Solving Multi-Follower Games

Lorenzo Lampariello[1], Simone Sagratella[2], Valerio Giuseppe Sasso[3], and Oliver Stein[4]

[1]Department of Business Studies, Roma Tre University, Italy, lorenzo.lampariello@uniroma3.it
[2]Department of Computer, Control and Management Engineering Antonio Ruberti, Sapienza University of Rