x^{-\nu,*},z^{\nu,*})$ and $(\lambda^{\nu,*},\mu^{\nu,*})$ given $x^{-\nu}=x^{-\nu,*}$, can be characterized by the saddle point of the P-Lagrangian.

Definition 2. Given $x^{-\nu,*}$, a point $(x^{\nu,*},x^{-\nu,*},z^{\nu,*},\lambda^{\nu,*},\mu^{\nu,*})$ is said to be a (parametrized) saddle point of the Proximal-Perturbed Lagrangian for $\alpha_\nu>0$ and $\beta_\nu>0$ if for every $\nu=1,\dots,N$,

$$\mathcal{L}^\nu_{\alpha\beta}(x^{\nu,*},x^{-\nu,*},z^{\nu,*},\lambda^\nu,\mu^\nu)\le\mathcal{L}^\nu_{\alpha\beta}(x^{\nu,*},x^{-\nu,*},z^{\nu,*},\lambda^{\nu,*},\mu^{\nu,*})\le\mathcal{L}^\nu_{\alpha\beta}(x^\nu,x^{-\nu,*},z^\nu,\lambda^{\nu,*},\mu^{\nu,*}) \tag{8}$$

for all $(x^\nu,z^\nu,\lambda^\nu,\mu^\nu)\in\mathcal{X}_\nu(x^{-\nu,*})\times\mathbb{R}^{m_\nu}\times\mathbb{R}^{m_\nu}_+\times\mathbb{R}^{m_\nu}$. Here, $x^{-\nu,*}$ are viewed as parameters.

We establish the equivalence between computing a saddle point of $\mathcal{L}^\nu_{\alpha\beta}$ and finding an equilibrium of the GNEP (1) by proving Theorems 1 and 2. Before studying the equivalence, let us observe the following properties of $\mathcal{L}^\nu_{\alpha\beta}(x^\nu,x^{-\nu},z^\nu,\lambda^\nu,\mu^\nu)$.

Observation 1. Notice that the inner minimization in (7) can be split into two parts as follows:

$$\max_{\lambda^\nu\in\mathbb{R}^{m_\nu}_+,\mu^\nu\in\mathbb{R}^{m_\nu}} \left\{\min_{x^\nu\in\mathcal{X}_\nu}\left[\theta_\nu\left(x^\nu,x^{-\nu}\right)+(\lambda^\nu)^T g^\nu\left(x^\nu,x^{-\nu}\right)\right]\right.$$
$$\left.+\min_{z^\nu\in\mathbb{R}^{m_\nu}}\left[-(\lambda^\nu-\mu^\nu)^T z^\nu+\frac{\alpha_\nu}{2}\|z^\nu\|^2\right]-\frac{\beta_\nu}{2}\|\lambda^\nu-\mu^\nu\|^2\right\}.$$

Denote by $z^\nu(\lambda^\nu,\mu^\nu)$ as a unique solution of the problem, $\min_{z^\nu\in\mathbb{R}^{m_\nu}}\left[-(\lambda^\nu-\mu^\nu)^T z^\nu+\frac{\alpha_\nu}{2}\|z^\nu\|^2\right]$ for given (λ^ν,μ^ν). If we minimize $\left[-(\lambda^\nu-\mu^\nu)^T z^\nu+\frac{\alpha_\nu}{2}\|z^\nu\|^2\right]$ with respect to z^ν, we have

$$z^\nu(\lambda^\nu,\mu^\nu)=\frac{1}{\alpha_\nu}(\lambda^\nu-\mu^\nu)\quad\Longleftarrow\quad(\mu^\nu-\lambda^\nu)+\alpha_\nu z^\nu=0.$$

Recall that based on the optimality condition for z^ν, we added a quadratic regularization term $-\frac{\beta_\nu}{2}\|\lambda^\nu-\mu^\nu\|^2$ to make the Lagrangian strongly concave in λ^ν (for fixed μ^ν) and in μ^ν (for fixed λ^ν) as it vanishes at $z^{\nu,*}=0$. Substituting $z^\nu(\lambda^\nu,\mu^\nu)$ into $\mathcal{L}^\nu_{\alpha\beta}(x^\nu,x^{-\nu},z^\nu,\lambda^\nu,\mu^\nu)$, $\mathcal{L}^\nu_{\alpha\beta}$ reduces to

$$\mathcal{L}^\nu_{\alpha\beta}\left(x^\nu,x^{-\nu},z^\nu(\lambda^\nu,\mu^\nu),\lambda^\nu,\mu^\nu\right)=\theta_\nu\left(x^\nu,x^{-\nu}\right)+(\lambda^\nu)^T g^\nu\left(x^\nu,x^{-\nu}\right)-\frac{1+\alpha_\nu\beta_\nu}{2\alpha_\nu}\|\lambda^\nu-\mu^\nu\|^2. \tag{9}$$

Then the P-Lagrangian dual problem can be expressed as

$$\max_{\lambda^\nu\in\mathbb{R}^{m_\nu}_+,\mu^\nu\in\mathbb{R}^{m_\nu}}\left\{\mathcal{D}^\nu_{\alpha\beta}(\lambda^\nu,\mu^\nu)\triangleq D^\nu_0(\lambda^\nu)-\frac{1+\alpha_\nu\beta_\nu}{2\alpha_\nu}\|\lambda^\nu-\mu^\nu\|^2\right\}, \tag{10}$$

where $D^\nu_0(\lambda^\nu)=\min_{x^\nu\in\mathcal{X}_\nu}\left\{\theta_\nu(x^\nu,x^{-\nu})+(\lambda^\nu)^T g^\nu(x^\nu,x^{-\nu})\right\}$, which is identical to the standard dual function associated with the original problem (1). Thus the P-Lagrangian dual function $\mathcal{D}^\nu_{\alpha\beta}(\lambda^\nu,\mu^\nu)$ is maximized jointly in λ^ν and μ^ν if and only if λ^ν maximizes $D^\nu_0(\lambda^\nu)$ and $\lambda^\nu=\mu^\nu$. This implies that the multiplier $\lambda^{\nu,*}$ for the constraint $g^\nu(x^\nu,x^{-\nu})-z^\nu\le 0$ in extended problem (5) is precisely to the multiplier $\eta^{\nu,*}$ for the constraint $g^\nu(x^\nu,x^{-\nu})\le 0$ in problem (1).

OBSERVATION 2. If we maximize $\mathcal{L}^{\nu}_{\alpha\beta}(x^\nu, x^{-\nu}, z^\nu, \lambda^\nu, \mu^\nu)$ with respect to μ^ν, we get

$$\nabla_{\mu^\nu} \mathcal{L}^{\nu}_{\alpha\beta}\left(x^\nu, x^{-\nu}, z^\nu, \lambda^\nu, \mu^\nu\right) = z^\nu + \beta_\nu\left(\lambda^\nu - \mu^\nu\right) = 0,$$

which, along with the fact $\lambda^{\nu,*} = \mu^{\nu,*}$, implies that $z^\nu = 0$ for maximizers $(\lambda^{\nu,*}, \mu^{\nu,*})$ and $\beta_\nu > 0$.

Using Observations 1 and 2, we now show the equivalence between a saddle point of $\mathcal{L}^{\nu}_{\alpha\beta}$ and an equilibrium of the GNEP (1).

THEOREM 1. *Let $(x^{\nu,*}, x^{-\nu,*}, z^{\nu,*}, \lambda^{\nu,*}, \mu^{\nu,*})$ be a saddle point of $\mathcal{L}^{\nu}_{\alpha\beta}(x^\nu, x^{-\nu}, z^\nu, \lambda^\nu, \mu^\nu)$ for a given $x^{-\nu} = x^{-\nu,*}$ and for some $\alpha_\nu > 0$ and $\beta_\nu > 0$. Then, $\mathbf{x}^* = (x^{\nu,*}, x^{-\nu,*})$ is an equilibrium of the GNEP (1) for every $\nu = 1, \ldots, N$.*

Proof. Using the reduced P-Lagrangian (9), we have

$$\begin{aligned}\mathcal{L}^{\nu}_{\alpha\beta}\left(\mathbf{x}^*, z^{\nu,*}\left(\lambda^\nu, \mu^\nu\right), \lambda^\nu, \mu^\nu\right) &= \theta_\nu\left(\mathbf{x}^*\right) + \left(\lambda^\nu\right)^T g^\nu\left(\mathbf{x}^*\right) - \frac{1+\alpha_\nu\beta_\nu}{2\alpha_\nu}\left\|\lambda^\nu - \mu^\nu\right\|^2 \\ &\le \mathcal{L}^{\nu}_{\alpha\beta}\left(\mathbf{x}^*, z^{\nu,*}\left(\lambda^{\nu,*}, \mu^{\nu,*}\right), \lambda^{\nu,*}, \mu^{\nu,*}\right).\end{aligned} \tag{11}$$

First, we prove that $\mathbf{x}^* = (x^{\nu,*}, x^{-\nu,*})$ is feasible for problem (1). Suppose by contradiction that $\mathbf{x}^*$ is infeasible, i.e., $g_i^\nu(\mathbf{x}^*) > 0$ for some i. Then there exist some λ_i^ν such that $\lambda_i^\nu g_i^\nu(\mathbf{x}^*) \to \infty$ as $\lambda_i^\nu \to \infty$. This implies that $\mathcal{L}^{\nu}_{\alpha\beta}(\mathbf{x}^*, z^{\nu,*}, \lambda^\nu, \mu^\nu) \to \infty$ by taking the limit as $\lambda_i^\nu \to \infty$ with $\lambda_i^\nu = \mu_i^\nu$ to maximize the left-hand side of the first inequality in (11), which is a contradiction with the first inequality in (8). Therefore, $g_i^\nu(\mathbf{x}^*) \le 0$ for all $i = 1, \ldots, m_\nu$. By the definition $\mathcal{L}^{\nu}_{\alpha\beta}(\mathbf{x}^*, z^{\nu,*}, \lambda^{\nu,*}, \mu^{\nu,*}) = \sup_{\lambda^\nu \ge 0, \mu^\nu} \mathcal{L}^{\nu}_{\alpha\beta}(\mathbf{x}^*, z^{\nu,*}, \lambda^\nu, \mu^\nu)$ with the fact that $g^\nu(\mathbf{x}^*) \le 0$ and $\lambda^{\nu,*} \ge 0$, we have $(\lambda^{\nu,*})^T g^\nu(\mathbf{x}^*) = 0$ and $\lambda^{\nu,*} = \mu^{\nu,*}$. It thus follows that

$$\mathcal{L}^{\nu}_{\alpha\beta}\left(\mathbf{x}^*, z^{\nu,*}, \lambda^{\nu,*}, \mu^{\nu,*}\right) = \theta_\nu\left(\mathbf{x}^*\right).$$

Next, let $x^\nu \in X_\nu(x^{-\nu,*})$ be any feasible solution to problem (1). For any feasible x^ν and $\lambda_i^\nu \ge 0$, since $g_i^\nu(x^\nu, x^{-\nu,*}) \le 0$, we have

$$\left(\lambda^\nu\right)^T g^\nu\left(x^\nu, x^{-\nu,*}\right) - \frac{1+\alpha_\nu\beta_\nu}{2\alpha_\nu}\left\|\lambda^\nu - \mu^\nu\right\|^2 \le \left(\lambda^\nu\right) \cdot 0 - \frac{1+\alpha_\nu\beta_\nu}{2\alpha_\nu}\left\|\lambda^\nu - \mu^\nu\right\|^2 \le 0. \tag{12}$$

From Observation 2 that $z^\nu = 0$ when $\lambda^{\nu,*} = \mu^{\nu,*}$ for any $\beta_\nu > 0$, we have

$$-\left(\lambda^{\nu,*} - \mu^{\nu,*}\right)^T z^\nu + \frac{\alpha_\nu}{2}\left\|z^\nu\right\|^2 = 0. \tag{13}$$

The second inequality of the saddle point condition (8) yields

$$\begin{aligned}\theta_\nu\left(\mathbf{x}^*\right) &\le \mathcal{L}^{\nu}_{\alpha\beta}\left(x^\nu, x^{-\nu,*}, z^\nu, \lambda^{\nu,*}, \mu^{\nu,*}\right) \\ &= \theta_\nu\left(x^\nu, x^{-\nu,*}\right) + \underbrace{\left(\lambda^{\nu,*}\right)^T g^\nu\left(x^\nu, x^{-\nu,*}\right) - \frac{\beta_\nu}{2}\left\|\lambda^{\nu,*} - \mu^{\nu,*}\right\|^2}_{\le 0} - \underbrace{\left(\lambda^{\nu,*} - \mu^{\nu,*}\right)^T z^\nu + \frac{\alpha_\nu}{2}\left\|z^\nu\right\|^2}_{=0} \\ &\le \theta_\nu\left(x^\nu, x^{-\nu,*}\right),\end{aligned}$$

where the last inequality is from (12) and (13). Hence, $\mathbf{x}^*$ is a GNE of problem (1). $\square$

THEOREM 2. *Assume that $\mathbf{x}^*=(x^{1,*},\ldots,x^{N,*})$ is an equilibrium of the GNEP (1) at which the KKT conditions (3) hold with some Lagrange multipliers $\eta^{\nu,*}$ for all players' problems, given $x^{-\nu}=x^{-\nu,*}$. Then for $\nu=1,\ldots,N$, there exist Lagrange multipliers $(\lambda^{\nu,*},\mu^{\nu,*})$ such that*

$$\mathcal{L}^{\nu}_{\alpha\beta}\left(\mathbf{x}^*,z^{\nu,*},\lambda^{\nu},\mu^{\nu}\right)\le\mathcal{L}^{\nu}_{\alpha\beta}\left(\mathbf{x}^*,z^{\nu,*},\lambda^{\nu,*},\mu^{\nu,*}\right)\le\mathcal{L}^{\nu}_{\alpha\beta}\left(x^{\nu},x^{-\nu,*},z^{\nu},\lambda^{\nu,*},\mu^{\nu,*}\right),\tag{14}$$

for any $(x^{\nu},z^{\nu},\lambda^{\nu},\mu^{\nu})\in\mathcal{X}_{\nu}(x^{-\nu,*})\times\mathbb{R}^{m_\nu}\times\mathbb{R}^{m_\nu}_{+}\times\mathbb{R}^{m_\nu}$,

Proof. From the feasibility of a GNE $\mathbf{x}^*$, we have for any $\lambda^{\nu}\in\mathbb{R}^{m_\nu}_{+}$, $\mu^{\nu}\in\mathbb{R}^{m_\nu}$ and $\alpha_\nu,\beta_\nu>0$

$$\left(\lambda^{\nu}\right)^{T}g^{\nu}\left(\mathbf{x}^*\right)-\frac{1+\alpha_\nu\beta_\nu}{2\alpha_\nu}\left\|\lambda^{\nu}-\mu^{\nu}\right\|^{2}\le 0,\tag{15}$$

implying that $\theta_\nu(\mathbf{x}^*)+(\lambda^{\nu})^Tg^{\nu}(\mathbf{x}^*)-\frac{1+\alpha_\nu\beta_\nu}{2\alpha_\nu}\left\|\lambda^{\nu}-\mu^{\nu}\right\|^{2}=\mathcal{L}^{\nu}_{\alpha\beta}(\mathbf{x}^*,z^{\nu,*}(\lambda^{\nu},\mu^{\nu}),\lambda^{\nu},\mu^{\nu})\le\theta_\nu(\mathbf{x}^*)$. On the other hand, since there exists a pair the Lagrange multipliers $(\lambda^{\nu,*},\mu^{\nu,*})$ maximizing $\mathcal{L}^{\nu}_{\alpha\beta}(\mathbf{x}^*,z^{\nu,*},\lambda^{\nu},\mu^{\nu})$, we also have that for $\lambda^{\nu}=\mu^{\nu}=0$

$$\mathcal{L}^{\nu}_{\alpha\beta}\left(\mathbf{x}^*,z^{\nu,*}\left(0,0\right),0,0\right)\le\mathcal{L}^{\nu}_{\alpha\beta}\left(\mathbf{x}^*,z^{\nu,*},\lambda^{\nu,*},\mu^{\nu,*}\right)=\theta_\nu\left(\mathbf{x}^*\right)+\left(\lambda^{\nu,*}\right)^{T}g^{\nu}\left(\mathbf{x}^*\right)-\frac{1+\alpha_\nu\beta_\nu}{2\alpha_\nu}\left\|\lambda^{\nu,*}-\mu^{\nu,*}\right\|^{2},$$

which together with the fact that $\mathcal{L}^{\nu}_{\alpha\beta}\left(\mathbf{x}^*,z^{\nu,*}\left(0,0\right),0,0\right)=\theta_\nu\left(\mathbf{x}^*\right)$ gives

$$\left(\lambda^{\nu,*}\right)^{T}g^{\nu}\left(\mathbf{x}^*\right)-\frac{1+\alpha_\nu\beta_\nu}{2\alpha_\nu}\left\|\lambda^{\nu,*}-\mu^{\nu,*}\right\|^{2}\ge 0.\tag{16}$$

Combining (15) and (16), we obtain

$$\left(\lambda^{\nu,*}\right)^{T}g^{\nu}\left(\mathbf{x}^*\right)-\frac{1+\alpha_\nu\beta_\nu}{2\alpha_\nu}\left\|\lambda^{\nu,*}-\mu^{\nu,*}\right\|^{2}=0,$$

which implies that $\mathcal{L}^{\nu}_{\alpha\beta}\left(\mathbf{x}^*,z^{\nu,*},\lambda^{\nu,*},\mu^{\nu,*}\right)=\theta_\nu\left(\mathbf{x}^*\right)$. Thus, the first inequality in (14) holds.

Using the facts that $g^{\nu}\left(\mathbf{x}^*\right)\le 0$, $\lambda^{\nu,*}\ge 0$, and (16), we have that $0\ge\left(\lambda^{\nu,*}\right)^{T}g^{\nu}\left(\mathbf{x}^*\right)\ge\frac{1+\alpha_\nu\beta_\nu}{2\alpha_\nu}\left\|\lambda^{\nu,*}-\mu^{\nu,*}\right\|^{2}\ge 0$, which implies that the multiplier $\lambda^{\nu,*}$ satisfies the complementarity slackness $\left(\lambda^{\nu,*}\right)^{T}g^{\nu}(x^{\nu,*},x^{-\nu,*})=0$ and $\lambda^{\nu,*}=\mu^{\nu,*}$. Therefore, the maximizer $\lambda^{\nu,*}$ is equivalent to the Lagrange multiplier $\eta^{\nu,*}$ satisfying the KKT conditions (3) for the original GNEP (1).

Next, noting that $\nabla_{z^{\nu}}L^{\nu}_{\alpha}\left(z^{\nu},\lambda^{\nu,*},\mu^{\nu,*}\right)=-\left(\lambda^{\nu,*}-\mu^{\nu,*}\right)+\alpha_\nu z^{\nu,*}=0$ and $\lambda^{\nu,*}=\mu^{\nu,*}$, we get the minimum $z^{\nu,*}=0$ for $\alpha_\nu>0$. By the convexity of $\theta_\nu\left(x^{\nu},x^{-\nu,*}\right)$ and $g^{\nu}\left(x^{\nu},x^{-\nu,*}\right)$ in x^{ν}, we have

$$\begin{aligned}\theta_\nu\left(x^{\nu},x^{-\nu,*}\right)&\ge\theta_\nu\left(\mathbf{x}^*\right)+\nabla_{x^{\nu}}\theta_\nu\left(\mathbf{x}^*\right)^{T}\left(x^{\nu}-x^{\nu,*}\right),\\ g^{\nu}\left(x^{\nu},x^{-\nu,*}\right)&\ge g^{\nu}\left(\mathbf{x}^*\right)+\nabla_{x^{\nu}}g^{\nu}\left(\mathbf{x}^*\right)^{T}\left(x^{\nu}-x^{\nu,*}\right).\end{aligned}$$

Then we have

$$\begin{aligned}\mathcal{L}^{\nu}_{\alpha\beta}\left(x^{\nu},x^{-\nu,*},z^{\nu},\lambda^{\nu,*},\mu^{\nu,*}\right)&\ge\theta_\nu\left(\mathbf{x}^*\right)+\left(\lambda^{\nu,*}\right)^{T}g^{\nu}\left(\mathbf{x}^*\right)+\left(\nabla_{x^{\nu}}\theta_\nu\left(\mathbf{x}^*\right)+\sum_{i=1}^{m_\nu}\lambda^{\nu,*}_{i}\nabla_{x^{\nu}}g^{\nu}_{i}\left(\mathbf{x}^*\right)\right)^{T}\left(x^{\nu}-x^{\nu,*}\right)\\&\quad-\left(\lambda^{\nu,*}-\mu^{\nu,*}\right)^{T}z^{\nu}+\frac{\alpha_\nu}{2}\left\|z^{\nu}\right\|^{2}-\frac{\beta_\nu}{2}\left\|\lambda^{\nu,*}-\mu^{\nu,*}\right\|^{2}\\&\ge\theta_\nu\left(\mathbf{x}^*\right)+\left(\lambda^{\nu,*}\right)^{T}g^{\nu}\left(\mathbf{x}^*\right)+\frac{\alpha_\nu}{2}\left\|z^{\nu}\right\|^{2}\\&\ge\theta_\nu\left(\mathbf{x}^*\right)=\mathcal{L}^{\nu}_{\alpha\beta}(\mathbf{x}^*,z^{\nu,*}=0,\lambda^{\nu,*},\mu^{\nu,*}).\end{aligned}$$

Hence, $(\mathbf{x}^*,z^{\nu,*}=0)$ satisfies the second inequality of (14). $\square$

3. Algorithm

In this section, we propose a simple first-order primal-dual algorithm for computing a saddle point of $\mathcal{L}^{\nu}_{\alpha\beta}$ based on a quadratic approximation of $\mathcal{L}^{\nu}_{\alpha\beta}$ for every $\nu = 1, \ldots, N$.

3.1. Motivation for approximation of subproblems

We begin by describing briefly why we need to consider an approximation scheme for updating $\mathbf{x} = (x^{\nu}, x^{-\nu})$. To compute a saddle point of $\mathcal{L}^{\nu}_{\alpha\beta}(\mathbf{x}, z^{\nu}, \lambda^{\nu}, \mu^{\nu})$ for every $\nu = 1, \ldots, N$, we should be able to determine a point $\widetilde{\mathbf{x}} = (\widetilde{x}^{\nu}, \widetilde{x}^{-\nu})$ that satisfies the following first-order optimality (or simultaneous stationarity) condition of subproblems for fixed $(z^{\nu}, \lambda^{\nu}, \mu^{\nu})$ and all for $\nu = 1, \ldots, N$:

$$\nabla_{x^{\nu}} \mathcal{L}^{\nu}_{\alpha\beta}(\widetilde{\mathbf{x}}, z^{\nu}, \lambda^{\nu}, \mu^{\nu})^T (x^{\nu} - \widetilde{x}^{\nu}) \geq 0, \quad \forall x^{\nu} \in \mathcal{X}_{\nu}.$$

It is well known (Facchinei and Pang 2007) that for given $(z^{\nu,k}, \lambda^{\nu,k}, \mu^{\nu,k})$, computing such a stationary point is equivalent to the variational inequality (VI) problem of finding $\widetilde{\mathbf{x}} \in \mathbf{X}$ such that

$$\mathbf{L}\left(\widetilde{\mathbf{x}}, z^k, \lambda^k, \mu^k\right)^T (\mathbf{x} - \widetilde{\mathbf{x}}) \geq 0, \quad \forall \mathbf{x} \in \mathbf{X},$$

where $\mathbf{X} := \prod_{\nu=1}^{N} \mathcal{X}_{\nu}$, the Cartesian product of the private strategy sets of all players, and the mapping $\mathbf{L}\left(\mathbf{x}, z^k, \lambda^k, \mu^k\right) : \mathbf{X} \to \mathbb{R}^n$ is given by

$$\mathbf{L}\left(\mathbf{x}, z^k, \lambda^k, \mu^k\right) = \begin{bmatrix} \nabla_{x^1} \mathcal{L}^1_{\alpha\beta}\left(x^1, x^{-1}, z^{1,k}, \lambda^{1,k}, \mu^{1,k}\right) \\ \vdots \\ \nabla_{x^N} \mathcal{L}^N_{\alpha\beta}\left(x^N, x^{-N}, z^{N,k}, \lambda^{N,k}, \mu^{N,k}\right) \end{bmatrix},$$

with $z = [(z^1)^T, \ldots, (z^N)^T]$, $\lambda = [(\lambda^1)^T, \ldots, (\lambda^N)^T]^T$, and $\mu = [(\mu^1)^T, \ldots, (\mu^N)^T]^T$.

However, it is challenging to compute the point $\widetilde{\mathbf{x}}$ using descent methods. In the GNEP setting, each player's choice of strategy affects the optimization problems of the others through coupling constraints and objectives. More specifically, the monotonicity of the mapping $\mathbf{L}(\mathbf{x}, z^k, \lambda^k, \mu^k)$ with respect to $\mathbf{x} = (x^{\nu}, x^{-\nu})$ does not hold in general (Facchinei and Kanzow 2010a, Section 5.2) even if each component $\nabla_{x^{\nu}} \mathcal{L}^{\nu}_{\alpha\beta}(x^{\nu}, x^{-\nu}, z^{\nu}, \lambda^{\nu}, \mu^{\nu})$ is convex in x^{ν}. This nonconvexity of each P-Lagrangian in the other players' decision variables makes it hard to preserve a descent direction for the convergence to the stationary point $\widetilde{\mathbf{x}}$ that satisfies all components of the variational inequality.

3.2. Construction of Quadratic Approximation Model

To overcome such a computational difficulty, we consider a monotone approximation, denoted by $\widehat{\mathbf{L}}^k$, to the nonmonotone mapping $\mathbf{L}$ in $\mathbf{x}$. The monotone approximation $\widehat{\mathbf{L}}^k$ of the mapping $\mathbf{L}$ can be always chosen even if $\mathbf{L}$ is nonmonotone (see e.g., Chung and Fuller 2010, Luna et al. 2014). Furthermore, strongly monotone approximation mapping can be derived by replacing each player's $\mathcal{L}^{\nu}_{\alpha\beta}$ by a simple approximation function and then constructing an approximation $\widehat{\mathbf{L}}^k$.

To this end, inspired by Beck and Teboulle (2009) and Bolte et al. (2014), we first employ the following quadratic approximation $\widehat{\mathcal{L}^{\nu}_{\alpha\beta}}$ in only $\mathbf{x}$ at a given point $\mathbf{y}$:

$$\widehat{\mathcal{L}^{\nu}_{\alpha\beta}}(\mathbf{x}, z^{\nu}, \lambda^{\nu}, \mu^{\nu}; \mathbf{y}) := \mathcal{L}^{\nu}_{\alpha\beta}(\mathbf{y}, z^{\nu}, \lambda^{\nu}, \mu^{\nu}) + \nabla_{x^{\nu}} \mathcal{L}^{\nu}_{\alpha\beta}(\mathbf{y}, z^{\nu}, \lambda^{\nu}, \mu^{\nu})^T (x^{\nu} - y^{\nu}) + \frac{\gamma_{\nu}}{2} \|x^{\nu} - y^{\nu}\|^2 + \sum_{\nu' \neq \nu} \nabla_{x^{\nu'}} \mathcal{L}^{\nu}_{\alpha\beta}(\mathbf{y}, z^{\nu}, \lambda^{\nu}, \mu^{\nu})^T (x^{\nu'} - y^{\nu'}) + \frac{\gamma_{\nu}}{2} \sum_{\nu' \neq \nu} \left\| x^{\nu'} - y^{\nu'} \right\|^2, \quad (17)$$

which is a linearized P-Lagrangian $\mathcal{L}^{\nu}_{\alpha\beta}$ with respect to all other players' strategies $\mathbf{x} = (x^{\nu}, x^{-\nu})$ at the point $\mathbf{y}$ combined with quadratic proximal terms that measure the local error in the linear approximation. Here, $\gamma_{\nu} > 0$ is a proximal parameter. The term $\sum_{\nu' \neq \nu} \nabla_{x^{\nu'}} \mathcal{L}^{\nu}_{\alpha\beta}(\mathbf{y}, z^{\nu}, \lambda^{\nu}, \mu^{\nu}) = \nabla_{x^{-\nu}} \mathcal{L}^{\nu}_{\alpha\beta}(\mathbf{y}, z^{\nu}, \lambda^{\nu}, \mu^{\nu})$ represents the gradient at a given point $\mathbf{y} \in \mathbb{R}^n$ in other players' strategies.

As a direct consequence of the above Lipschitz continuity of $\nabla_{\mathbf{x}} \theta_{\nu}(\mathbf{x})$ and $\nabla_{\mathbf{x}} g^{\nu}(\mathbf{x})$, (2a) and (2b), respectively, we have the well-known descent Lemma.

LEMMA 1 **(Bertsekas 1999, Proposition A.24)**. *Let Assumptions 1 and 2 hold. Then for $\nu = 1, \ldots, N$ and for fixed $(z^{\nu}, \lambda^{\nu}, \mu^{\nu})$, $\nabla_{\mathbf{x}} \mathcal{L}^{\nu}_{\alpha\beta}$ is Lipschitz continuous with constant $L_{\nu} > 0$. Thus*

$$\mathcal{L}^{\nu}_{\alpha\beta}(\mathbf{x}_1) \leq \mathcal{L}^{\nu}_{\alpha\beta}(\mathbf{x}_2) + \nabla_{\mathbf{x}} \mathcal{L}^{\nu}_{\alpha\beta}(\mathbf{x}_2)^T (\mathbf{x}_1 - \mathbf{x}_2) + \frac{L_{\nu}}{2} \|\mathbf{x}_1 - \mathbf{x}_2\|^2, \quad \forall \mathbf{x}_1, \mathbf{x}_2 \in \mathbf{X}.$$

Here, we omit fixed $(z^{\nu}, \lambda^{\nu}, \mu^{\nu})$ for simplicity.

With the parameter $\gamma_{\nu} > 0$ such that $\gamma_{\nu} \geq L_{\nu}$, $\widehat{\mathcal{L}^{\nu}_{\alpha\beta}}(\mathbf{x}, z^{\nu}, \lambda^{\nu}, \mu^{\nu}; \mathbf{y})$ (17) is an *upper quadratic approximation* of $\mathcal{L}^{\nu}_{\alpha\beta}(\bullet, z^{\nu}, \lambda^{\nu}, \mu^{\nu})$ around the point $\mathbf{y}$ with respect to $\mathbf{x} = (x^{\nu}, x^{-\nu})$ and it has the following properties (see e.g., Beck and Teboulle 2009, Razaviyayn et al. 2013, Scutari et al. 2016).

REMARK 1. The quadratic approximation function $\widehat{\mathcal{L}^{\nu}_{\alpha\beta}}$ with $\gamma_{\nu} \geq L_{\nu}$ satisfies the properties:

(P1) $\widehat{\mathcal{L}^{\nu}_{\alpha\beta}}(\mathbf{y}, z^{\nu}, \lambda^{\nu}, \mu^{\nu}; \mathbf{y}) = \mathcal{L}^{\nu}_{\alpha\beta}(\mathbf{y}, z^{\nu}, \lambda^{\nu}, \mu^{\nu})$ for $\forall \mathbf{y} \in \mathbf{X}$.

(P2) $\widehat{\mathcal{L}^{\nu}_{\alpha\beta}}(\mathbf{x}, z^{\nu}, \lambda^{\nu}, \mu^{\nu}; \mathbf{y}) \geq \mathcal{L}^{\nu}_{\alpha\beta}(\mathbf{y}, z^{\nu}, \lambda^{\nu}, \mu^{\nu})$ for $\forall \mathbf{x}, \mathbf{y} \in \mathbf{X}$.

(P3) $\widehat{\mathcal{L}^{\nu}_{\alpha\beta}}(\bullet, z^{\nu}, \lambda^{\nu}, \mu^{\nu}; \mathbf{y})$ is strongly convex in $\mathbf{x} = (x^{\nu}, x^{-\nu})$ with constant $c_{\nu} > 0$, i.e.,

$$\left(\nabla_{\mathbf{x}} \widehat{\mathcal{L}^{\nu}_{\alpha\beta}}(\mathbf{x}_1, z^{\nu}, \lambda^{\nu}, \mu^{\nu}; \mathbf{y}) - \nabla_{\mathbf{x}} \widehat{\mathcal{L}^{\nu}_{\alpha\beta}}(\mathbf{x}_2, z^{\nu}, \lambda^{\nu}, \mu^{\nu}; \mathbf{y}) \right)^T (\mathbf{x}_1 - \mathbf{x}_2) \geq c_{\nu} \|\mathbf{x}_1 - \mathbf{x}_2\|^2, \ \forall \mathbf{x}_1, \mathbf{x}_2 \in \mathbf{X}.$$

(P4) $\nabla_{\mathbf{x}} \widehat{\mathcal{L}^{\nu}_{\alpha\beta}} = \left(\nabla_{x^1} \widehat{\mathcal{L}^{\nu}_{\alpha\beta}}, \ldots, \nabla_{x^N} \widehat{\mathcal{L}^{\nu}_{\alpha\beta}} \right)$ is Lipschitz continuous on $\mathbf{X}$ with constant $\widehat{L}_{\nu} \geq \gamma_{\nu}$, i.e.,

$$\left\| \nabla_{\mathbf{x}} \widehat{\mathcal{L}^{\nu}_{\alpha\beta}}(\mathbf{x}_1, z^{\nu}, \lambda^{\nu}, \mu^{\nu}; \mathbf{y}) - \nabla_{\mathbf{x}} \widehat{\mathcal{L}^{\nu}_{\alpha\beta}}(\mathbf{x}_2, z^{\nu}, \lambda^{\nu}, \mu^{\nu}; \mathbf{y}) \right\| \leq \widehat{L}_{\nu} \|\mathbf{x}_1 - \mathbf{x}_2\|, \ \forall \mathbf{x}_1, \mathbf{x}_2 \in \mathbf{X}.$$

The properties (P1) and (P2) imply that $\widehat{\mathcal{L}^{\nu}_{\alpha\beta}}$ with $\gamma_{\nu} \geq L_{\nu}$ is a tight upper bound of $\mathcal{L}^{\nu}_{\alpha\beta}$ around the given point $\mathbf{y}$. The properties (P3) and (P4) are from the structure of $\widehat{\mathcal{L}^{\nu}_{\alpha\beta}}$ that is the first-order approximation of $\mathcal{L}^{\nu}_{\alpha\beta}$ in $\mathbf{x}$ at $\mathbf{y}$ with quadratic term $\frac{\gamma_{\nu}}{2} \|\mathbf{x} - \mathbf{y}\|^2$.

Noting that for $\mathbf{y}=\mathbf{x}^k$ and $(z^{\nu,k},\lambda^{\nu,k},\mu^{\nu,k})$ and for $\gamma_\nu \geq L_\nu$, $\widehat{\mathcal{L}^\nu_{\alpha\beta}}(\bullet,z^{\nu,k},\lambda^{\nu,k},\mu^{\nu,k};\mathbf{x}^k)$ is strongly convex on $\mathbf{X}$, there must exist a unique minimizer $\widehat{\mathbf{x}}^k=(\widehat{x}^{\nu,k},\widehat{x}^{-\nu,k})$ at each iteration k such that

$$\nabla_{x^\nu}\widehat{\mathcal{L}^\nu_{\alpha\beta}}\left(\widehat{\mathbf{x}}^k,z^{\nu,k},\lambda^{\nu,k},\mu^{\nu,k};\mathbf{x}^k\right)^T\left(x^\nu-\widehat{x}^{\nu,k}\right)\geq 0,\quad \nu=1,\ldots,N.$$

It also follows from (P1) that $\widehat{\mathcal{L}^\nu_{\alpha\beta}}\left(\widehat{\mathbf{x}}^k,z^{\nu,k},\lambda^{\nu,k},\mu^{\nu,k};\mathbf{x}^k\right)\leq \mathcal{L}^\nu_{\alpha\beta}\left(\mathbf{x}^k,z^{\nu,k},\lambda^{\nu,k},\mu^{\nu,k}\right)$. We can construct a (strongly) monotone approximation mapping $\widehat{\mathbf{L}}^k:\mathbf{X}\to\mathbb{R}^n$ given by

$$\widehat{\mathbf{L}}^k\left(\mathbf{x},z^k,\lambda^k,\mu^k;\mathbf{x}^k\right):=\begin{bmatrix}\nabla_{x^1}\widehat{\mathcal{L}^1_{\alpha\beta}}\left(x^1,x^{-1},z^{1,k},\lambda^{1,k},\mu^{1,k};\mathbf{x}^k\right)\\ \vdots \\ \nabla_{x^N}\widehat{\mathcal{L}^N_{\alpha\beta}}\left(x^N,x^{-N},z^{N,k},\lambda^{N,k},\mu^{N,k};\mathbf{x}^k\right)\end{bmatrix}.$$

Let us now consider solving the approximate variational inequality $\mathrm{VI}^k(\mathbf{X},\widehat{\mathbf{L}}^k)$ of finding $\widehat{\mathbf{x}}^k$:

$$\mathrm{VI}^k(\mathbf{X},\widehat{\mathbf{L}}^k):\quad \widehat{\mathbf{L}}^k\left(\widehat{\mathbf{x}}^k,z^k,\lambda^k,\mu^k;\mathbf{x}^k\right)^T\left(\mathbf{x}-\widehat{\mathbf{x}}^k\right)\geq 0,\quad \forall \mathbf{x}\in\mathbf{X}. \tag{18}$$

It is well known (Facchinei and Pang 2007, Proposition 1.5.8) that $\widehat{\mathbf{x}}^k$ is also a solution to the system of fixed-point subproblem (or system of nonlinear projected equations) at iteration k:

$$\widehat{\mathbf{x}}^k-\mathcal{P}_{\mathbf{X}}\left[\widehat{\mathbf{x}}^k-\sigma\widehat{\mathbf{L}}^k\left(\widehat{\mathbf{x}}^k,z^k,\lambda^k,\mu^k;\mathbf{x}^k\right)\right]=0, \tag{19}$$

where $\mathcal{P}_{\mathbf{X}}(x)=\operatorname{argmin}\{\|x-y\|:y\in\mathbf{X}\}$ denotes the projection operator onto $\mathbf{X}$ and $\sigma>0$ is a constant. The constant $\sigma>0$ is defined as $\sigma=\max_{\nu=1,\ldots,N}\sigma_\nu$ such that $0<\sigma<(2\gamma^2_{\min})/\widehat{L}_{\max}$, where $\gamma_{\min}=\min_{\nu=1,\ldots,N}\gamma_\nu$ and $\widehat{L}_{\max}=\max_{\nu=1,\ldots,N}\widehat{L}_\nu$. The choice of σ will be further discussed in the context of the convergence condition in Lemma 2. For fixed $(\mathbf{x}^k,z^k,\lambda^k,\mu^k)$ at iteration k, we use the following gradient projection to generate a sequence $\left\{\mathbf{u}^{k,l}\right\}$ in inner iterations $l=0,1,2,\ldots$

$$\mathbf{u}^{k,l+1}=\mathcal{P}_{\mathbf{X}}\left[\mathbf{u}^{k,l}-\sigma\widehat{\mathbf{L}}^k\left(\mathbf{u}^{k,l},z^k,\lambda^k,\mu^k;\mathbf{x}^k\right)\right], \tag{20}$$

equivalently,

$$\mathbf{u}^{k,l+1}=\begin{pmatrix}u^{1,k,l+1}\\ \vdots\\ u^{\nu,k,l+1}\\ \vdots\\ u^{N,k,l+1}\end{pmatrix}=\begin{pmatrix}\mathcal{P}_{\mathcal{X}_1}\left[u^{1,k,l}-\sigma\left(\nabla_{x^1}\mathcal{L}^1_{\alpha\beta}\left(\mathbf{x}^k,z^{1,k},\lambda^{1,k},\mu^{1,k}\right)+\gamma_1\left(u^{1,k,l}-x^{1,k}\right)\right)\right]\\ \vdots\\ \mathcal{P}_{\mathcal{X}_\nu}\left[u^{\nu,k,l}-\sigma\left(\nabla_{x^\nu}\mathcal{L}^\nu_{\alpha\beta}\left(\mathbf{x}^k,z^{\nu,k},\lambda^{\nu,k},\mu^{\nu,k}\right)+\gamma_\nu\left(u^{\nu,k,l}-x^{\nu,k}\right)\right)\right]\\ \vdots\\ \mathcal{P}_{\mathcal{X}_N}\left[u^{N,k,l}-\sigma\left(\nabla_{x^N}\mathcal{L}^N_{\alpha\beta}\left(\mathbf{x}^k,z^{N,k},\lambda^{N,k},\mu^{N,k}\right)+\gamma_N\left(u^{N,k,l}-x^{N,k}\right)\right)\right]\end{pmatrix}, \tag{21}$$

Notice that the structure of $\nabla_{x^\nu}\widehat{\mathcal{L}^\nu_{\alpha\beta}}$ allows for the inner gradient projection (21) to be implemented in a distributed way since each ν can update its own $u^{\nu,k,l}$ while keeping $\mathbf{x}^k=(x^{\nu,k},x^{-\nu,k})$ fixed. Thus, Algorithm 1 allows each ν to choose its own step size σ_ν, $\nu=1,\ldots,N$. We also note that when the private strategy set of each player ν includes functional constraints $c^\nu_j(x^\nu)\leq 0,\ j=$

$1,\ldots,p_\nu$, they are treated in the same way to handle $g^\nu(x^\nu,x^{-\nu})\le 0$. It follows that the set $\mathcal{X}_\nu$ remains as a simple constraint, and thus the projection onto $\mathcal{X}_\nu$ is computationally cheap.

The following Lemma shows that the inner gradient projection scheme (21) converges to the solution $\widehat{\mathbf{x}}^k$ of the subproblem (18) at each iteration k and thus enables us to compute a point satisfying the desired decrease property for every $\mathcal{L}^\nu_{\alpha\beta}$ during inner iterations.

LEMMA 2. *Let $\widehat{\mathbf{x}}^k$ be the unique solution to* $\mathrm{VI}^k(\mathbf{X},\widehat{\mathbf{L}}^k)$ (18) *and $\mathbf{x}^k\neq\widehat{\mathbf{x}}^k$. Let $\{\mathbf{u}^{k,l}\}_{l\ge1}$ be the sequence generated by gradient projection* (21) *with the step size σ_ν. Suppose that for $\nu=1,\ldots,N$, the parameter γ_ν is chosen such that $\gamma_\nu\ge L_\nu$ where L_ν is the Lipschitz constant of $\nabla_{\mathbf{x}}\mathcal{L}^\nu_{\alpha\beta}$. Then,*

(a) for $\widehat{\sigma}:=\max_{\nu=1,\ldots,N}\sigma_\nu$ satisfying $0<\widehat{\sigma}<(2\gamma_{\min}^2)/\widehat{L}_{\max}$, where $\gamma_{\min}=\min_{\nu=1,\ldots,N}\gamma_\nu$, $\widehat{L}_{\max}=\max_{\nu=1,\ldots,N}\widehat{L}_\nu$, and $\widehat{L}_\nu$ is the Lipschitz constant of $\nabla_{\mathbf{x}}\widehat{\mathcal{L}^\nu_{\alpha\beta}}$, the sequence $\{\mathbf{u}^{k,l}\}_{l\ge1}$ converges to $\widehat{\mathbf{x}}^k$. That is,

$$\left\|\mathbf{u}^{k,l+1}-\widehat{\mathbf{x}}^k\right\|\le\boldsymbol{\tau}\left\|\mathbf{u}^{k,l}-\widehat{\mathbf{x}}^k\right\|,\quad 0<\boldsymbol{\tau}<1, \tag{22}$$

where $\boldsymbol{\tau}=\sqrt{1-2\gamma_{\min}\widehat{\sigma}+\widehat{\sigma}^2\widehat{L}_{\max}}$.

(b) thus, the inner gradient projection (21) *can compute $\mathbf{u}^{k,l+1}$ sufficiently close to $\widehat{\mathbf{x}}^k$ such that*

$$\mathcal{L}^\nu_{\alpha\beta}(\mathbf{u}^{k,l+1},z^{\nu,k},\lambda^{\nu,k},\mu^{\nu,k})<\mathcal{L}^\nu_{\alpha\beta}(\mathbf{x}^k,z^{\nu,k},\lambda^{\nu,k},\mu^{\nu,k})$$

for every $\nu=1,\ldots,N$ in a finite number of iterations.

Proof. (a) Fix $k\ge0$ and omit the iterates $(z^{\nu,k},\lambda^{\nu,k},\mu^{\nu,k})$ for simplicity in the proof. Let $\widehat{x}^{\nu,k}$ be νth component of $\widehat{\mathbf{x}}^k$. By the fixed-point characterization of $\widehat{x}^{\nu,k}$

$$\widehat{x}^{\nu,k}=\mathcal{P}_{\mathcal{X}_\nu}\left[\widehat{x}^{\nu,k}-\sigma_\nu\nabla_{x^\nu}\widehat{\mathcal{L}^\nu_{\alpha\beta}}\left(\widehat{\mathbf{x}}^{\nu,k};\mathbf{x}^k\right)\right]$$

and the contraction property of projection operator $\mathcal{P}_{\mathcal{X}_\nu}[\bullet]$, we have that for all $\nu=1,\ldots,N$,

$$\begin{aligned}\left\|u^{\nu,k,l+1}-\widehat{x}^{\nu,k}\right\|^2&=\left\|\mathcal{P}_{\mathcal{X}_\nu}\left[u^{\nu,k,l}-\sigma_\nu\nabla_{x^\nu}\widehat{\mathcal{L}^\nu_{\alpha\beta}}\left(\mathbf{u}^{k,l};\mathbf{x}^k\right)\right]-\mathcal{P}_{\mathcal{X}_\nu}\left[\widehat{x}^{\nu,k}-\sigma_\nu\nabla_{x^\nu}\widehat{\mathcal{L}^\nu_{\alpha\beta}}\left(\widehat{\mathbf{x}}^k;\mathbf{x}^k\right)\right]\right\|^2\\&\le\left\|\left[u^{\nu,k,l}-\sigma_\nu\nabla_{x^\nu}\widehat{\mathcal{L}^\nu_{\alpha\beta}}\left(\mathbf{u}^{k,l};\mathbf{x}^k\right)\right]-\left[\widehat{x}^{\nu,k}-\sigma_\nu\nabla_{x^\nu}\widehat{\mathcal{L}^\nu_{\alpha\beta}}\left(\widehat{\mathbf{x}}^k;\mathbf{x}^k\right)\right]\right\|^2.\end{aligned}\tag{23}$$

By expanding the last term on the right, the above inequality can be rewritten as

$$\begin{aligned}\left\|u^{\nu,k,l+1}-\widehat{x}^{\nu,k}\right\|^2\le\left\|u^{\nu,k,l}-\widehat{x}^{\nu,k}\right\|^2&-2\sigma_\nu\left(\nabla_{x^\nu}\widehat{\mathcal{L}^\nu_{\alpha\beta}}(\mathbf{u}^{k,l};\mathbf{x}^k)-\nabla_{x^\nu}\widehat{\mathcal{L}^\nu_{\alpha\beta}}(\widehat{\mathbf{x}}^k;\mathbf{x}^k)\right)^T\left(u^{\nu,k,l}-\widehat{x}^{\nu,k}\right)\\&+\sigma_\nu^2\left\|\nabla_{x^\nu}\widehat{\mathcal{L}^\nu_{\alpha\beta}}(\mathbf{u}^{k,l};\mathbf{x}^k)-\nabla_{x^\nu}\widehat{\mathcal{L}^\nu_{\alpha\beta}}(\widehat{\mathbf{x}}^k;\mathbf{x}^k)\right\|^2.\end{aligned}\tag{24}$$

Since $\widehat{\mathcal{L}^\nu_{\alpha\beta}}$ is strongly convex in $\mathbf{x}$ with constant c_ν and $\nabla_{\mathbf{x}}\widehat{\mathcal{L}^\nu_{\alpha\beta}}$ is Lipschitz continuous with constant $\widehat{L}_\nu$ ((P3) and (P4) in Remark 1), we can estimate the second and third terms on the RHS of (24):

$$\left(\nabla_{x^\nu}\widehat{\mathcal{L}^\nu_{\alpha\beta}}\left(\mathbf{u}^{k,l};\mathbf{x}^k\right)-\nabla_{x^\nu}\widehat{\mathcal{L}^\nu_{\alpha\beta}}\left(\widehat{\mathbf{x}}^k;\mathbf{x}^k\right)\right)^T\left(u^{\nu,k,l}-\widehat{x}^{\nu,k}\right)\ge c_\nu\left\|u^{\nu,k,l}-\widehat{x}^{\nu,k}\right\|^2, \tag{25}$$

$$\left\|\nabla_{x^\nu}\widehat{\mathcal{L}^\nu_{\alpha\beta}}\left(\mathbf{u}^{k,l};\mathbf{x}^k\right)-\nabla_{x^\nu}\widehat{\mathcal{L}^\nu_{\alpha\beta}}\left(\widehat{\mathbf{x}}^k;\mathbf{x}^k\right)\right\|\leq\widehat{L}_\nu\left\|u^{\nu,k,l}-\widehat{x}^{\nu,k}\right\|. \tag{26}$$

Note that since $\widehat{\mathcal{L}^\nu_{\alpha\beta}}$ is a proximal linearized function with the quadratic term $\frac{\gamma_\nu}{2}\left\|\mathbf{x}-\mathbf{x}^k\right\|^2$, we can take $\gamma_\nu=c_\nu$. Substituting (25) with $\gamma_\nu=c_\nu$ and (26) into (24) yields

$$\left\|u^{\nu,k,l+1}-\widehat{x}^{\nu,k}\right\|^2\leq\left(1-2\gamma_\nu\sigma_\nu+\sigma_\nu^2\widehat{L}_\nu^2\right)\left\|u^{\nu,k,l}-\widehat{x}^{\nu,k}\right\|^2.$$

Notice that $\left(1-2\gamma_\nu\sigma_\nu+\sigma_\nu^2\widehat{L}_\nu^2\right)\geq 0$ is satisfied since $\widehat{L}_\nu\geq\gamma_\nu$. Now, setting $\widehat{\sigma}:=\max\limits_{\nu=1,\ldots,N}\sigma_\nu$ and observing that $\left(1-2\gamma_\nu\widehat{\sigma}+\widehat{\sigma}^2\widehat{L}_\nu^2\right)\leq\left(1-2\gamma_{\min}\widehat{\sigma}+\widehat{\sigma}^2\widehat{L}_{\max}^2\right)$, where $\gamma_{\min}=\min\limits_{\nu=1,\ldots,N}\gamma_\nu$ and $\widehat{L}_{\max}=\max\limits_{\nu=1,\ldots,N}\widehat{L}_\nu$, it immediately follows that

$$\left\|u^{\nu,k,l+1}-\widehat{x}^{\nu,k}\right\|^2\leq\left(1-2\gamma_{\min}\widehat{\sigma}+\widehat{\sigma}^2\widehat{L}_{\max}^2\right)\left\|u^{\nu,k,l}-\widehat{x}^{\nu,k}\right\|^2.$$

Thus, for $0<\widehat{\sigma}<\left(2\gamma_{\min}^2\right)/\widehat{L}_{\max}$ implying that $\left(1-2\gamma_{\min}\sigma+\sigma^2\widehat{L}_{\max}^2\right)<1$, we obtain

$$\left\|u^{\nu,k,l+1}-\widehat{x}^{\nu,k}\right\|\leq\boldsymbol{\tau}\left\|u^{\nu,k,l}-\widehat{x}^{\nu,k}\right\|,\quad 0<\boldsymbol{\tau}<1, \tag{27}$$

where $\boldsymbol{\tau}=\sqrt{1-2\gamma_{\min}\sigma+\sigma^2\widehat{L}_{\max}^2}$. Therefore, by summing over the above inequality for all players from $\nu=1$ to N, we deduce the desired result (22).

(b) From the property (P1) in Remark 1 with $\mathbf{y}=\mathbf{x}^k$, we know that

$$\widehat{\mathcal{L}^\nu_{\alpha\beta}}\left(\mathbf{x}^k,z^{\nu,k},\lambda^{\nu,k},\mu^{\nu,k};\mathbf{x}^k\right)=\mathcal{L}^\nu_{\alpha\beta}\left(\mathbf{x}^k,z^{\nu,k},\lambda^{\nu,k},\mu^{\nu,k}\right),\quad\nu=1,\ldots,N. \tag{28}$$

Since $\mathbf{x}^k\neq\widehat{\mathbf{x}}^k$ and $\mathbf{u}^{k,l}\to\widehat{\mathbf{x}}^k$ by the result (a), the inner gradient projection (21) can find a point $\mathbf{u}^{k,l+1}$ close to $\widehat{\mathbf{x}}^k$ such that

$$\widehat{\mathcal{L}^\nu_{\alpha\beta}}\left(\mathbf{u}^{k,l+1},z^{\nu,k},\lambda^{\nu,k},\mu^{\nu,k};\mathbf{x}^k\right)<\widehat{\mathcal{L}^\nu_{\alpha\beta}}\left(\mathbf{x}^k,z^{\nu,k},\lambda^{\nu,k},\mu^{\nu,k};\mathbf{x}^k\right),\quad\nu=1,\ldots,N, \tag{29}$$

in a finite number of iterations. By (P2) in Remark 1 with $\mathbf{y}=\mathbf{u}^{k,l+1}$, we have that for any $\gamma_\nu\geq L_\nu$,

$$\mathcal{L}^\nu_{\alpha\beta}\left(\mathbf{u}^{k,l+1},z^{\nu,k},\lambda^{\nu,k},\mu^{\nu,k}\right)\leq\widehat{\mathcal{L}^\nu_{\alpha\beta}}\left(\mathbf{u}^{k,l+1},z^{\nu,k},\lambda^{\nu,k},\mu^{\nu,k};\mathbf{x}^k\right),\quad\nu=1,\ldots,N. \tag{30}$$

Combining (28), (29), and (30) yields

$$\mathcal{L}^\nu_{\alpha\beta}\left(\mathbf{u}^{k,l+1},z^{\nu,k},\lambda^{\nu,k},\mu^{\nu,k}\right)<\mathcal{L}^\nu_{\alpha\beta}\left(\mathbf{x}^k,z^{\nu,k},\lambda^{\nu,k},\mu^{\nu,k}\right),\quad\nu=1,\ldots,N.$$

Hence, we can derive the desired decrease property of every $\mathcal{L}^\nu_{\alpha\beta}$ during inner iterations. $\square$

3.3. Description of Algorithm

We are ready to formally present our distributed algorithm that exploits all the features discussed. The steps of the proposed algorithm are summarized in Algorithm 1, where the choice of the proximal parameter $\gamma_\nu \geq L_\nu + \frac{3L_{g^\nu}^2}{\beta_\nu}$ for updating x will be discussed in detail later (Lemma 4).

Algorithm 1: P-Lagrangian based Alternating Direction Algorithm (PL-ADA)

Input parameters: $\sigma_\nu > 0$, $\alpha_\nu > 0$, $\beta_\nu > 0$, $\sigma \in \left(0,\ 2\gamma_{\min}^2/\widehat{L}_{\max}\right)$, and $\gamma_\nu \geq L_\nu + \frac{3L_{g^\nu}^2}{\beta_\nu}$.

Initialization: Set $k=0$, and define $(x^{\nu,0}, z^{\nu,0}, \lambda^{\nu,0}, \mu^{\nu,0})$ with $\lambda^{\nu,0} = \mu^{\nu,0}$, $\nu = 1, \ldots, N$.

Step 1. Let iteration k be fixed and set $\mathbf{u}^{k,0} = \mathbf{x}^k$.
For every $\nu = 1, \ldots, N$, and for fixed $(\mathbf{x}^k, z^{\nu,k}, \lambda^{\nu,k}, \mu^{\nu,k})$, compute $u^{\nu,k,l+1}$ according to the following gradient projection scheme on $\widehat{\mathcal{L}_{\alpha\beta}^\nu}$ for inner iterations $l = 0, 1, 2, \ldots$

while $\left\|\mathcal{P}_{\mathbf{X}}\left[\mathbf{u}^{k,l+1} - \sigma\widehat{\mathbf{L}}^k(\mathbf{u}^{k,l+1}, z^k, \lambda^k, \mu^k; \mathbf{x}^k)\right] - \mathbf{u}^{k,l+1}\right\| > \varepsilon$ **or**
$\widehat{\mathcal{L}_{\alpha\beta}^\nu}(\mathbf{u}^{k,l+1}, z^{\nu,k}, \lambda^{\nu,k}, \mu^{\nu,k}; \mathbf{x}^k) - \mathcal{L}_{\alpha\beta}^\nu(\mathbf{x}^k, z^{\nu,k}, \lambda^{\nu,k}, \mu^{\nu,k}) \geq 0$ **do**

$$\begin{aligned} u^{\nu,k,l+1} &= \mathcal{P}_{\mathcal{X}_\nu}\left[u^{\nu,k,l} - \sigma_\nu \nabla_{x^\nu}\widehat{\mathcal{L}_{\alpha\beta}^\nu}\left(u^{\nu,k,l}, z^{\nu,k}, \lambda^{\nu,k}, \mu^{\nu,k}; \mathbf{x}^k\right)\right] \\ &= \mathcal{P}_{\mathcal{X}_\nu}\left[u^{\nu,k,l} - \sigma_\nu\left(\nabla_{x^\nu}\mathcal{L}_{\alpha\beta}^\nu\left(\mathbf{x}^k, z^{\nu,k}, \lambda^{\nu,k}, \mu^{\nu,k}\right) + \gamma_\nu\left(u^{\nu,k,l} - x^{\nu,k}\right)\right)\right]. \end{aligned}$$

end while
Set $\mathbf{x}^{k+1} = \mathbf{u}^{k,l+1} := \left[(u^{1,k,l+1})^T, \ldots, (u^{N,k,l+1})^T\right]^T$.

Step 2. For $\nu = 1, \ldots, N$, compute $z^{\nu,k+1}$ by an exact minimization step on $\mathcal{L}_{\alpha\beta}^\nu$

$$z^{\nu,k+1} = \operatorname*{arg\,min}_{z^\nu \in \mathbb{R}^{m_\nu}}\left\{\mathcal{L}_{\alpha\beta}^\nu\left(\mathbf{x}^{k+1}, z^\nu, \lambda^{\nu,k}, \mu^{\nu,k}\right)\right\} = \left(\lambda^{\nu,k} - \mu^{\nu,k}\right)/\alpha_\nu.$$

Step 3. For $\nu = 1, \ldots, N$, update $(\lambda^{\nu,k+1}, \mu^{\nu,k+1})$ by exact maximization steps on $\mathcal{L}_{\alpha\beta}^\nu$

$$\begin{aligned} \lambda^{\nu,k+1} &= \operatorname*{arg\,max}_{\lambda^\nu \in \mathbb{R}_+^{m_\nu}}\left\{\mathcal{L}_{\alpha\beta}^\nu\left(\mathbf{x}^{k+1}, z^{\nu,k+1}, \lambda^\nu, \mu^{\nu,k}\right)\right\} = \left[\mu^{\nu,k} + \frac{1}{\beta_\nu} g^\nu(\mathbf{x}^{k+1})\right]^+. \\ \mu^{\nu,k+1} &= \operatorname*{arg\,max}_{\mu^\nu \in \mathbb{R}^{m_\nu}}\left\{\mathcal{L}_{\alpha\beta}^\nu\left(\mathbf{x}^{k+1}, z^{\nu,k+1}, \lambda^{\nu,k+1}, \mu^\nu\right)\right\} = \lambda^{\nu,k+1}. \end{aligned}$$

Step 4. Set $k \leftarrow k+1$ and go to Step 1.

The main computational effort of Algorithm 1 is involved in Step 1 to update primal iterates from $\mathbf{x}^k$ to $\mathbf{x}^{k+1}$. If $\widehat{\mathbf{x}}^k \neq \mathbf{x}^k$, by Lemma 2, we can find a point $\mathbf{u}^{k,l+1}$ satisfying both conditions:

$$\left\|\mathcal{P}_{\mathbf{X}}\left[\mathbf{u}^{k,l+1} - \sigma\widehat{\mathbf{L}}^k\left(\mathbf{u}^{k,l+1}, z^k, \lambda^k, \mu^k; \mathbf{x}^k\right)\right] - \mathbf{u}^{k,l+1}\right\| \leq \varepsilon \tag{31}$$

$$\widehat{\mathcal{L}_{\alpha\beta}^\nu}\left(\mathbf{u}^{k,l+1}, z^{\nu,k}, \lambda^{\nu,k}, \mu^{\nu,k}; \mathbf{x}^k\right) < \mathcal{L}_{\alpha\beta}^\nu\left(\mathbf{x}^k, z^{\nu,k}, \lambda^{\nu,k}, \mu^{\nu,k}\right), \quad \nu = 1, \ldots, N, \tag{32}$$

in a finite number of inner iterations. When the descent condition (32) is satisfied, $\mathbf{u}^{k,l+1}$ is set to $\mathbf{x}^{k+1}$. Consequently, the decrease of $\mathcal{L}_{\alpha\beta}^\nu(\mathbf{x}^k, z^{\nu,k}, \lambda^{\nu,k}, \mu^{\nu,k})$ value is obtained, that is,

$$\mathcal{L}_{\alpha\beta}^\nu\left(\mathbf{x}^{k+1}, z^{\nu,k}, \lambda^{\nu,k}, \mu^{\nu,k}\right) \leq \widehat{\mathcal{L}_{\alpha\beta}^\nu}\left(\mathbf{x}^{k+1}, z^{\nu,k}, \lambda^{\nu,k}, \mu^{\nu,k}; \mathbf{x}^k\right) < \mathcal{L}_{\alpha\beta}^\nu\left(\mathbf{x}^k, z^{\nu,k}, \lambda^{\nu,k}, \mu^{\nu,k}\right).$$

for $\nu = 1, \ldots, N$ (see Lemma 2 (b)).

The next step is to update z^ν by taking exact minimization step (Step 2) on $\mathcal{L}_{\alpha\beta}^\nu$.

After the minimization steps have been carried out, given $(\mathbf{x}^{k+1}, z^{\nu,k+1})$, the multipliers are updated by exact maximization steps on $\mathcal{L}^{\nu}_{\alpha\beta}$. The updates of λ^ν and μ^ν take the explicit forms:

$$\lambda^{\nu,k+1} = \left[\mu^{\nu,k} + \frac{1}{\beta_\nu} g^\nu(\mathbf{x}^{k+1})\right]^+, \qquad \mu^{\nu,k+1} = \lambda^{\nu,k+1},$$

We remark that a point satisfying the fixed-point condition (31) does not necessarily guarantee that the descent condition (32) holds. Hence, the algorithm keeps updating iterates $\mathbf{u}^{k,l}$ until the condition (32) is satisfied even after condition (31) is met, which may require many inner iterations.

4. Convergence Analysis

In this section, we establish the convergence results of Algorithm 1. We prove that the sequence generated by Algorithm 1 converges to a saddle point of $\mathcal{L}^{\nu}_{\alpha\beta}(x^\nu, x^{-\nu}, z^\nu, \lambda^\nu, \mu^\nu)$ for $\nu = 1, \ldots, N$. In particular, our analysis proceeds with the steps:

1. We show that $\left\|\lambda^{\nu,k+1} - \lambda^{\nu,k}\right\|$ can be bounded by $\left\|\mathbf{x}^{k+1} - \mathbf{x}^k\right\|$ (Lemma 3), which is exploited to show $\{\mathcal{L}^{\nu}_{\alpha\beta}\}$ is nonincreasing and convergent (Lemma 4). Then we establish key results; boundedness of $\left\{\mathbf{x}^k\right\}$, $\lim_{k\to\infty}\left\|\mathbf{x}^{k+1} - \mathbf{x}^k\right\| = 0$, and then boundedness of $\left\{\lambda^{\nu,k}\right\}$ (Theorem 3).
2. With the bounded sequences, convergence to an equilibrium of the GNEP is proven; we show that any limit point of the sequence is a saddle point of $\mathcal{L}^{\nu}_{\alpha\beta}$ (Theorem 4).
3. We establish the global convergence; the generated *whole* sequence converges to the saddle point under the assumption of *Kurdyka-Łojasiewicz* (KŁ) inequality (Theorem 5).

4.1. Key Properties of Algorithm 1

We first derive an important relation on the dual iterates $\lambda^{\nu,k}$ with the primal iterates $\mathbf{x}^k$; the difference of two consecutive dual iterates can be bounded by that of the primal iterates.

LEMMA 3. *Let $\left\{(x^{\nu,k}, z^{\nu,k}, \lambda^{\nu,k}, \mu^{\nu,k})\right\}_{\nu=1}^N$ be the sequence generated by Algorithm 1. Then,*

$$\left\|\lambda^{\nu,k+1} - \lambda^{\nu,k}\right\|^2 \le \frac{L^2_{g^\nu}}{\beta^2_\nu}\left\|\mathbf{x}^{k+1} - \mathbf{x}^k\right\|^2, \tag{33}$$

where L_{g^ν} is the Lipschitz constant of g^ν and $\beta_\nu > 0$ is the parameter of $-\frac{\beta_\nu}{2}\left\|\lambda^\nu - \mu^\nu\right\|^2$ in $\mathcal{L}^{\nu}_{\alpha\beta}$.

Proof. Note that since $\mathcal{L}^{\nu}_{\alpha\beta}$ is strongly concave in λ^ν for fixed $(\mathbf{x}, z^\nu, \mu^\nu)$, there exists a unique maximizer, denoted by $\widehat{\lambda}^\nu(\mathbf{x}, z^\nu, \mu^\nu)$, such that $\mathcal{L}^{\nu}_{\alpha\beta}(\mathbf{x}, z^\nu, \widehat{\lambda}^\nu(\mathbf{x}, z^\nu, \mu^\nu), \mu^\nu) = \max_{\lambda^\nu \in \mathbb{R}^{m_\nu}_+} \mathcal{L}^{\nu}_{\alpha\beta}(\mathbf{x}, z^\nu, \lambda^\nu, \mu^\nu)$. From the update of $\lambda^{\nu,k}$ defined (as maximizer) in Step 3, we have

$$\nabla_{\lambda^\nu}\mathcal{L}^{\nu}_{\alpha\beta}\left(\mathbf{x}^{k+1}, z^{\nu,k+1}, \lambda^{\nu,k+1}, \mu^{\nu,k}\right)^T\left(\lambda^{\nu,k} - \lambda^{\nu,k+1}\right) \le 0,$$
$$\nabla_{\lambda^\nu}\mathcal{L}^{\nu}_{\alpha\beta}\left(\mathbf{x}^{k}, z^{\nu,k}, \lambda^{\nu,k}, \mu^{\nu,k-1}\right)^T\left(\lambda^{\nu,k+1} - \lambda^{\nu,k}\right) \le 0.$$

By the definition $z^{\nu,k+1} = \frac{\lambda^{\nu,k}-\mu^{\nu,k}}{\alpha_\nu}$ in Step 2 with $\lambda^{\nu,k+1} = \mu^{\nu,k+1}$ in Step 3, we have $z^{\nu,k+1} = z^{\nu,k} = 0$ for $\lambda^{\nu,0} = \mu^{\nu,0}$. Adding the above inequalities and a direct computation of $\nabla_{\lambda^\nu}\mathcal{L}^\nu_{\alpha\beta}$ give

$$\begin{aligned}
&\left(\nabla_{\lambda^\nu}\mathcal{L}^\nu_{\alpha\beta}\left(\mathbf{x}^{k+1}, z^{\nu,k+1}, \lambda^{\nu,k+1}, \mu^{\nu,k}\right) - \nabla_{\lambda^\nu}\mathcal{L}^\nu_{\alpha\beta}\left(\mathbf{x}^{k}, z^{\nu,k}, \lambda^{\nu,k}, \mu^{\nu,k-1}\right)\right)^T\left(\lambda^{\nu,k} - \lambda^{\nu,k+1}\right) \\
&= \left(g^\nu\left(\mathbf{x}^{k+1}\right) - g^\nu\left(\mathbf{x}^{k}\right) - \beta_\nu\left(\lambda^{\nu,k+1} - \lambda^{\nu,k}\right) + \beta_\nu\left(\mu^{\nu,k} - \mu^{\nu,k-1}\right)\right)^T\left(\lambda^{\nu,k} - \lambda^{\nu,k+1}\right) \\
&= \left(g^\nu\left(\mathbf{x}^{k+1}\right) - g^\nu\left(\mathbf{x}^{k}\right)\right)^T\left(\lambda^{\nu,k} - \lambda^{\nu,k+1}\right) + \beta_\nu\left\|\lambda^{\nu,k+1} - \lambda^{\nu,k}\right\|^2 + \beta_\nu\underbrace{\left(\mu^{\nu,k} - \mu^{\nu,k-1}\right)^T\left(\lambda^{\nu,k} - \lambda^{\nu,k+1}\right)}_{\geq\left\|\lambda^{\nu,k+1}-\lambda^{\nu,k}\right\|^2} \leq 0,
\end{aligned} \tag{34}$$

where, to bound the third term, we used Lemma 1(a) in Nedic et al. (2010); $(x-y)^T(x-P[x]) \geq \|P[x]-x\|^2$ with $x = \mu^{\nu,k}$, $y = \mu^{\nu,k-1}$, $P[x] = \lambda^{\nu,k+1}$, and the fact $\mu^{\nu,k} = \lambda^{\nu,k}$. Specifically, since $\lambda^{\nu,k+1}$ maximizes $\mathcal{L}^\nu_{\alpha\beta}(\mathbf{x}^{k+1}, z^{\nu,k+1}, \lambda^\nu, \mu^{\nu,k}) = \theta_\nu(\mathbf{x}^{k+1}) + (\lambda^\nu)^T g^\nu(\mathbf{x}^{k+1}) - \frac{\beta_\nu}{2}\left\|\lambda^\nu - \mu^{\nu,k}\right\|^2$, we have

$$\begin{aligned}
&\theta_\nu(\mathbf{x}^{k+1}) + \left(\lambda^{\nu,k+1}\right)^T g^\nu(\mathbf{x}^{k+1}) - \frac{\beta_\nu}{2}\left\|\lambda^{\nu,k+1} - \mu^{\nu,k}\right\|^2 \\
&\qquad \geq \theta_\nu(\mathbf{x}^{k+1}) + \left(\widehat{\lambda}^\nu(\mathbf{x}^{k+1}, z^{k+1}, \mu^{\nu,k})\right)^T g^\nu(\mathbf{x}^{k+1}) - \frac{\beta_\nu}{2}\left\|\widehat{\lambda}^\nu(\mathbf{x}^{k+1}, z^{k+1}, \mu^{\nu,k}) - \mu^{\nu,k}\right\|^2,
\end{aligned}$$

and $\left(\lambda^{\nu,k+1}\right)^T g^\nu(\mathbf{x}^{k+1}) = \widehat{\lambda}^\nu\left(\mathbf{x}^{k+1}, z^{k+1}, \mu^{\nu,k}\right)^T g^\nu(\mathbf{x}^{k+1})$. It thus follows that

$$\left\|\lambda^{\nu,k+1} - \mu^{\nu,k}\right\| \leq \left\|\widehat{\lambda}^\nu(\mathbf{x}^{k+1}, z^{k+1}, \mu^{\nu,k}) - \mu^{\nu,k}\right\|,$$

which, by definition of projection (Bertsekas and Tsitsiklis 1989, Section 3.4), means that $\lambda^{\nu,k+1}$ can be viewed as the projection of μ^k onto the solution set $\widehat{\lambda}^\nu(\mathbf{x}^{k+1}, z^{k+1}, \mu^{\nu,k})$. We thus see that

$$\left(\mu^{\nu,k} - \mu^{\nu,k-1}\right)^T\left(\lambda^{\nu,k} - \lambda^{\nu,k+1}\right) = \left(\lambda^{\nu,k} - \lambda^{\nu,k-1}\right)^T\left(\lambda^{\nu,k} - \lambda^{\nu,k+1}\right) \geq \left\|\lambda^{\nu,k+1} - \lambda^{\nu,k}\right\|^2 \geq 0.$$

By the Cauchy-Schwarz inequality, we also get that $\left\|\lambda^{\nu,k} - \lambda^{\nu,k-1}\right\| \geq \left\|\lambda^{\nu,k+1} - \lambda^{\nu,k}\right\|$, implying the stable sequence of the multipliers. Rearranging terms in (34), we obtain

$$\beta_\nu\left\|\lambda^{\nu,k+1} - \lambda^{\nu,k}\right\|^2 \leq \left(g^\nu(\mathbf{x}^{k+1}) - g^\nu(\mathbf{x}^{k})\right)^T\left(\lambda^{\nu,k+1} - \lambda^{\nu,k}\right),$$

which leads to

$$\left\|\lambda^{\nu,k+1} - \lambda^{\nu,k}\right\| \overset{(i)}{\leq} \frac{1}{\beta_\nu}\left\|g^\nu(\mathbf{x}^{k+1}) - g^\nu(\mathbf{x}^{k})\right\| \overset{(ii)}{\leq} \frac{L_{g^\nu}}{\beta_\nu}\left\|\mathbf{x}^{k+1} - \mathbf{x}^{k}\right\|,$$

where (i) follows from the Cauchy-Schwarz inequality; (ii) is from the continuous differentiability of $g^\nu(\mathbf{x})$ (Assumption 1), implying that $g^\nu(\mathbf{x})$ is locally Lipschitz continuous with constant L_{g^ν}. Squaring both sides of the inequality gives the desired result (33). □

With Lemmas 2 and 3, we prove that $\{\mathcal{L}^\nu_{\alpha\beta}\}$ can be monotonically decreasing and convergent.

LEMMA 4 **(Sufficient Decrease and Convergence of $\{\mathcal{L}^{\nu}_{\alpha\beta}\}$)**. *Suppose that Assumptions 1 and 2 hold true, and let $\left\{(x^{\nu,k}, z^{\nu,k}, \lambda^{\nu,k}, \mu^{\nu,k})\right\}_{\nu=1}^{N}$ be the sequence generated by Algorithm 1. Then for $\nu = 1, \ldots, N$, we have*

$$\mathcal{L}^{\nu}_{\alpha\beta}\left(\mathbf{x}^{k+1}, z^{\nu,k+1}, \lambda^{\nu,k+1}, \mu^{\nu,k+1}\right) \leq \mathcal{L}^{\nu}_{\alpha\beta}\left(\mathbf{x}^{k}, z^{\nu,k}, \lambda^{\nu,k}, \mu^{\nu,k}\right) - \frac{1}{2}\left(\gamma_\nu - L_\nu - \frac{3L_{g^\nu}^2}{\beta_\nu}\right)\left\|\mathbf{x}^{k+1} - \mathbf{x}^k\right\|^2,$$

where L_ν is the Lipschitz constant of $\nabla_{\mathbf{x}}\mathcal{L}^{\nu}_{\alpha\beta}$, $\gamma_\nu > 0$ is the parameter of $\frac{\gamma_\nu}{2}\left\|\mathbf{x} - \mathbf{x}^k\right\|^2$ in $\widehat{\mathcal{L}^{\nu}_{\alpha\beta}}$, and $\beta_\nu > 0$ is the parameter of $-\frac{\beta_\nu}{2}\left\|\lambda - \mu\right\|^2$ in $\mathcal{L}^{\nu}_{\alpha\beta}$. If $\gamma_\nu > 0$ is chosen such that $\gamma_\nu \geq L_\nu + \frac{3L_{g^\nu}^2}{\beta_\nu}$, then the sequence $\{\mathcal{L}^{\nu}_{\alpha\beta}\}$ is nonincreasing and convergent.

Proof. Consider the difference of two consecutive sequences of $\mathcal{L}^{\nu}_{\alpha\beta}$:

$$\begin{aligned}
&\mathcal{L}^{\nu}_{\alpha\beta}\left(\mathbf{x}^{k+1}, z^{\nu,k+1}, \lambda^{\nu,k+1}, \mu^{\nu,k+1}\right) - \mathcal{L}^{\nu}_{\alpha\beta}\left(\mathbf{x}^{k}, z^{\nu,k}, \lambda^{\nu,k}, \mu^{\nu,k}\right) \\
&= \left[\mathcal{L}^{\nu}_{\alpha\beta}\left(\mathbf{x}^{k+1}, z^{\nu,k+1}, \lambda^{\nu,k}, \mu^{\nu,k}\right) - \mathcal{L}^{\nu}_{\alpha\beta}\left(\mathbf{x}^{k}, z^{\nu,k}, \lambda^{\nu,k}, \mu^{\nu,k}\right)\right] \\
&\quad + \left[\mathcal{L}^{\nu}_{\alpha\beta}\left(\mathbf{x}^{k+1}, z^{\nu,k+1}, \lambda^{\nu,k+1}, \mu^{\nu,k+1}\right) - \mathcal{L}^{\nu}_{\alpha\beta}\left(\mathbf{x}^{k+1}, z^{\nu,k+1}, \lambda^{\nu,k}, \mu^{\nu,k}\right)\right].
\end{aligned} \tag{35}$$

For the first term in (35), recalling the descent Lemma (Lemma 1), we have

$$\mathcal{L}^{\nu}_{\alpha\beta}\left(\mathbf{x}^{k+1}\right) \leq \mathcal{L}^{\nu}_{\alpha\beta}\left(\mathbf{x}^{k}\right) + \nabla_{\mathbf{x}}\mathcal{L}^{\nu}_{\alpha\beta}\left(\mathbf{x}^{k}\right)^T\left(\mathbf{x}^{k+1} - \mathbf{x}^k\right) + \frac{L_\nu}{2}\left\|\mathbf{x}^{k+1} - \mathbf{x}^k\right\|^2. \tag{36}$$

Here, we omitted $(z^{\nu,k}, \lambda^{\nu,k}, \mu^{\nu,k})$ for simplicity. Since $\gamma_\nu \geq L_\nu$, by Lemma 2 and Step 1, we have

$$\widehat{\mathcal{L}^{\nu}_{\alpha\beta}}\left(\mathbf{x}^{k+1}; \mathbf{x}^k\right) = \mathcal{L}^{\nu}_{\alpha\beta}\left(\mathbf{x}^{k}\right) + \nabla_{\mathbf{x}}\mathcal{L}^{\nu}_{\alpha\beta}\left(\mathbf{x}^{k}\right)^T\left(\mathbf{x}^{k+1} - \mathbf{x}^k\right) + \frac{\gamma_\nu}{2}\left\|\mathbf{x}^{k+1} - \mathbf{x}^k\right\|^2 \leq \widehat{\mathcal{L}^{\nu}_{\alpha\beta}}\left(\mathbf{x}^{k}; \mathbf{x}^k\right) = \mathcal{L}^{\nu}_{\alpha\beta}\left(\mathbf{x}^{k}\right).$$

Thus,

$$\nabla_{\mathbf{x}}\mathcal{L}^{\nu}_{\alpha\beta}\left(\mathbf{x}^{k}\right)^T\left(\mathbf{x}^{k+1} - \mathbf{x}^k\right) \leq -\frac{\gamma_\nu}{2}\left\|\mathbf{x}^{k+1} - \mathbf{x}^k\right\|^2.$$

By substituting the above expression into (36) and using the definition of $z^{\nu,k+1}$ in Step 2, we get

$$\mathcal{L}^{\nu}_{\alpha\beta}\left(\mathbf{x}^{k+1}, z^{\nu,k+1}, \lambda^{\nu,k}, \mu^{\nu,k}\right) - \mathcal{L}^{\nu}_{\alpha\beta}\left(\mathbf{x}^{k}, z^{\nu,k}, \lambda^{\nu,k}, \mu^{\nu,k}\right) \leq -\frac{1}{2}\left(\gamma_\nu - L_\nu\right)\left\|\mathbf{x}^{k+1} - \mathbf{x}^k\right\|^2. \tag{37}$$

Next, consider the second term i (35):

$$\begin{aligned}
&\mathcal{L}^{\nu}_{\alpha\beta}\left(\mathbf{x}^{k+1}, z^{\nu,k+1}, \lambda^{\nu,k+1}, \mu^{\nu,k+1}\right) - \mathcal{L}^{\nu}_{\alpha\beta}\left(\mathbf{x}^{k+1}, z^{\nu,k+1}, \lambda^{\nu,k}, \mu^{\nu,k}\right) \\
&= \left(\lambda^{\nu,k+1} - \lambda^{\nu,k}\right)^T g^\nu\left(\mathbf{x}^{k+1}\right) - \frac{\beta_\nu}{2}\underbrace{\left\|\lambda^{\nu,k+1} - \mu^{\nu,k+1}\right\|^2}_{=0} + \frac{\beta_\nu}{2}\underbrace{\left\|\lambda^{\nu,k} - \mu^{\nu,k}\right\|^2}_{=0},
\end{aligned} \tag{38}$$

where it follows from Step 3 that the second and third terms on the right-hand side are zero.

We now focus on deriving an upper bound for the term $\left(\lambda^{\nu,k+1} - \lambda^{\nu,k}\right)^T g^\nu\left(\mathbf{x}^{k+1}\right)$. To this end, we need to consider two cases: $\mu^{\nu,k} + \frac{1}{\beta_\nu}g^\nu(\mathbf{x}^{k+1}) \geq 0$ and $\mu^{\nu,k} + \frac{1}{\beta_\nu}g^\nu(\mathbf{x}^{k+1}) < 0$.

Case 1. $\mu^{\nu,k} + \frac{1}{\beta_\nu}g^\nu(\mathbf{x}^{k+1}) \geq 0$. Since $\lambda^{\nu,k+1} = \left[\mu^{\nu,k} + \frac{1}{\beta_\nu}g^\nu(\mathbf{x}^{k+1})\right]^+$ and $\lambda^{\nu,k} = \mu^{\nu,k}$, we obtain

$$\left(\lambda^{\nu,k+1} - \lambda^{\nu,k}\right)^T g^\nu(\mathbf{x}^{k+1}) = \beta_\nu\left\|\lambda^{\nu,k+1} - \lambda^{\nu,k}\right\|^2. \tag{39}$$

Case 2. $\mu^{\nu,k} + \frac{1}{\beta_\nu} g^\nu(\mathbf{x}^{k+1}) < 0$. In this case, $\lambda^{\nu,k+1} = 0$ and $\mathbf{x}^{k+1}$ is feasible because $g(\mathbf{x}^{k+1}) < 0$. For convenience, we define

$$\triangle_k := (\lambda^{\nu,k+1})^T g^\nu(\mathbf{x}^{k+1}) - (\lambda^{\nu,k})^T g^\nu(\mathbf{x}^k).$$

By subtracting and adding $\left(\lambda^{\nu,k}\right)^T g^\nu(\mathbf{x}^{k+1})$ to the RHS and using the fact that $\lambda^{\nu,k+1} = 0$, we have

$$\begin{aligned} \|\triangle_k\| &= \left\| \left(\lambda^{\nu,k+1} - \lambda^{\nu,k}\right)^T g^\nu(\mathbf{x}^{k+1}) + (\lambda^{\nu,k})^T \left(g^\nu(\mathbf{x}^{k+1}) - g^\nu(\mathbf{x}^k)\right) \right\| \\ &\geq \left\| \left(\lambda^{\nu,k+1} - \lambda^{\nu,k}\right)^T g^\nu(\mathbf{x}^{k+1}) \right\| - \left\| \left(\lambda^{\nu,k} - \lambda^{\nu,k+1}\right)^T \left(g^\nu(\mathbf{x}^{k+1}) - g^\nu(\mathbf{x}^k)\right) \right\|. \end{aligned} \tag{40}$$

From the feasibility of $\mathbf{x}^{k+1}$, we have that for any $\lambda^\nu \in \mathbb{R}_+^{m_\nu}$, $\mu^\nu \in \mathbb{R}^{m_\nu}$ and $\beta_\nu > 0$

$$(\lambda^\nu)^T g^\nu(\mathbf{x}^{k+1}) - \frac{\beta_\nu}{2} \|\lambda^\nu - \mu^\nu\|^2 \leq 0.$$

Thus, we can get with $\lambda^\nu = \lambda^{\nu,k+1}$ and $\mu^\nu = \lambda^{\nu,k}$ that

$$\left(\lambda^{\nu,k+1}\right)^T g^\nu(\mathbf{x}^{k+1}) \leq \frac{\beta_\nu}{2} \left\|\lambda^{\nu,k+1} - \lambda^{\nu,k}\right\|^2.$$

On the other hand, since $\lambda^{\nu,k} \geq 0$ maximizes $\mathcal{L}^\nu_{\alpha\beta}(\mathbf{x}^k, z^{\nu,k}, \lambda^\nu, \mu^{\nu,k-1}) = \theta_\nu\left(\mathbf{x}^k\right) + (\lambda^\nu)^T g^\nu\left(\mathbf{x}^k\right) - \frac{\beta_\nu}{2}\left\|\lambda^\nu - \mu^{\nu,k-1}\right\|^2$ for given $(\mathbf{x}^k, \mu^{\nu,k-1})$ and the third term,

$$-\frac{\beta_\nu}{2}\left\|\lambda^{\nu,k} - \mu^{\nu,k-1}\right\|^2 = \begin{cases} -\frac{1}{2\beta_\nu}\left\|g^\nu(\mathbf{x}^k)\right\|^2 & \text{if} \quad \mu^{\nu,k-1} + \frac{1}{\beta_\nu} g^\nu(\mathbf{x}^k) \geq 0 \\ -\frac{\beta_\nu}{2}\left\|\mu^{\nu,k-1}\right\|^2 & \text{otherwise,} \end{cases}$$

is a given constant, we have that $(\lambda^{\nu,k})^T g^\nu\left(\mathbf{x}^k\right) \geq 0$. Hence,

$$\|\triangle_k\| = \left\| \left(\lambda^{\nu,k+1}\right)^T g^\nu\left(\mathbf{x}^{k+1}\right) - \left(\lambda^{\nu,k}\right)^T g^\nu\left(\mathbf{x}^k\right) \right\| \leq \frac{\beta_\nu}{2}\left\|\lambda^{\nu,k+1} - \lambda^{\nu,k}\right\|^2, \tag{41}$$

Combining (40) and (41) and invoking Lemma 3, we obtain

$$\begin{aligned} \left\| \left(\lambda^{\nu,k+1} - \lambda^{\nu,k}\right)^T g^\nu(\mathbf{x}^{k+1}) \right\| &\leq \left\| \left(\lambda^{\nu,k} - \lambda^{\nu,k+1}\right)^T \left(g^\nu(\mathbf{x}^{k+1}) - g^\nu(\mathbf{x}^k)\right) \right\| + \frac{\beta_\nu}{2}\left\|\lambda^{\nu,k+1} - \lambda^{\nu,k}\right\|^2 \\ &\leq L_g \left\|\lambda^{\nu,k+1} - \lambda^{\nu,k}\right\| \left\|\mathbf{x}^{k+1} - \mathbf{x}^k\right\| + \frac{L^2_{g^\nu}}{2\beta_\nu}\left\|\mathbf{x}^{k+1} - \mathbf{x}^k\right\|^2 \\ &\leq \frac{3L^2_{g^\nu}}{2\beta_\nu}\left\|\mathbf{x}^{k+1} - \mathbf{x}^k\right\|^2. \end{aligned} \tag{42}$$

Notice that the above upper bound on $\left\|(\lambda^{\nu,k+1} - \lambda^{\nu,k})^T g^\nu(\mathbf{x}^{k+1})\right\|$ includes the upper bound in Case 1. Therefore, by combining (37), (38) and (42), we obtain the desired result:

$$\mathcal{L}^\nu_{\alpha\beta}(\mathbf{x}^{k+1}, z^{\nu,k+1}, \lambda^{\nu,k+1}, \mu^{\nu,k+1}) \leq \mathcal{L}^\nu_{\alpha\beta}(\mathbf{x}^k, z^{\nu,k}, \lambda^{\nu,k}, \mu^{\nu,k}) - \frac{1}{2}\left(\gamma_\nu - L_\nu - \frac{3L_g^2}{\beta_\nu}\right)\left\|\mathbf{x}^{k+1} - \mathbf{x}^k\right\|^2,$$

which implies that the sequence $\left\{\mathcal{L}^\nu_{\alpha\beta}\left(\mathbf{x}^k, z^{\nu,k}, \lambda^{\nu,k}, \mu^{\nu,k}\right)\right\}$ is monotonically decreasing if γ_ν is chosen such that $\gamma_\nu > L_\nu + \frac{3L^2_{g^\nu}}{\beta_\nu}$ with a suitable choice of $\beta_\nu > 0$.

Next, we show that $\{\mathcal{L}^{\nu}_{\alpha\beta}\}$ is convergent. We know from Theorem 2 that a saddle point of $\mathcal{L}^{\nu}_{\alpha\beta}$ exists. Let $(\mathbf{x}^*, z^{\nu,*}, \lambda^{\nu,*}, \mu^{\nu,*})$ be a saddle point of $\mathcal{L}^{\nu}_{\alpha\beta}(\mathbf{x}, z^{\nu}, \lambda^{\nu}, \mu^{\nu})$. By the updating rules for $(\lambda^{\nu,k+1}, \mu^{\nu,k+1})$ defined as maximizers for the updated $(\mathbf{x}^{k+1}, z^{k+1})$, we see that

$$\mathcal{L}^{\nu}_{\alpha\beta}\left(\mathbf{x}^{k+1}, z^{\nu,k+1}, \lambda^{\nu,k+1}, \mu^{\nu,k+1}\right) \geq \mathcal{L}^{\nu}_{\alpha\beta}\left(\mathbf{x}^{k+1}, z^{\nu,k+1}, \lambda^{\nu,*}\mu^{\nu,*}\right) \geq \mathcal{L}^{\nu}_{\alpha\beta}\left(\mathbf{x}^{*}, z^{\nu,*}, \lambda^{\nu,*}, \mu^{\nu,*}\right) > -\infty, \quad (43)$$

which implies that the sequence $\left\{\mathcal{L}^{\nu}_{\alpha\beta}\left(\mathbf{x}^{k}, z^{\nu,k}, \lambda^{\nu,k}, \mu^{\nu,k}\right)\right\}$ is lower bounded by a finite value of $\mathcal{L}^{\nu}_{\alpha\beta}\left(\mathbf{x}^{*}, z^{\nu,*}, \lambda^{\nu,*}, \mu^{\nu,*}\right)$. Thus, with the choice of $\gamma_\nu > 0$ such that $\gamma_\nu \geq L_\nu + \frac{3L^2_{g^\nu}}{\beta_\nu}$, the sequence $\left\{\mathcal{L}^{\nu}_{\alpha\beta}\left(\mathbf{x}^{k}, z^{\nu,k}, \lambda^{\nu,k}, \mu^{\nu,k}\right)\right\}$ converges to a finite limit, denoted by $\underline{\mathcal{L}^{\nu}}$, as $k \to \infty$. □

Next, we provide our key results that the generated sequence is bounded and asymptotic regular.

Theorem 3. *Suppose that Assumptions 1–3 hold and that there exists a GNE of the GNEP (1) satisfying the KKT conditions (3) for* $\nu = 1, \ldots, N$. *Let* $\left\{(x^{\nu,k}, z^{\nu,k}, \lambda^{\nu,k}, \mu^{\nu,k})\right\}_{\nu=1}^{N}$ *be the sequence generated by Algorithm 1 with the parameters set to* $\gamma_\nu > 0$ *such that* $\gamma_\nu \geq L_\nu + \frac{3L^2_{g^\nu}}{\beta_\nu}$. *Then,*

(a) the primal sequence $\left\{\mathbf{x}^k\right\}$ *is bounded;*

(b) the sequence of the multiplier $\left\{\lambda^{\nu,k}\right\}$ *is bounded;*

(c) it holds that $\sum_{k=1}^{\infty}\left\|\mathbf{x}^{k+1} - \mathbf{x}^k\right\|^2 < \infty$ *and* $\sum_{k=1}^{\infty}\left\|\lambda^{\nu,k+1} - \lambda^{\nu,k}\right\|^2 < \infty$, *and hence*

$$\lim_{k\to\infty}\left\|\mathbf{x}^{k+1} - \mathbf{x}^k\right\| = 0, \quad \lim_{k\to\infty}\left\|\lambda^{\nu,k+1} - \lambda^{\nu,k}\right\| = 0, \quad \text{and} \quad \lim_{k\to\infty}\left\|\mu^{\nu,k+1} - \mu^{\nu,k}\right\| = 0. \quad (44)$$

Proof. (a) Recall from Theorem 2 that a saddle point $(\mathbf{x}^*, z^{\nu,*}, \lambda^{\nu,*}, \mu^{\nu,*})$ of $\mathcal{L}^{\nu}_{\alpha\beta}(\mathbf{x}, z^{\nu}, \lambda^{\nu}, \mu^{\nu})$ exists. From (43) in Lemma 4, we know that $\mathcal{L}^{\nu}_{\alpha\beta}(\mathbf{x}^{k}, z^{\nu,k}, \lambda^{\nu,k}, \mu^{\nu,k})$ is lower bounded by $\mathcal{L}^{\nu}_{\alpha\beta}\left(\mathbf{x}^{*}, z^{\nu,*}, \lambda^{\nu,*}, \mu^{\nu,*}\right)$. Since $\mathcal{L}^{\nu}_{\alpha\beta}\left(\mathbf{x}^{k}, z^{\nu,k}, \lambda^{\nu,k}, \mu^{\nu,k}\right)$ is nonincreasing, it is also upper bounded by a finite value, i.e, $\mathcal{L}^{\nu}_{\alpha\beta}(\mathbf{x}^{k}, z^{\nu,k}, \lambda^{\nu,k}, \mu^{\nu,k}) < \infty$. We thus have

$$-\infty < \mathcal{L}^{\nu}_{\alpha\beta}(\mathbf{x}^{k+1}, z^{\nu,k+1}, \lambda^{\nu,k+1}, \mu^{\nu,k+1}) = \theta(\mathbf{x}^{k+1}) + (\lambda^{\nu,k+1})^T g(\mathbf{x}^{k+1}) < +\infty.$$

Hence, $\left\{\mathbf{x}^k\right\}$ is bounded due to the coercivity of $\theta_\nu(\mathbf{x})$ (Assumption 3) with the facts $\lambda^{\nu,k+1} \geq 0$ and $(\lambda^{\nu,k+1})^T g^{\nu}(\mathbf{x}^{k+1}) \geq 0$.

(b) Note that since the function $\lambda^\nu \to \mathcal{L}_{\alpha\beta}(\mathbf{x}, z^\nu, \lambda^\nu, \mu^\nu)$ is strongly concave, there exists parameter $c^\lambda_\nu > 0$ such that for $\lambda^{\nu,*}, \lambda^{\nu,k+1} \in \mathbb{R}^{m_\nu}_+$ and for given $(\mathbf{x}^{k+1}, z^{\nu,k+1}, \mu^{\nu,k})$

$$\mathcal{L}^{\nu}_{\alpha\beta}(\lambda^{\nu,*}, \mu^{\nu,k}) \leq \mathcal{L}^{\nu}_{\alpha\beta}(\lambda^{\nu,k+1}, \mu^{\nu,k}) + \nabla_{\lambda^\nu}\mathcal{L}^{\nu}_{\alpha\beta}(\lambda^{\nu,k+1}, \mu^{\nu,k})^T(\lambda^{\nu,*} - \lambda^{\nu,k+1}) - \frac{c^\lambda_\nu}{2}\left\|\lambda^{\nu,k+1} - \lambda^{\nu,*}\right\|^2$$

$$\mathcal{L}^{\nu}_{\alpha\beta}(\lambda^{\nu,k+1}, \mu^{\nu,k}) \leq \mathcal{L}^{\nu}_{\alpha\beta}(\lambda^{\nu,*}, \mu^{\nu,k}) + \nabla_{\lambda^\nu}\mathcal{L}^{\nu}_{\alpha\beta}(\lambda^{\nu,*}, \mu^{\nu,k})^T(\lambda^{\nu,k+1} - \lambda^{\nu,*}) - \frac{c^\lambda_\nu}{2}\left\|\lambda^{\nu,k+1} - \lambda^{\nu,*}\right\|^2,$$

where we omitted $(\mathbf{x}^{k+1}, z^{\nu,k+1})$ for notational simplicity. Adding the above two inequalities yields

$$c^\lambda_\nu\left\|\lambda^{\nu,k+1} - \lambda^{\nu,*}\right\|^2 \leq \left(\nabla_{\lambda^\nu}\mathcal{L}^{\nu}_{\alpha\beta}(\lambda^{\nu,k+1}, \mu^{\nu,k}) - \nabla_{\lambda^\nu}\mathcal{L}^{\nu}_{\alpha\beta}(\lambda^{\nu,*}, \mu^{\nu,k})\right)^T\left(\lambda^{\nu,*} - \lambda^{\nu,k+1}\right).$$

Using the Cauchy-Schwarz inequality and the triangle inequality, we obtain

$$\begin{aligned}\left\|\lambda^{\nu,k+1}-\lambda^{\nu,*}\right\| &\leq \frac{1}{c_\nu^\lambda}\left\|\nabla_{\lambda^\nu}\mathcal{L}^\nu_{\alpha\beta}\left(\lambda^{\nu,k+1},\mu^{\nu,k}\right)-\nabla_{\lambda^\nu}\mathcal{L}^\nu_{\alpha\beta}(\lambda^{\nu,*},\mu^{\nu,k})\right\| \\ &\overset{(a)}{\leq} \frac{1}{c_\nu^\lambda}\left\|\beta_\nu\left(\lambda^{\nu,*}-\mu^{\nu,k}\right)-g^\nu(\mathbf{x}^{k+1})\right\| \leq \frac{\beta_\nu}{c_\nu^\lambda}\left\|\lambda^{\nu,k}-\lambda^{\nu,*}\right\|+\frac{1}{c_\nu^\lambda}\left\|g^\nu(\mathbf{x}^{k+1})\right\|.\end{aligned}$$

where the inequality (a) comes from the definition of $\lambda^{\nu,k+1}$ in Step 3, implying that $\lambda^{\nu,k+1} \geq \mu^{\nu,k}+\frac{1}{\beta_\nu}g(\mathbf{x}^{k+1})$. Since $\{\mathbf{x}^k\}$ is bounded and $g^\nu(\mathbf{x})$ is continuous differentiable (Assumption 1), there exists $D_\nu > 0$ such that $\left\|g^\nu\left(\mathbf{x}^{k+1}\right)\right\| \leq D_\nu$. From the update of $\mu^{\nu,k+1}=\lambda^{\nu,k+1}$ in Step 3, we have $\lambda^k=\mu^k$ for any $k\geq 1$. By taking $c_\nu^\lambda=\beta_\nu$ we have

$$\left\|\lambda^{\nu,k+1}-\lambda^{\nu,*}\right\| \leq \left\|\lambda^{\nu,k}-\lambda^{\nu,*}\right\|+D_\nu/\beta_\nu.$$

Therefore, the sequence $\left\{\lambda^{\nu,k}\right\}$ is bounded on any subset of $\mathbb{R}^{m_\nu}_+$.

(c) Invoking Lemma 4, we have that for all $k\geq 1$

$$\rho_\nu\left\|\mathbf{x}^{k+1}-\mathbf{x}^k\right\|^2 \leq \mathcal{L}^\nu_{\alpha\beta}\left(\mathbf{x}^k,z^{\nu,k},\lambda^{\nu,k},\mu^{\nu,k}\right)-\mathcal{L}^\nu_{\alpha\beta}\left(\mathbf{x}^{k+1},z^{\nu,k+1},\lambda^{\nu,k+1},\mu^{\nu,k+1}\right),$$

where $\rho_\nu := \frac{1}{2}\left(\gamma_\nu - L_\nu - \frac{3L_{g^\nu}}{2\beta_\nu}\right) > 0$. Summing the above inequality over $k=1,\ldots,K$, we obtain

$$\begin{aligned}\sum_{k=1}^K\left\|\mathbf{x}^{k+1}-\mathbf{x}^k\right\|^2 &\leq \frac{1}{\rho_\nu}\left(\mathcal{L}^\nu_{\alpha\beta}\left(\mathbf{x}^1,z^{\nu,1},\lambda^{\nu,1},\mu^{\nu,1}\right)-\mathcal{L}^\nu_{\alpha\beta}\left(\mathbf{x}^{K+1},z^{\nu,K+1},\lambda^{\nu,K+1},\mu^{\nu,K+1}\right)\right) \\ &\leq \frac{1}{\rho_\nu}\left(\mathcal{L}^\nu_{\alpha\beta}\left(\mathbf{x}^1,z^{\nu,1},\lambda^{\nu,1},\mu^{\nu,1}\right)-\theta_\nu\left(\mathbf{x}^*\right)\right),\end{aligned}$$

where the last inequality is from (43) and $\mathcal{L}^\nu_{\alpha\beta}(\mathbf{x}^*,z^{\nu,*},\lambda^{\nu,*},\mu^{\nu,*})=\theta_\nu\left(\mathbf{x}^*\right)$. Letting $K\to\infty$ yields

$$\sum_{k=1}^\infty\left\|\mathbf{x}^{k+1}-\mathbf{x}^k\right\|^2 < \infty,$$

from which, along with Lemma 3, it also follows immediately that $\sum_{k=1}^\infty\left\|\lambda^{\nu,k+1}-\lambda^{\nu,k}\right\|^2<\infty$ and $\sum_{k=1}^\infty\left\|\mu^{\nu,k+1}-\mu^{\nu,k}\right\|^2<\infty$. Therefore, we can deduce the desired results in (44). □

4.2. Main Convergence Results

We are ready to establish our main convergence results. We first show that any limit point of the sequence produced by Algorithm 1 is a saddle point of $\mathcal{L}^\nu_{\alpha\beta}$ for all $\nu=1,\ldots,N$.

THEOREM 4 **(Subsequence Convergence)**. *Suppose that Assumptions 1–3 hold and that there exists a GNE of the GNEP (1) such that the KKT conditions (3) are satisfied for every $\nu=1,\ldots,N$. Let $\left\{(x^{\nu,k},z^{\nu,k},\lambda^{\nu,k},\mu^{\nu,k})\right\}_{\nu=1}^N$ be the sequence generated by Algorithm 1. Then, the sequence $\left\{(x^{\nu,k},z^{\nu,k},\lambda^{\nu,k},\mu^{\nu,k})\right\}_{\nu=1}^N$ converges to a point $(\overline{\mathbf{x}},\overline{z}^\nu,\overline{\lambda}^\nu,\overline{\mu}^\nu)$ satisfying the saddle point condition (8).*

Proof. Since the sequence $\{(\mathbf{x}^k, z^{\nu,k}, \lambda^{\nu,k}, \mu^{\nu,k})\}$ is bounded, there exists at least one limit point. Let $(\overline{\mathbf{x}}, \overline{z}, \overline{\lambda}, \overline{\mu})$ be a limit point of $\{(\mathbf{x}^k, z^{\nu,k}, \lambda^{\nu,k}, \mu^{\nu,k})\}$, and let $\{(\mathbf{x}^{k_j}, z^{\nu,k_j}, \lambda^{\nu,k_j}, \mu^{\nu,k_j})\}$ be a subsequence converging to $(\overline{\mathbf{x}}, \overline{z}, \overline{\lambda}, \overline{\mu})$ as $j \to \infty$. From Theorem 3(c), it also follows that $\{(\mathbf{x}^{k_j+1}, z^{\nu,k_j+1}, \lambda^{\nu,k_j+1}, \mu^{\nu,k_j+1})\} \to (\overline{\mathbf{x}}, \overline{z}, \overline{\lambda}, \overline{\mu})$ as $j \to \infty$.

First, we show that a limit point $(\overline{\mathbf{x}}, \overline{z}^\nu, \overline{\lambda}^\nu, \overline{\mu}^\nu)$ satisfies the second inequality of the saddle point condition (8). Because $\{x^{\nu,k_j+1}\} \to \overline{x}^\nu$ and $\{x^{\nu,k_j}\} \to \overline{x}^\nu$ as $j \to \infty$, we have from Step 1 that

$$\overline{x}^\nu = \mathcal{P}_{\mathcal{X}_\nu}\left[\overline{x}^\nu - \sigma_\nu \nabla_{x^\nu} \widehat{\mathcal{L}^\nu_{\alpha\beta}}(\overline{\mathbf{x}}, \overline{z}^\nu, \overline{\lambda}^\nu, \overline{\mu}^\nu; \overline{\mathbf{x}})\right].$$

The limit point $\overline{x}^\nu$ is equivalent to a solution of the VI (Facchinei and Pang 2007, Prop. 1.5.8):

$$\nabla_{x^\nu} \widehat{\mathcal{L}^\nu_{\alpha\beta}}(\overline{\mathbf{x}}, \overline{z}^\nu, \overline{\lambda}^\nu, \overline{\mu}^\nu; \overline{\mathbf{x}})^T (x^\nu - \overline{x}^\nu) \geq 0, \quad \forall x^\nu \in \mathcal{X}_\nu.$$

Using the fact that $\nabla_{x^\nu} \widehat{\mathcal{L}^\nu_{\alpha\beta}}(\overline{\mathbf{x}}, \overline{z}^\nu, \overline{\lambda}^\nu, \overline{\mu}^\nu; \overline{\mathbf{x}}) = \nabla_{x^\nu} \mathcal{L}^\nu_{\alpha\beta}(\overline{\mathbf{x}}, \overline{z}^\nu, \overline{\lambda}^\nu, \overline{\mu}^\nu)$ and the convexity of $\mathcal{L}^\nu_{\alpha\beta}$ with respect to x^ν, we obtain the first-order optimality condition for $\mathcal{L}^\nu_{\alpha\beta}$:

$$\nabla_{x^\nu} \mathcal{L}^\nu_{\alpha\beta}(\overline{\mathbf{x}}, \overline{z}^\nu, \overline{\lambda}^\nu, \overline{\mu}^\nu)^T (x^\nu - \overline{x}^\nu) \geq 0, \quad \forall x^\nu \in \mathcal{X}_\nu,$$

which implies that $(\overline{\mathbf{x}}, \overline{z}^\nu, \overline{\lambda}^\nu, \overline{\mu}^\nu)$ satisfies the second inequality of the saddle point condition (8):

$$\mathcal{L}^\nu_{\alpha\beta}(\overline{x}^\nu, \overline{x}^{-\nu}, \overline{z}^\nu, \overline{\lambda}^\nu, \overline{\mu}^\nu) \leq \mathcal{L}^\nu_{\alpha\beta}(x^\nu, \overline{x}^{-\nu}, \overline{z}^\nu, \overline{\lambda}^\nu, \overline{\mu}^\nu).$$

Similarly, by the definitions $\lambda^{\nu,k+1}$ and $\mu^{\nu,k+1}$ (as maximizers) in Step 3, the limit points $(\overline{\lambda}^\nu, \overline{\mu}^\nu)$ maximize $\mathcal{L}^\nu_{\alpha\beta}(\overline{\mathbf{x}}, z^\nu(\lambda^\nu, \mu^\nu), \lambda^\nu, \mu^\nu)$. We thus have that

$$\nabla_{\lambda^\nu} \mathcal{L}^\nu_{\alpha\beta}(\overline{\mathbf{x}}, \overline{z}^\nu, \overline{\lambda}^\nu, \overline{\mu}^\nu)^T (\lambda^\nu - \overline{\lambda}^\nu) \leq 0, \quad \forall \lambda^\nu \in \mathbb{R}^{m_\nu}_+,$$
$$\nabla_{\mu^\nu} \mathcal{L}^\nu_{\alpha\beta}(\overline{\mathbf{x}}, \overline{z}^\nu, \overline{\lambda}^\nu, \overline{\mu}^\nu)^T (\mu^\nu - \overline{\mu}^\nu) \leq 0, \quad \forall \mu^\nu \in \mathbb{R}^{m_\nu}.$$

Consequently, $(\overline{\mathbf{x}}, \overline{z}^\nu, \overline{\lambda}^\nu, \overline{\mu}^\nu)$ satisfies the first inequality of the saddle point condition (8). □

We now strengthen the subsequence convergence result, under an additional assumption that θ_ν and g^ν_i satisfy the *Kurdyka-Łojasiewicz (KŁ) property* (Łojasiewicz 1963, Kurdyka 1998). Before proceeding with global convergence, we briefly review the KŁ property.

Definition 3 (KŁ Property & KŁ function). Let $\delta \in (0, +\infty]$. Denote by Φ_δ the class of all concave and continuous functions $\varphi : [0, \delta) \to \mathbb{R}_+$, which satisfy the following conditions:

(i) $\varphi(0) = 0$;

(ii) φ is continuously differentiable (C^1) on $[0, \delta)$ and continuous at 0;

(iii) for all $s \in (0, \delta) : \varphi' > 0$.

A proper and lower semicontinuous function $\Psi : \mathbb{R}^n \to (-\infty, +\infty]$ is said to have the *Kurdyka-Łojasiewicz (KŁ) property* at $\overline{u} \in \operatorname{dom} \partial\Psi := \{u \in \mathbb{R}^n : \partial\Psi(u) = \emptyset\}$ if there exist $\delta \in (0, +\infty]$, a neighborhood U of $\overline{u}$ and a function $\varphi \in \Phi_\delta$, such that

$$\varphi'(\Psi(u) - \Psi(\overline{u})) \cdot \operatorname{dist}(0, \partial\Psi(u)) \geq 1$$

for all $u \in U(\overline{u}) \cap \{u : \Psi(\overline{u}) < \Psi(u) < \Psi(\overline{u}) + \delta\}$. The function Ψ satisfying the KŁ property at each point of $\operatorname{dom} \partial\Psi$ is called a *KŁ function*.

Lemma 5 **(Uniformized KŁ Property (Bolte et al. 2014, Lemma 6))**. *Let Ω be a compact set and let $\Psi : \mathbb{R}^n \to (-\infty, \infty]$ be proper, lower semicontinuous function. Assume that Ψ is constant on Ω and satisfies the KŁ property at each point of Ω. Then there exist $\varepsilon > 0$, δ and $\varphi \in \Phi_\delta$ such that for all $\overline{u}$ in Ω and all u in the following intersection:*

$$\{u \in \mathbb{R}^n : \operatorname{dist}(u, \Omega) < \varepsilon\} \cap [\Psi(\overline{u}) < \Psi(u) < \Psi(\overline{u}) + \delta] \tag{45}$$

one has,

$$\varphi'(\Psi(u) - \Psi(\overline{u})) \cdot \operatorname{dist}(0, \partial\Psi(u)) \geq 1. \tag{46}$$

If Ψ is continuously differentiable and $\Psi(\overline{u}) = 0$, the inequality (46) can be rewritten as

$$\varphi'(\Psi(u) - \Psi(\overline{u})) \, \|\nabla\Psi(u)\| \geq 1.$$

With the uniformized KŁ property, we can prove that the generated sequence has finite length, and hence the *whole* sequence converges to a saddle point. The techniques developed in Bolte et al. (2014) are extended to our smooth constrained game setting with some modifications.

In order to exploit Lemma 5 for proving global convergence, we use the size of the gradient of the P-Lagrangian, denoted by $\widetilde{\nabla}\mathcal{L}_{\alpha\beta}$, and derive an upper bound on the gradient. Noting that θ_ν and g^ν are continuously differentiable, $x^\nu \in \mathcal{X}_\nu$, and $\lambda^\nu \in \mathbb{R}^{m_\nu}_+$, we consider the projected gradients of $\mathcal{L}^\nu_{\alpha\beta}$ in x^ν and λ^ν for x^ν-component and λ^ν-component of $\widetilde{\nabla}\mathcal{L}_{\alpha\beta}$:

$$\begin{aligned}
\widetilde{\nabla}_{x^\nu}\mathcal{L}^\nu_{\alpha\beta}(\mathbf{x}, z^\nu, \lambda^\nu, \mu^\nu) &:= x^\nu - \mathcal{P}_{\mathcal{X}_\nu}\left[x^\nu - \nabla_{x^\nu}\mathcal{L}^\nu_{\alpha\beta}(\mathbf{x}, z^\nu, \lambda^\nu, \mu^\nu)\right], \\
\widetilde{\nabla}_{\lambda^\nu}\mathcal{L}^\nu_{\alpha\beta}(\mathbf{x}, z^\nu, \lambda^\nu, \mu^\nu) &:= \lambda^\nu - \left[\lambda^\nu + \nabla_{\lambda^\nu}\mathcal{L}^\nu_{\alpha\beta}(\mathbf{x}, z^\nu, \lambda^\nu, \mu^\nu)\right]^+.
\end{aligned}$$

Let us now define the *projected gradient* of $\mathcal{L}^\nu_{\alpha\beta}$ at $(\mathbf{x}^{k+1}, z^{\nu,k+1}, \lambda^{\nu,k+1}, \mu^{\nu,k+1})$ as

$$\widetilde{\nabla}\mathcal{L}^\nu_{\alpha\beta}(\mathbf{w}^{\nu,k+1}) := \begin{pmatrix} q^{k+1}_{x^\nu} \\ q^{k+1}_{z^\nu} \\ q^{k+1}_{\lambda^\nu} \\ q^{k+1}_{\mu^\nu} \end{pmatrix} = \begin{pmatrix} x^{\nu,k+1} - \mathcal{P}_{\mathcal{X}_\nu}\left[x^{\nu,k+1} - \nabla_{x^\nu}\mathcal{L}^\nu_{\alpha\beta}(\mathbf{x}^{k+1}, z^{\nu,k+1}, \lambda^{\nu,k+1}, \mu^{\nu,k+1})\right] \\ \nabla_{z^\nu}\mathcal{L}^\nu_{\alpha\beta}(\mathbf{x}^{k+1}, z^{\nu,k+1}, \lambda^{\nu,k+1}, \mu^{\nu,k+1}) \\ \lambda^{\nu,k+1} - \left[\lambda^{\nu,k+1} + \nabla_{\lambda^\nu}\mathcal{L}^\nu_{\alpha\beta}(\mathbf{x}^{k+1}, z^{\nu,k+1}, \lambda^{\nu,k+1}, \mu^{\nu,k+1})\right]^+ \\ \nabla_{\mu^\nu}\mathcal{L}^\nu_{\alpha\beta}(\mathbf{x}^{k+1}, z^{\nu,k+1}, \lambda^{\nu,k+1}, \mu^{\nu,k+1}) \end{pmatrix}. \tag{47}$$

It is clear that if $\widetilde{\nabla}\mathcal{L}^{\nu}_{\alpha\beta}(\mathbf{w}^{\nu,k+1}) \to 0$, a saddle point of $\mathcal{L}^{\nu}_{\alpha\beta}(\mathbf{w}^{\nu})$ is obtained. We derive an upper bound on $\widetilde{\nabla}\mathcal{L}^{\nu}_{\alpha\beta}(\mathbf{w}^{\nu,k+1})$ in terms of the generated iterates. In addition, we recall that by Assumption 2 imply there exist constants $M_{\nabla\theta_\nu}$ and $M_{\nabla g^\nu}$ such that

$$\left\|\nabla_{x^\nu}\theta_\nu\left(\mathbf{x}_1\right) - \nabla_{x^\nu}\theta_\nu\left(\mathbf{x}_2\right)\right\| \le M_{\nabla\theta_\nu}\left\|\mathbf{x}_1 - \mathbf{x}_2\right\|, \quad \forall \mathbf{x}_1, \mathbf{x}_2 \in \mathbf{X}, \tag{48a}$$

$$\left\|\nabla_{x^\nu} g^\nu\left(\mathbf{x}_1\right) - \nabla_{x^\nu} g^\nu\left(\mathbf{x}_2\right)\right\| \le M_{\nabla g^\nu}\left\|\mathbf{x}_1 - \mathbf{x}_2\right\|, \quad \forall \mathbf{x}_1, \mathbf{x}_2 \in \mathbf{X}_\nu, \tag{48b}$$

Lemma 6. *Let $\left\{\mathbf{w}^{\nu,k}\right\}_{\nu=1}^{N}$ be the sequence generated by Algorithm 1. Then, for every $\nu = 1, \ldots, N$, there exist constant $C_\nu > 0$ such that for all $k \ge 0$*

$$\|\widetilde{\nabla}\mathcal{L}^{\nu}_{\alpha\beta}(\mathbf{w}^{\nu,k+1})\| \le C_\nu \|\mathbf{x}^{k+1} - \mathbf{x}^k\|. \tag{49}$$

Proof. We first estimate an upper bound for $q^{k+1}_{x^\nu}$ in $\widetilde{\nabla}\mathcal{L}^{\nu}_{\alpha\beta}(\mathbf{w}^{\nu,k+1})$. Recall that there exists a unique solution $\widehat{\mathbf{x}}^k$ of $\mathrm{VI}^k(\mathbf{X}, \widehat{\mathbf{L}}^k)$ in (18) at each iteration k (Lemma 2), and denote by $\widehat{x}^{\nu,k}$ the νth component of $\widehat{\mathbf{x}}^k$. From the fixed-point characterization of $\widehat{x}^{\nu,k}$, we know that for every $\nu = 1, \ldots, N$,

$$\widehat{x}^{\nu,k} = \mathcal{P}_{\mathcal{X}_\nu}\left[\widehat{x}^{\nu,k} - \left(\nabla_{x^\nu}\theta_\nu(\mathbf{x}^k) + \nabla_{x^\nu} g^\nu(\mathbf{x}^k)\lambda^{\nu,k} + \gamma_\nu(\widehat{x}^{\nu,k} - x^{\nu,k})\right)\right].$$

Hence,

$$\begin{aligned}
\left\|q^{k+1}_{x^\nu}\right\| &= \left\|x^{\nu,k+1} - \widehat{x}^{\nu,k}\right\| + \left\|\mathcal{P}_{\mathcal{X}_\nu}\left[\widehat{x}^{\nu,k} - \nabla_{x^\nu}\theta_\nu(\mathbf{x}^k) - \nabla_{x^\nu} g^\nu(\mathbf{x}^k)\lambda^{\nu,k} - \gamma_\nu(\widehat{x}^{\nu,k} - x^{\nu,k})\right]\right. \\
&\qquad\qquad \left. - \mathcal{P}_{\mathcal{X}_\nu}\left[x^{\nu,k+1} - \nabla_{x^\nu}\theta_\nu(\mathbf{x}^{k+1}) - \nabla_{x^\nu} g^\nu(\mathbf{x}^{k+1})\lambda^{\nu,k+1}\right]\right\| \\
&\overset{(a)}{\le} \left\|x^{\nu,k+1} - \widehat{x}^{\nu,k}\right\| + \left\|\left[\widehat{x}^{\nu,k} - \nabla_{x^\nu}\theta_\nu(\mathbf{x}^k) - \nabla_{x^\nu} g^\nu(\mathbf{x}^k)\lambda^{\nu,k} - \gamma_\nu(\widehat{x}^{\nu,k} - x^{\nu,k})\right]\right. \\
&\qquad\qquad \left. - \left[x^{\nu,k+1} - \nabla_{x^\nu}\theta_\nu(\mathbf{x}^{k+1}) - \nabla_{x^\nu} g^\nu(\mathbf{x}^{k+1})\lambda^{\nu,k+1}\right]\right\| \\
&\overset{(b)}{\le} (2+\gamma_\nu)\left\|\mathbf{x}^{k+1} - \mathbf{x}^k\right\| + \left\|\nabla_{x^\nu}\theta_\nu(\mathbf{x}^{k+1}) - \nabla_{x^\nu}\theta_\nu(\mathbf{x}^k) + \nabla_{x^\nu} g^\nu(\mathbf{x}^{k+1})\lambda^{\nu,k+1} - \nabla_{x^\nu} g^\nu(\mathbf{x}^k)\lambda^{\nu,k}\right\|,
\end{aligned}$$

where (a) follows from the non-expansive property of the projection operator, and (b) is due to the facts that $\left\|x^{\nu,k+1} - \widehat{x}^{\nu,k}\right\| \le \left\|x^{\nu,k+1} - x^{\nu,k}\right\|$ and $\left\|x^{\nu,k+1} - x^{\nu,k}\right\| \le \left\|\mathbf{x}^{k+1} - \mathbf{x}^k\right\|$. Then, by adding and subtracting $g^\nu(\mathbf{x}^k)\lambda^{\nu,k+1}$ and using the triangle inequality, we obtain

$$\begin{aligned}
\left\|q^{k+1}_{x^\nu}\right\| &\le (2+\gamma_\nu)\left\|\mathbf{x}^{k+1} - \mathbf{x}^k\right\| + \left\|\nabla_{x^\nu}\theta_\nu(\mathbf{x}^{k+1}) - \nabla_{x^\nu}\theta_\nu(\mathbf{x}^k)\right\| \\
&\qquad + \left\|\nabla_{x^\nu} g^\nu(\mathbf{x}^{k+1})\lambda^{\nu,k+1} - \nabla_{x^\nu} g^\nu(\mathbf{x}^k)\lambda^{\nu,k+1}\right\| + \left\|\nabla_{x^\nu} g^\nu(\mathbf{x}^k)\lambda^{\nu,k+1} - \nabla_{x^\nu} g^\nu(\mathbf{x}^k)\lambda^{\nu,k}\right\| \\
&\le (2+\gamma_\nu)\left\|\mathbf{x}^{k+1} - \mathbf{x}^k\right\| + M_{\nabla\theta_\nu}\left\|\mathbf{x}^{k+1} - \mathbf{x}^k\right\| + M_{\nabla g^\nu} B_{\lambda^\nu}\left\|\mathbf{x}^{k+1} - \mathbf{x}^k\right\| + R_{g^\nu}\left\|\lambda^{\nu,k+1} - \lambda^{\nu,k}\right\| \\
&\le \left(2 + \gamma_\nu + M_{\nabla\theta_\nu} + M_{\nabla g^\nu} B_{\lambda^\nu} + \frac{R_{g^\nu} L_{g^\nu}}{\beta_\nu}\right)\left\|\mathbf{x}^{k+1} - \mathbf{x}^k\right\|,
\end{aligned} \tag{50}$$

where the second inequality is due to the Lipschitz continuity of $\nabla_{x^\nu}\theta_\nu$ and $\nabla_{x^\nu} g^\nu$ and the boundedness of $\left\{\mathbf{x}^k\right\}$ and $\left\{\lambda^{\nu,k}\right\}$, implying there exist constants $B_{\lambda^\nu} := \max_{k\in\mathbb{N}}\left\|\lambda^{\nu,k}\right\|$ and $R_{g^\nu} := \max_{k\in\mathbb{N}}\left\|\nabla_{x^\nu} g^\nu(\mathbf{x}^k)\right\|$; the last inequality is from $\left\|\lambda^{\nu,k+1} - \lambda^{\nu,k}\right\| \le \frac{L_{g^\nu}}{\beta_\nu}\left\|\mathbf{x}^{k+1} - \mathbf{x}^k\right\|$ (Lemma 3).

Next, by the definition $\lambda^{\nu,k+1} \in \mathbb{R}^{m_\nu}_+$ as a maximizer, $\lambda^{\nu,k+1}$ is characterized by

$$\lambda^{\nu,k+1} = \left[\lambda^{\nu,k+1} + \nabla_{\lambda^\nu} \mathcal{L}^\nu_{\alpha\beta}(\mathbf{x}^{k+1}, z^{\nu,k+1}, \lambda^{\nu,k+1}, \mu^{\nu,k})\right]^+,$$

which, together with the nonexpansive property of the projection onto $\mathbb{R}^{m_\nu}_+$ and Lemma 3, yields

$$\begin{aligned}
\left\|q^{k+1}_{\lambda^\nu}\right\| &= \Big\| \left[\lambda^{\nu,k+1} + (g^\nu(\mathbf{x}^{k+1}) - z^{\nu,k+1}) - \beta_\nu(\lambda^{\nu,k+1} - \mu^{\nu,k})\right]^+ \\
&\qquad - \left[\lambda^{\nu,k+1} + (g^\nu(\mathbf{x}^{k+1}) - z^{\nu,k+1}) - \beta_\nu(\lambda^{\nu,k+1} - \mu^{\nu,k+1})\right]^+ \Big\| \\
&\le \left\|\beta_\nu(\mu^{\nu,k+1} - \mu^{\nu,k})\right\| \le L_{g^\nu} \left\|\mathbf{x}^{k+1} - \mathbf{x}^k\right\|.
\end{aligned} \tag{51}$$

In addition, recalling that the definitions of $z^{\nu,k+1}$ in Step 2 and $\mu^{\nu,k+1}$ in Step 3, we have

$$\left\|q^{k+1}_{z^\nu}\right\| = \left\|(\mu^{\nu,k+1} - \lambda^{\nu,k+1}) + \alpha_\nu z^{\nu,k+1}\right\| = 0, \tag{52}$$

$$\left\|q^{k+1}_{\mu^\nu}\right\| = \left\|z^{\nu,k+1} + \beta_\nu(\lambda^{\nu,k+1} - \mu^{\nu,k+1})\right\| = 0. \tag{53}$$

Therefore, summing the inequalities (50) and (51), we deduce that for all $k \ge 0$

$$\left\|\widetilde{\nabla}\mathcal{L}(\mathbf{w}^{\nu,k+1})\right\| = \sum_{i = x^\nu, z^\nu, \lambda^\nu, \mu^\nu} \left\|q^{\nu,k+1}_i\right\| \le C_\nu \left\|\mathbf{x}^{k+1} - \mathbf{x}^k\right\|$$

with positive constant $C_\nu = 2 + \gamma_\nu + M_{\nabla\theta_\nu} + M_{\nabla g^\nu} B_{\lambda^\nu} + \frac{R_{g^\nu} L_{g^\nu}}{\beta_\nu} + L_{g^\nu}$. □

THEOREM 5 **(Global Convergence)**. *Suppose that the assumptions required for Theorem 4 hold. Let* $\left\{\mathbf{w}^{\nu,k} := (\mathbf{x}^k, z^{\nu,k}, \lambda^{\nu,k}, \mu^{\nu,k})\right\}_{\nu=1}^N$ *be the sequence generated by Algorithm 1. If* θ_ν *and* g^ν*,* $\nu = 1, \ldots, N$*, satisfy the KŁ property, then* $\left\{\mathbf{w}^{\nu,k}\right\}_{\nu=1}^N$ *has finite length, i.e.,*

$$\sum_{k=1}^{\infty} \left\|\mathbf{w}^{\nu,k+1} - \mathbf{w}^{\nu,k}\right\| < +\infty,$$

and the whole sequence $\left\{(x^{\nu,k}, z^{\nu,k}, \lambda^{\nu,k}, \mu^{\nu,k})\right\}_{\nu=1}^N$ *converges to a saddle point* $(\overline{\mathbf{x}}, \overline{z}^\nu, \overline{\lambda}^\nu, \overline{\mu}^\nu)$ *of* $\mathcal{L}^\nu_{\alpha\beta}$*.*

Proof. Let $\overline{\mathbf{w}}^\nu := (\overline{\mathbf{x}}, \overline{z}^\nu, \overline{\lambda}^\nu, \overline{\mu}^\nu)$ be a limit point of $\left\{\mathbf{w}^{\nu,k} = (\mathbf{x}^k, z^{\nu,k}, \lambda^{\nu,k}, \mu^{\nu,k})\right\}$ that is bounded for every $\nu = 1, \ldots, N$. Then, by the continuity of $\mathcal{L}^\nu_{\alpha\beta}$, we have

$$\lim_{k\to\infty} \mathcal{L}^\nu_{\alpha\beta}(\mathbf{w}^{\nu,k}) = \mathcal{L}^\nu_{\alpha\beta}(\overline{\mathbf{w}}^\nu). \tag{54}$$

In the following, we consider two cases:

Case 1. Suppose that there exists an integer $\bar{k}$ such that $\mathcal{L}^\nu_{\alpha\beta}(\mathbf{w}^{\nu,\bar{k}}) = \mathcal{L}^\nu_{\alpha\beta}(\overline{\mathbf{w}}^\nu)$ for $\nu = 1, \ldots, N$. Since the sequence $\left\{\mathcal{L}^\nu_{\alpha\beta}\right\}$ is nonincreasing, we have that $\mathcal{L}^\nu_{\alpha\beta}(\mathbf{w}^{\nu,k}) = \mathcal{L}^\nu_{\alpha\beta}(\overline{\mathbf{w}}^\nu)$ for all $k \ge \bar{k}$. Then, we have from Lemma 4 that for any $t \ge 0$

$$\rho_\nu \left\|\mathbf{x}^{k+t} - \mathbf{x}^k\right\|^2 \le \mathcal{L}^\nu_{\alpha\beta}(\mathbf{w}^{\nu,k}) - \mathcal{L}^\nu_{\alpha\beta}(\mathbf{w}^{\nu,k+t}) = 0,$$

which leads to

$$\mathbf{x}^{k+1} - \mathbf{x}^k = 0, \qquad \forall k \geq \bar{k}, \tag{55}$$

From Lemma 3, we also obtain that $\lambda^{\nu,k+1} - \lambda^{\nu,k} = 0$ and $\mu^{\nu,k+1} - \mu^{\nu,k} = 0$ for all $k \geq \bar{k}$. Therefore, $\left\{\mathbf{w}^{\nu,k} = (\mathbf{x}^k, z^{\nu,k}, \lambda^{\nu,k}, \mu^{\nu,k})\right\}$ must be eventually constant (stationary), and it thus has finite length.

Case 2. Consider the case where such an integer $\bar{k}$ does not exist (and every $\left\{\mathbf{w}^{\nu,k}\right\}$ is non-stationary) for $\nu = 1, \ldots, N$. In this case, we first show that the P-Lagrangian $\mathcal{L}^\nu_{\alpha\beta}$ is finite and constant on the set of all limit points $\omega_\nu(\mathbf{w}^\nu_0)$ of $\left\{\mathbf{w}^{\nu,k}\right\}$, and then apply Lemma 5 to show that $\left\{\mathbf{w}^{\nu,k}\right\}$ is a Cauchy sequence and convergent.

First, since the sequence $\{\mathcal{L}^\nu_{\alpha\beta}\}$ is nonincreasing, we have $\mathcal{L}^\nu_{\alpha\beta}(\mathbf{w}^{\nu,k}) > \mathcal{L}^\nu_{\alpha\beta}(\overline{\mathbf{w}}^\nu)$ for all k. This, along with (54), implies that there exists an integer k_0 such that for any $\varepsilon > 0$ and $\delta > 0$:

$$\mathcal{L}^\nu_{\alpha\beta}(\overline{\mathbf{w}}^\nu) < \mathcal{L}^\nu_{\alpha\beta}(\mathbf{w}^{\nu,k}) < \mathcal{L}^\nu_{\alpha\beta}(\overline{\mathbf{w}}^\nu) + \delta \quad \text{and} \quad \operatorname{dist}(\mathbf{w}^{\nu,k}, \omega(\mathbf{w}^\nu_0)) < \varepsilon \quad \text{for all } k \geq k_0, \tag{56}$$

where the second comes from the fact that $\lim_{k\to\infty}\operatorname{dist}(\mathbf{w}^{\nu,k}, \omega(\mathbf{w}^\nu_0)) = 0$ (see Theorem 4). Thus $\left\{\mathbf{w}^{\nu,k}\right\}$ belongs to the intersection in (45) with $\Omega = \omega_\nu(\mathbf{w}^\nu_0)$ for all $k \geq k_0$, and $\Omega = \omega_\nu(\mathbf{w}^\nu_0)$ is nonempty and compact. Recall that $\{\mathcal{L}^\nu_{\alpha\beta}\}$ is bounded below by the value of $\mathcal{L}^\nu_{\alpha\beta}$ at a saddle point, and hence $\{\mathcal{L}^\nu_{\alpha\beta}\}$ converges to a finite limit, denoted by $\underline{\mathcal{L}^\nu}$. It then follows from (54) that $\underline{\mathcal{L}^\nu} = \mathcal{L}^\nu_{\alpha\beta}(\overline{\mathbf{w}}^\nu)$, which shows that $\mathcal{L}^\nu_{\alpha\beta}$ is finite and constant on $\omega^\nu(\overline{\mathbf{w}}^\nu_0)$.

Thus, since $\mathcal{L}^\nu_{\alpha\beta}$ is a KL function, by applying Lemma 5 with $\Omega = \omega^\nu(\mathbf{w}^\nu_0)$ and $\partial\Psi(u) = \widetilde{\nabla}\mathcal{L}^\nu_{\alpha\beta}(\mathbf{w}^{\nu,k})$, we get that for any $k > k_0$

$$\varphi'\left(\mathcal{L}^\nu_{\alpha\beta}(\mathbf{w}^{\nu,k}) - \mathcal{L}^\nu_{\alpha\beta}(\overline{\mathbf{w}}^\nu)\right) \cdot \operatorname{dist}\left(0, \widetilde{\nabla}\mathcal{L}^\nu_{\alpha\beta}(\mathbf{w}^{\nu,k})\right) \geq 1,$$

which combined with Lemma 6 gives

$$\varphi'\left(\mathcal{L}^\nu_{\alpha\beta}(\mathbf{w}^{\nu,k}) - \mathcal{L}^\nu_{\alpha\beta}(\overline{\mathbf{w}}^\nu)\right) \geq \frac{1}{\operatorname{dist}\left(0, \widetilde{\nabla}\mathcal{L}^\nu_{\alpha\beta}(\mathbf{w}^{\nu,k})\right)} \geq \frac{1}{C_\nu \left\|\mathbf{x}^k - \mathbf{x}^{k-1}\right\|}. \tag{57}$$

On the other hand, since φ is concave function, we know that

$$\begin{aligned}\varphi\left(\mathcal{L}^\nu_{\alpha\beta}(\mathbf{w}^{\nu,k}) - \mathcal{L}^\nu_{\alpha\beta}(\overline{\mathbf{w}}^\nu)\right) - \varphi\left(\mathcal{L}^\nu_{\alpha\beta}(\mathbf{w}^{\nu,k+1}) - \mathcal{L}^\nu_{\alpha\beta}(\overline{\mathbf{w}}^\nu)\right)& \\ \geq \varphi'\left(\mathcal{L}^\nu_{\alpha\beta}(\mathbf{w}^{\nu,k}) - \mathcal{L}^\nu_{\alpha\beta}(\overline{\mathbf{w}}^\nu)\right)\left(\mathcal{L}^\nu_{\alpha\beta}(\mathbf{w}^{\nu,k}) - \mathcal{L}^\nu_{\alpha\beta}(\mathbf{w}^{\nu,k+1})\right)&.\end{aligned}$$

For convenience, we define for any $p, q \in \mathbb{N}$

$$\triangle_{p,q} := \varphi\left(\mathcal{L}^\nu_{\alpha\beta}(\mathbf{w}^{\nu,p}) - \mathcal{L}^\nu_{\alpha\beta}(\overline{\mathbf{w}}^\nu)\right) - \varphi\left(\mathcal{L}^\nu_{\alpha\beta}(\mathbf{w}^{\nu,q}) - \mathcal{L}^\nu_{\alpha\beta}(\overline{\mathbf{w}}^\nu)\right).$$

Then we get

$$\triangle_{k,k+1} \geq \varphi'\left(\mathcal{L}^\nu_{\alpha\beta}(\mathbf{w}^{\nu,k}) - \mathcal{L}^\nu_{\alpha\beta}(\overline{\mathbf{w}}^\nu)\right)\left(\mathcal{L}^\nu_{\alpha\beta}(\mathbf{w}^{\nu,k}) - \mathcal{L}^\nu_{\alpha\beta}(\mathbf{w}^{\nu,k+1})\right). \tag{58}$$

Recalling that $\mathcal{L}^{\nu}_{\alpha\beta}(\mathbf{w}^{\nu,k}) - \mathcal{L}^{\nu}_{\alpha\beta}(\mathbf{w}^{\nu,k+1}) \geq \rho_\nu \left\|\mathbf{x}^{k+1} - \mathbf{x}^k\right\|^2$, we combine (57) and (58) to obtain

$$\triangle_{k,k+1} \geq \frac{\rho_\nu \left\|\mathbf{x}^{k+1} - \mathbf{x}^k\right\|^2}{C_\nu \left\|\mathbf{x}^k - \mathbf{x}^{k-1}\right\|}.$$

Multiplying the above inequality by $\frac{C_\nu}{\rho_\nu} \left\|\mathbf{x}^k - \mathbf{x}^{k-1}\right\|$ gives

$$\left\|\mathbf{x}^{k+1} - \mathbf{x}^k\right\|^2 \leq \xi_\nu \triangle_{k,k+1} \left\|\mathbf{x}^k - \mathbf{x}^{k-1}\right\| \quad \text{where} \;\; \xi_\nu = C_\nu/\rho_\nu,$$

and hence $2\left\|\mathbf{x}^{k+1} - \mathbf{x}^k\right\| \leq 2\sqrt{\xi_\nu \triangle_{k,k+1} \left\|\mathbf{x}^k - \mathbf{x}^{k-1}\right\|}$. Using the inequality $2\sqrt{ab} \leq a + b$ for any $a, b \geq 0$ with $a = \left\|\mathbf{x}^k - \mathbf{x}^{k-1}\right\|$ and $b = \xi_\nu \triangle_{k,k+1}$, we have

$$2\left\|\mathbf{x}^{k+1} - \mathbf{x}^k\right\| \leq \left\|\mathbf{x}^k - \mathbf{x}^{k-1}\right\| + \xi_\nu \triangle_{k,k+1}. \tag{59}$$

Now we show that for any $k > k_0$ the following inequality holds:

$$2\sum_{l=k_0+1}^{k} \left\|\mathbf{x}^{l+1} - \mathbf{x}^l\right\| \leq \left\|\mathbf{x}^{k_0+1} - \mathbf{x}^{k_0}\right\| + \xi_\nu \triangle_{k_0+1,k+1}.$$

By summing (59) over $l = k_0 + 1, \ldots, k$, we have

$$\begin{aligned} 2\sum_{l=k_0+1}^{k} \left\|\mathbf{x}^{l+1} - \mathbf{x}^l\right\| &\leq \sum_{l=k_0+1}^{k} \left\|\mathbf{x}^l - \mathbf{x}^{l-1}\right\| + \xi_\nu \sum_{l=k_0+1}^{k} \triangle_{l,l+1} \\ &\leq \sum_{l=k_0+1}^{k} \left\|\mathbf{x}^{l+1} - \mathbf{x}^l\right\| + \left\|\mathbf{x}^{k_0+1} - \mathbf{x}^{k_0}\right\| + \xi_\nu \sum_{l=k_0+1}^{k} \triangle_{l,l+1} \end{aligned} \tag{60}$$

and using fact that $\triangle_{p,q} + \triangle_{q,r} = \triangle_{p,r}$ for all $p, q, r \in \mathbb{N}$, we get

$$\begin{aligned} \sum_{l=k_0+1}^{k} \triangle_{l,l+1} = \triangle_{k_0+1,k+1} &= \varphi\left(\mathcal{L}^{\nu}_{\alpha\beta}(\mathbf{w}^{\nu,k_0+1}) - \mathcal{L}^{\nu}_{\alpha\beta}(\overline{\mathbf{w}}^{\nu})\right) - \varphi\left(\mathcal{L}^{\nu}_{\alpha\beta}(\mathbf{w}^{\nu,k_0+2}) - \mathcal{L}^{\nu}_{\alpha\beta}(\overline{\mathbf{w}}^{\nu})\right) \\ &\leq \varphi\left(\mathcal{L}^{\nu}_{\alpha\beta}(\mathbf{w}^{\nu,k_0+1}) - \mathcal{L}^{\nu}_{\alpha\beta}(\overline{\mathbf{w}}^{\nu})\right) < \infty, \end{aligned} \tag{61}$$

where the last inequality is from the fact that $\varphi \geq 0$. Plugging (61) into (60), we obtain

$$\sum_{l=k_0+1}^{k} \left\|\mathbf{x}^{l+1} - \mathbf{x}^l\right\| \leq \left\|\mathbf{x}^{k_0+1} - \mathbf{x}^{k_0}\right\| + \xi_\nu \varphi\left(\mathcal{L}^{\nu}_{\alpha\beta}(\mathbf{w}^{\nu,k_0+1}) - \mathcal{L}^{\nu}_{\alpha\beta}(\overline{\mathbf{w}}^{\nu})\right) < \infty. \tag{62}$$

Since the right-hand side of (62) does not depends k, the sequence $\left\{\mathbf{x}^k\right\}$ has finite length, i.e.,

$$\sum_{k=1}^{\infty} \left\|\mathbf{x}^{k+1} - \mathbf{x}^k\right\| < \infty.$$

This implies $\left\{\mathbf{x}^k\right\}$ is a Cauchy sequence and thus a convergent sequence. By Lemma 3, the multiplier sequences $\{\lambda^{\nu,k}\}$ and $\{\mu^{\nu,k}\}$ are also Cauchy. Therefore, we conclude that the whole sequence $\left\{(\mathbf{x}^k, z^{\nu,k}, \lambda^{\nu,k}, \mu^{\nu,k})\right\}$ converges to a saddle point $(\overline{\mathbf{x}}, \overline{z}^{\nu}, \overline{\lambda}^{\nu}, \overline{\mu}^{\nu})$ of $\mathcal{L}^{\nu}_{\alpha\beta}$, $\nu = 1, \ldots, N$. $\square$

Note that verifying the KŁ property of a function might be difficult. However, it is known that *semi-algebraic* and *real-analytic* functions, which capture many applications, are classes of functions that satisfy the KŁ property; see e.g., Attouch and Bolte (2009), Attouch et al. (2013), Xu and Yin (2013), Li and Pong (2018) for an in-depth study of the KŁ functions and illustrating examples.

5. Computational Results

We present computational results to demonstrate the effectiveness of Algorithm 1. We conducted experiments on test problems taken from a library of GNEPs, as used in Facchinei and Kanzow (2010b), Dreves et al. (2011), Kanzow and Steck (2016). The experiments were carried out using MATLAB (R2018a) on a laptop with an Intel Core i5-6300U CPU 2.50GHz 8GB RAM. Two classes of instances were considered in the experiments: *general* GNEPs (A.1-A.10) and *jointly-convex* GNEPs (A.11-A.18). We refer the readers to Facchinei and Kanzow (2009) for a detailed description of the problems with data. Before present the results, it is noteworthy to mention how our test settings for the Arrow-Debreu equilibrium problems (A.10 (a)-(e)) differ from those in Facchinei and Kanzow (2009). Specifically, our setup includes production variables in consumers' constraints ($p^T x^i \le p^T \xi^i + \sum_{j=1}^J q_{ij} p^T y^j$), while the constraints in Facchinei and Kanzow (2009) were set to $p^T x^i \le p^T \xi^i$. This reflects the original Arrow-Debreu model better.

In the numerical experiments, we used the starting points listed in Facchinei and Kanzow (2010b), and the other variables' initial points were set to $(z^{\nu,0}, \lambda^{\nu,0}, \mu^{\nu,0}) = (\mathbf{0}, \mathbf{0}, \mathbf{0})$ for every $\nu = 1, \dots, N$. As for the parameters, we used fixed parameters set to $\alpha_\nu = 10$ and $\beta_\nu = 1$ for each player's P-Lagrangian across all test problems. In addition, a large parameter γ_ν was used so that $\gamma_\nu \ge L_\nu + \frac{3L_{g^\nu}^2}{\beta_\nu}$ and a diminishing step size σ_ν was simply used for every player ν in each problem. The stopping criterion is set as

$$\max_{\nu=1,\dots,N} \left\{ \left\| x^{\nu,k+1} - x^{\nu,k} \right\|_\infty, \left\| \lambda^{\nu,k+1} - \lambda^{\nu,k} \right\|_\infty \right\} \le 10^{-4}.$$

The computational results of our algorithm for the test problems are summarized in Table 1. The notations used in the table are as follows: the number of players 'N', the number of variables 'n', the number of constraints 'm', the starting point '$\mathbf{x}^0$' (a specific reported number indicates that all primal variables are uniformly initialized to that value), the total (cumulative) number of inner iterations 'Iter.', and the computation time in CPU seconds 'Time (s)'.

We make some remarks on the computational results. Algorithm 1 successfully solved all test problems. On the other hand, the exact penalty algorithm (Facchinei and Kanzow 2010b) failed to find solutions for problems A.2, A.7, and A.8, and the interior-point algorithm (Dreves et al. 2011) and the augmented Lagrangian method (Kanzow and Steck 2016) were unable to find a GNE for the instance A.8. This is attributed to their sensitivity to initial points and choices of parameters, while Algorithm 1 is insensitive to initialization and does not require (penalty) parameter updates. Additionally, algorithm 1 converges to a GNE in fewer iterations, due to its favorable structure: it employs a first-order scheme on strongly (smooth) convex approximations for the $\mathbf{x}$-update, along with exact maximization steps with fixed step sizes for the multiplier updates.

Furthermore, our algorithm consistently demonstrates fast convergence to a GNE in each problem. The convergence speed is primarily determined by how efficiently the subproblems are solved. Our algorithm employs a first-order scheme combined with Jacobi-type decomposition on a strongly convex approximation, and it includes a cost-effective projection onto the simple set $\mathcal{X}_\nu$ for the $\mathbf{x}$-update. These features enable the algorithm to circumvent the computational burden of solving a nonlinear system of equations during each (outer) iteration. Consequently, our algorithm achieves convergence to a GNE within a very short CPU time for each instance.

Illustrative Examples

To see how Algorithm 1 performs on GNEPs well, we provide numerical results for three important and practical instances with graphical illustrations.

Problem A.9 (a) (Power allocation in telecommunications). This model is described in detail in Pang et al. (2008) and represents a realistic communication system subject to Quality-of-Service (QoS) constraints. There are N links transmitting to K different Base Stations by using K different channels. Link ν transmits with power $x^\nu = (x_1^\nu, \ldots, x_K^\nu)$, and denote by $\mathbf{x} = (x^1, \ldots, x^N)$ the power allocation of all links. The GNEP model is defined by

$$\underset{x^\nu}{\text{minimize}} \quad \sum_{i=1}^{K} x_i^\nu \quad \text{subject to} \quad \sum_{i=1}^{K} \log_2\left(1 + \frac{h_i^{\nu\nu} x_i^\nu}{(\sigma_i^\nu)^2 + \sum_{\mu \neq \nu} h_i^{\nu\mu} x_i^\mu}\right) \geq L^\nu, \quad x^\nu \geq 0,$$

where $h_i^{\nu\mu}$ is the power gain between transmitter μ and receiver ν on the ith channel, $(\sigma_i^\nu)^2$ is the noise of link ν on the ith channel, and L^ν is the minimum transmission target rate for link ν.

This instance sets $\sigma_i^\nu = 0.3162$ for all ν and i, $K = 8$, $L^\nu = 8$ for all players, and the starting point was set to $(0, \ldots, 0)$. The data of coefficient h is given in Facchinei and Kanzow (2009). As shown in Figure 1, the P-Lagrangian values $\{\mathcal{L}_{\alpha\beta}^\nu\}$, $\nu = 1, \ldots, 7$, are monotonically decreasing and convergent, as expected. Additionally, Figure 2 illustrates that the iterates of x^ν, $\nu = 1, 3, 5$, converge to a point satisfying the minimum target rate of 8. Note that since the coupling constraints are relaxed into the objective with the multipliers, the projection onto $\mathcal{X}_\nu = \{x^\nu \in \mathbb{R}^{n_\nu} : x^\nu \geq 0\}$ performs efficiently, which leads to convergence to a GNE within a short CPU time of 0.32 seconds.

Problem A.10 (a) (Arrow-Debreu general equilibrium model). This model is introduced by Arrow and Debreu (1954) and described in Facchinei and Kanzow (2010a) in detail. In this instance, there are 8 players ($I = 5$, $J = 2$, and one market player) and 3 goods ($K = 3$). The market player sets (normalized) prices $p \in \mathbb{R}_+^K$ for the market clearing problem. The jth firm maximizes its profit by determining production quantity $y^j \in Y_j$, where $Y_j \subseteq \mathbb{R}^K$ is a production

Table 1 Computational results for Algorithm 1.

general GNEP	N	n	m	$\mathbf{x}^0$	Iter.	Time (s)
A.1	10	10	20	0.01	38	< 0.01
				0.1	36	< 0.01
				1	38	< 0.01
A.2	10	10	24	0.01	610	0.04
				0.1	536	0.04
				1	683	0.05
A.3	3	7	18	0	51	0.01
				1	51	0.01
A.4	3	7	18	0	7	< 0.01
				1	7	< 0.01
				10	7	< 0.01
A.5	3	7	18	0	82	0.02
				1	82	0.02
				10	82	0.02
A.6	3	7	21	0	49	0.02
				1	49	0.02
				10	49	0.02
A.7	4	20	44	0	48	0.02
				1	48	0.02
				10	48	0.02
A.8	3	3	8	0	45	< 0.01
				1	45	< 0.01
				10	45	< 0.01
A.9 (a)	7	56	63	0	108	0.32
A.9 (b)	7	112	119	0	135	1.24
A.10 (a)	8	24	33	0	780	0.10
A.10 (b)	25	125	151	1	1374	0.67
A.10 (c)	37	222	260	0	2154	1.12
A.10 (d)	37	370	408	1	3251	1.35
A.10 (e)	48	576	625	1	4728	2.54
jointly-convex GNEP	N	n	m	$\mathbf{x}^0$	Iter.	Time (s)
A.11	2	2	2	0	12	< 0.01
A.12	2	2	4	(2,0)	10	< 0.01
A.13	3	3	9	0	15	< 0.01
A.14	10	10	20	0.01	38	< 0.01
A.15	3	6	12	0	145	< 0.01
A.16 (P=75)	5	5	10	10	52	0.02
A.16 (P=100)	5	5	10	10	52	0.02
A.16 (P=150)	5	5	10	10	52	0.02
A.16 (P=200)	5	5	10	10	52	0.02
A.17	2	3	7	0	9	< 0.01
A.18	2	12	28	0	114	0.02

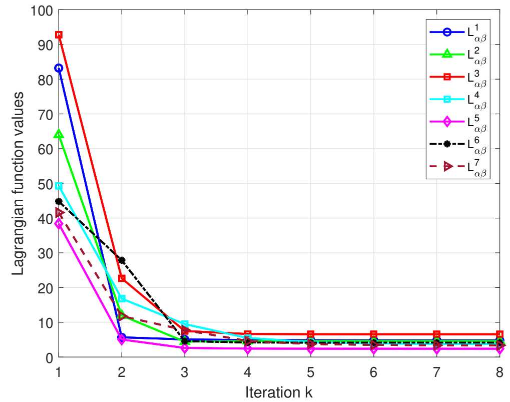
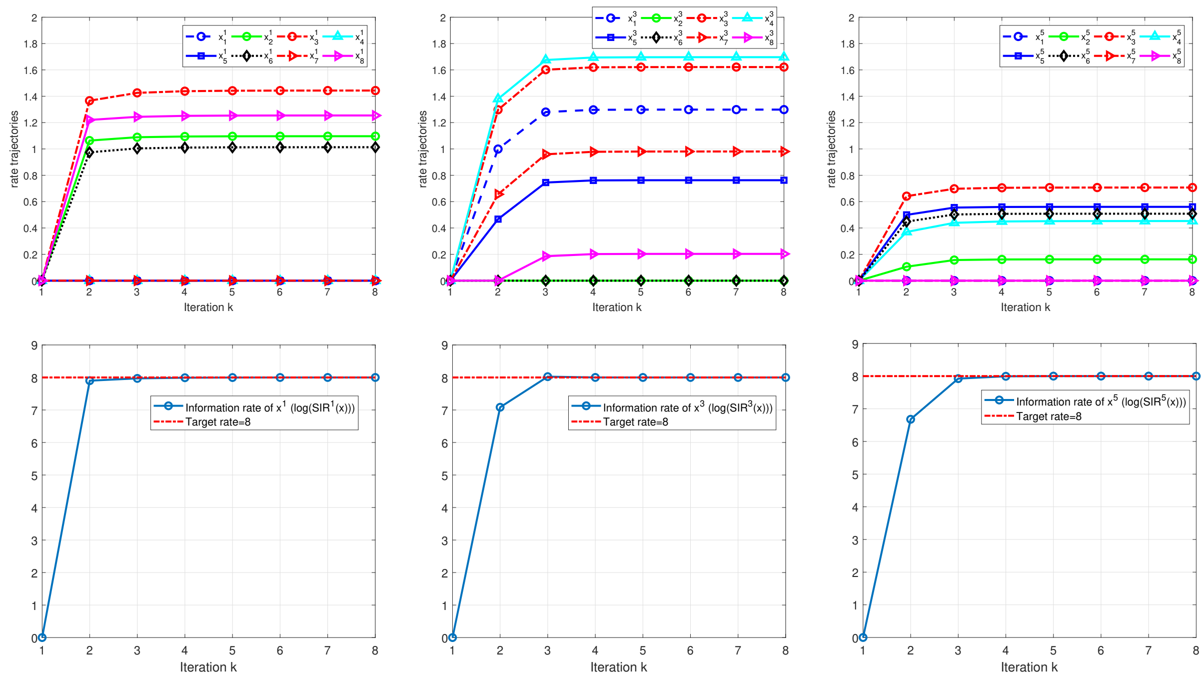


Figure 1 **Convergence behaviors of P-Lagrangian $\mathcal{L}^{\nu}_{\alpha\beta}$, $\nu = 1, \ldots, 7$.**



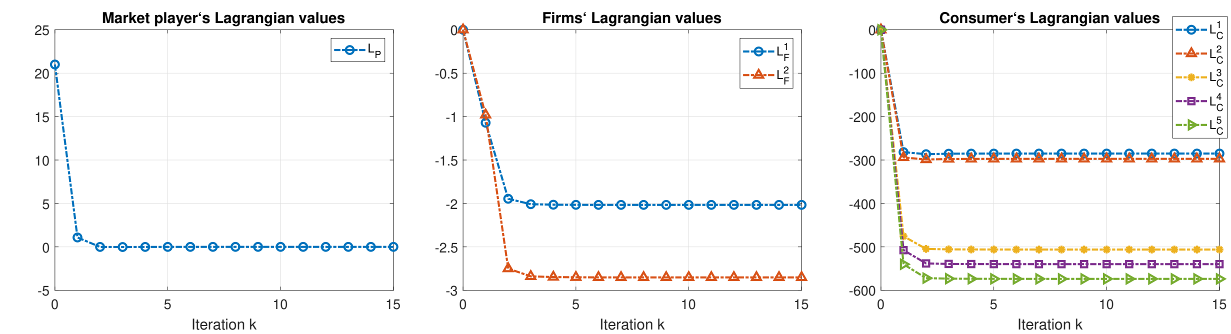
(a) Link 1 x^1_i (b) Link 3 x^3_i (c) Link 5 x^5_i

Figure 2 **Trajectories of the iterates of variables x^1_i, x^3_i, and x^5_i, $i = 1, \ldots, 8$, with sum-rates.**

set. The ith consumer decides on goods quantity $x^i \in X_i$ to maximize its utility, where $X_i \subseteq \mathbb{R}^K$ is a consumption set. The GNEP is defined as the set of problems of three types of players:

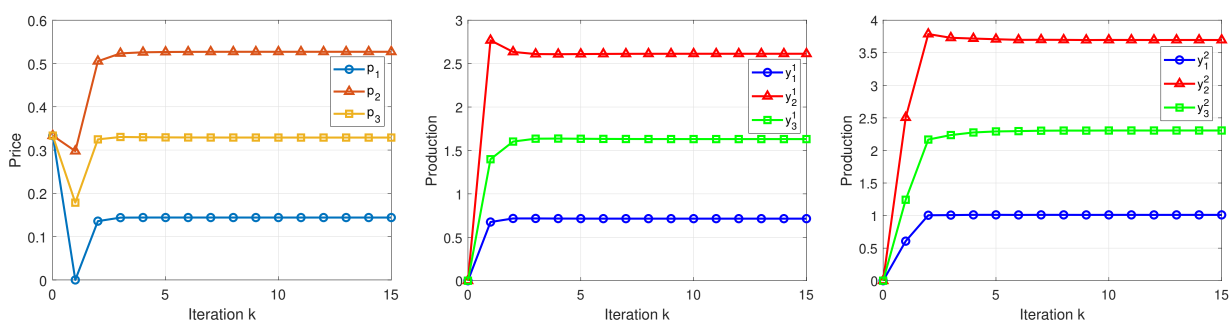
$$\begin{array}{llll} \max\limits_{y^j} & p^T y^j & \max\limits_{x^i \in X_i} & u_i(x^i) & \max\limits_{p} & p^T \left(\sum\limits_{i=1}^{I} x^i - \sum\limits_{j=1}^{J} y^j - \sum\limits_{i=1}^{I} \xi^i \right) \\ \text{s.t.} & y^j \in Y_j, & \text{s.t.} & p^T x^i \le p^T \left(\xi^i + \sum\limits_{j=1}^{J} q_{ij} y^j \right), & \text{s.t.} & \sum\limits_{k=1}^{K} p_k = 1, \quad p_k \ge 0, \end{array}$$

where $q_{ij} \ge 0$ is the fraction of the profit of the jth production owned by consumer i such that $\sum_{i=1}^{I} q_{ij} = 1$, and $\xi^i \in \mathbb{R}^K_+$ is an initial endowment of goods. The utility functions u_i are quadratic and concave, $u_i(x^i) = -\frac{1}{2}(x^i)^T Q^i x^i + (b^i)^T x^i$, and jth firm's production set is defined by $Y_j = \left\{ y^j \,\middle|\, y^j \geq 0,\ \sum_{k=1}^{K} (y_k^j)^2 \leq 10 \cdot j \right\}$. The detailed data is given in Facchinei and Kanzow (2009). Starting point is set to $x^{i,0} = \mathbf{0}$, $y^{j,0} = \mathbf{0}$, and $p^0 = (1/3, 1/3, 1/3)$. The convergence behaviors are shown in Figures 3 and 4. We see that the results also verify our theoretical findings. Figure 3 shows that all P-Lagrangian values are decreasing and convergent to finite values. Figure 4 illustrates that the iterates generated by Algorithm 1 converge to the equilibrium price $\overline{p} = (0.1441, 0.5270, 0.3289)$, as well as to the equilibrium production and consumption.

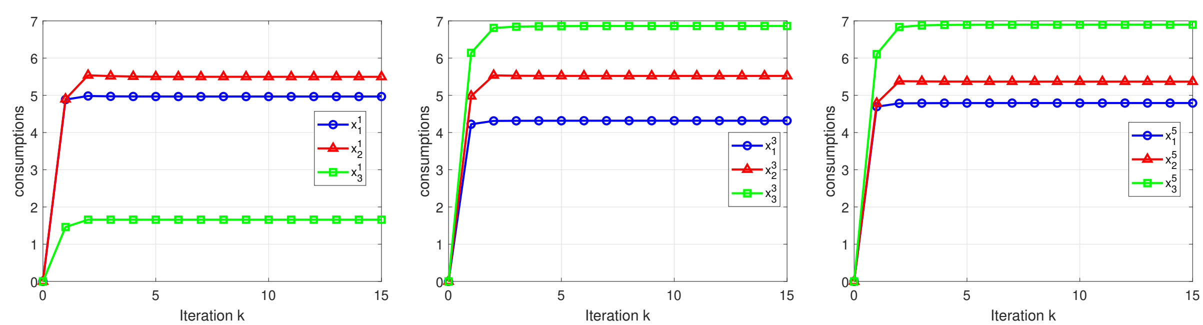


(a) Market player (b) Firms 1 and 2 (c) Consumers $1, \ldots, 5$

Figure 3 **Convergence of P-Lagrangian values.**



(a) Market player (b) Firm 1 (c) Firm 2



(d) Consumer 1 (e) Consumer 3 (f) Consumer 5

Figure 4 **Convergence of the sequence of decision variables for each player.**

6. Conclusions

In this paper, we proposed a novel algorithmic framework for computing an equilibrium of generalized continuous Nash games (GNEPs) with theoretical guarantees based on the Proximal-Perturbed Lagrangian function. We have shown that the proposed method has significant advantages over existing approaches from both theoretical and computational perspectives; it does not require boundedness assumptions and is the first development of an algorithm to solve *a general class of GNEPs* in a *distributed* manner. The numerical results supported our theoretical findings. Possible future research is to extend our methodology to compute equilibria in nonconvex games or stochastic games with coupling constraints that arise in economics and operations research, which will result in a broader application domain.

References

R. Andreani, E. G. Birgin, J. M. Martínez, and M. L. Schuverdt. On augmented Lagrangian methods with general lower-level constraints. *SIAM Journal on Optimization*, 18(4):1286–1309, 2007.

R. Andreani, E. G. Birgin, J. M. Martínez, and M. L. Schuverdt. Augmented Lagrangian methods under the constant positive linear dependence constraint qualification. *Mathematical Programming*, 111(1-2): 5–32, 2008.

K. J. Arrow and G. Debreu. Existence of an equilibrium for a competitive economy. *Econometrica*, 22: 265–290, 1954.

H. Attouch and J. Bolte. On the convergence of the proximal algorithm for nonsmooth functions involving analytic features. *Mathematical Programming*, 116(1-2):5–16, 2009.

H. Attouch, J. Bolte, and B. F. Svaiter. Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods. *Mathematical Programming*, 137(1-2):91–129, 2013.

Q. Ba and J.-S. Pang. Exact penalization of generalized Nash equilibrium problems. *Operations Research*, 70(3):1448–1464, 2022.

A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009.

G. Belgioioso, P. Yi, S. Grammatico, and L. Pavel. Distributed generalized Nash equilibrium seeking: An operator-theoretic perspective. *IEEE Control Systems Magazine*, 42(4):87–102, 2022.

D. P. Bertsekas. *Nonlinear programming.* Athena scientific, Belmont, MA, 1999.

D. P. Bertsekas. *Constrained optimization and Lagrange multiplier methods.* Academic press, 2014.

D. P. Bertsekas and J. N. Tsitsiklis. *Parallel and distributed computation: numerical methods*, volume 23. Prentice hall Englewood Cliffs, NJ, 1989.

E. G. Birgin and J. M. Martínez. *Practical augmented Lagrangian methods for constrained optimization.* SIAM, Philadelphia, PA, 2014.

J. Bolte, S. Sabach, and M. Teboulle. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Mathematical Programming*, 146(1-2):459–494, 2014.

M. Breton, G. Zaccour, and M. Zahaf. A game-theoretic formulation of joint implementation of environmental projects. *European Journal of Operational Research*, 168(1):221–239, 2006.

V. Cardellini, V. D. N. Personé, V. Di Valerio, F. Facchinei, V. Grassi, F. L. Presti, and V. Piccialli. A game-theoretic approach to computation offloading in mobile cloud computing. *Mathematical Programming*, 157(2):421–449, 2016.

C. Cenedese, G. Belgioioso, S. Grammatico, and M. Cao. An asynchronous distributed and scalable generalized nash equilibrium seeking algorithm for strongly monotone games. *European Journal of Control*, 58:143–151, 2021.

W. Chung and J. D. Fuller. Subproblem approximation in Dantzig-Wolfe decomposition of variational inequality models with an application to a multicommodity economic equilibrium model. *Operations Research*, 58(5):1318–1327, 2010.

J. Contreras, M. Klusch, and J. B. Krawczyk. Numerical solutions to Nash-Cournot equilibria in coupled constraint electricity markets. *IEEE Transactions on Power Systems*, 19(1):195–206, 2004.

G. Debreu. A social equilibrium existence theorem. *Proceedings of the National Academy of Sciences*, 38 (10):886–893, 1952.

Z. Deng. Distributed generalized Nash equilibrium seeking algorithm for nonsmooth aggregative games. *Automatica*, 132:109794, 2021.

D. Dorsch, H. T. Jongen, and V. Shikhman. On structure and computation of generalized nash equilibria. *SIAM Journal on Optimization*, 23(1):452–474, 2013.

A. Dreves, F. Facchinei, C. Kanzow, and S. Sagratella. On the solution of the KKT conditions of generalized Nash equilibrium problems. *SIAM Journal on Optimization*, 21(3):1082–1108, 2011.

F. Facchinei and C. Kanzow. Penalty methods for the solution of generalized Nash equilibrium problems (with complete test problems). *Sapienza University of Rome*, 2009.

F. Facchinei and C. Kanzow. Generalized Nash equilibrium problems. *Annals of Operations Research*, 175 (1):177–211, 2010a.

F. Facchinei and C. Kanzow. Penalty methods for the solution of generalized Nash equilibrium problems. *SIAM Journal on Optimization*, 20(5):2228–2253, 2010b.

F. Facchinei and J.-S. Pang. *Finite-dimensional variational inequalities and complementarity problems.* Springer Science & Business Media, 2007.

F. Facchinei, A. Fischer, and V. Piccialli. On generalized Nash games and variational inequalities. *Operations Research Letters*, 35(2):159–164, 2007.

F. Facchinei, A. Fischer, and V. Piccialli. Generalized Nash equilibrium problems and Newton methods. *Mathematical Programming*, 117(1-2):163–194, 2009.

S. Grammatico. Dynamic control of agents playing aggregative games with coupling constraints. *IEEE Transactions on Automatic Control*, 62(9):4537–4548, 2017.

P. T. Harker. Generalized Nash games and quasi-variational inequalities. *European journal of Operational research*, 54(1):81–94, 1991.

M. R. Hestenes. Multiplier and gradient methods. *Journal of optimization theory and applications*, 4(5): 303–320, 1969.

B. F. Hobbs and J.-S. Pang. Nash-Cournot equilibria in electric power markets with piecewise linear demand functions and joint constraints. *Operations Research*, 55(1):113–127, 2007.

A. F. Izmailov and M. V. Solodov. On error bounds and newton-type methods for generalized nash equilibrium problems. *Computational Optimization and Applications*, 59(1-2):201–218, 2014.

W. Jing-Yuan and Y. Smeers. Spatial oligopolistic electricity models with cournot generators and regulated transmission prices. *Operations Research*, 47(1):102–112, 1999.

C. Kanzow and D. Steck. Augmented Lagrangian methods for the solution of generalized Nash equilibrium problems. *SIAM Journal on Optimization*, 26(4):2034–2058, 2016.

C. Kanzow and D. Steck. Augmented lagrangian and exact penalty methods for quasi-variational inequalities. *Computational Optimization and Applications*, 69:801–824, 2018.

J. B. Krawczyk and S. Uryasev. Relaxation algorithms to find Nash equilibria with economic applications. *Environmental Modeling & Assessment*, 5(1):63–73, 2000.

A. A. Kulkarni and U. V. Shanbhag. On the variational equilibrium as a refinement of the generalized nash equilibrium. *Automatica*, 48(1):45–55, 2012.

K. Kurdyka. On gradients of functions definable in o-minimal structures. In *Annales de l'institut Fourier*, volume 48, pages 769–783, 1998.

G. Li and T. K. Pong. Calculus of the exponent of kurdyka–łojasiewicz inequality and its applications to linear convergence of first-order methods. *Foundations of computational mathematics*, 18(5):1199–1232, 2018.

S. Łojasiewicz. Une propriété topologique des sous-ensembles analytiques réels. *Les équations aux dérivées partielles*, 117:87–89, 1963.

J. P. Luna, C. Sagastizábal, and M. Solodov. A class of Dantzig–Wolfe type decomposition methods for variational inequality problems. *Mathematical Programming*, 143(1-2):177–209, 2014.

T. Migot and M.-G. Cojocaru. A parametrized variational inequality approach to track the solution set of a generalized Nash equilibrium problem. *European Journal of Operational Research*, 283(3):1136–1147, 2020.

T. Migot and M.-G. Cojocaru. A decomposition method for a class of convex generalized Nash equilibrium problems. *Optimization and Engineering*, 22:1653–1679, 2021.

K. Nabetani, P. Tseng, and M. Fukushima. Parametrized variational inequality approaches to generalized Nash equilibrium problems with shared constraints. *Computational Optimization and Applications*, 48 (3):423–452, 2011.

J. F. Nash. Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences*, 36(1): 48–49, 1950.

A. Nedic, A. Ozdaglar, and P. A. Parrilo. Constrained consensus and optimization in multi-agent networks. *IEEE Transactions on Automatic Control*, 55(4):922–938, 2010.

H. Nikaidô and K. Isoda. Note on non-cooperative convex games. 1955.

J. Nocedal and S. Wright. *Numerical optimization*. Springer Science & Business Media, 2006.

D. Paccagnan, B. Gentile, F. Parise, M. Kamgarpour, and J. Lygeros. Nash and Wardrop equilibria in aggregative games with coupling constraints. *IEEE Transactions on Automatic Control*, 64(4):1373–1388, 2018.

J.-S. Pang and M. Fukushima. Quasi-variational inequalities, generalized Nash equilibria, and multi-leader-follower games. *Computational Management Science*, 2(1):21–56, 2005.

J.-S. Pang, G. Scutari, F. Facchinei, and C. Wang. Distributed power allocation with rate constraints in Gaussian parallel interference channels. *IEEE Transactions on Information Theory*, 54(8):3471–3489, 2008.

M. J. Powell. A method for nonlinear constraints in minimization problems. *Optimization*, pages 283–298, 1969.

M. Razaviyayn, M. Hong, and Z.-Q. Luo. A unified convergence analysis of block successive minimization methods for nonsmooth optimization. *SIAM Journal on Optimization*, 23(2):1126–1153, 2013.

R. T. Rockafellar. Augmented Lagrange multiplier functions and duality in nonconvex programming. *SIAM Journal on Control*, 12(2):268–285, 1974.

J. B. Rosen. Existence and uniqueness of equilibrium points for concave n-person games. *Econometrica*, 33: 520–534, 1965.

D. A. Schiro, J.-S. Pang, and U. V. Shanbhag. On the solution of affine generalized nash equilibrium problems with shared constraints by lemke's method. *Mathematical Programming*, 142(1-2):1–46, 2013.

G. Scutari, F. Facchinei, J.-S. Pang, and D. P. Palomar. Real and complex monotone communication games. *IEEE Transactions on Information Theory*, 60(7):4197–4231, 2014.

G. Scutari, F. Facchinei, and L. Lampariello. Parallel and distributed methods for constrained nonconvex optimization—Part I: Theory. *IEEE Transactions on Signal Processing*, 65(8):1929–1944, 2016.

O. Stein and N. Sudermann-Merx. The noncooperative transportation problem and linear generalized nash games. *European Journal of Operational Research*, 266(2):543–553, 2018.

S. Uryas'ev and R. Y. Rubinstein. On relaxation algorithms in computation of noncooperative equilibria. *IEEE Transactions on Automatic Control*, 39(6):1263–1267, 1994.

A. Von Heusinger and C. Kanzow. Optimization reformulations of the generalized Nash equilibrium problem using Nikaido-Isoda-type functions. *Computational Optimization and Applications*, 43(3):353–377, 2009a.

A. Von Heusinger and C. Kanzow. Relaxation methods for generalized Nash equilibrium problems with inexact line search. *Journal of Optimization Theory and Applications*, 143(1):159–183, 2009b.

A. von Heusinger, C. Kanzow, and M. Fukushima. Newton's method for computing a normalized equilibrium in the generalized nash game through fixed point formulation. *Mathematical Programming*, 132(1-2): 99–123, 2012.

Y. Xu and W. Yin. A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. *SIAM Journal on Imaging Sciences*, 6(3):1758–1789, 2013.

P. Yi and L. Pavel. Distributed generalized Nash equilibria computation of monotone games via double-layer preconditioned proximal-point algorithms. *IEEE Transactions on Control of Network Systems*, 6(1): 299–311, 2018.

P. Yi and L. Pavel. An operator splitting approach for distributed generalized nash equilibria computation. *Automatica*, 102:111–121, 2019.

H. Yin, U. V. Shanbhag, and P. G. Mehta. Nash equilibrium problems with scaled congestion costs and shared constraints. *IEEE Transactions on Automatic Control*, 56(7):1702–1708, 2011.

M. Zhu and E. Frazzoli. Distributed robust adaptive equilibrium computation for generalized convex games. *Automatica*, 63:82–91, 2016.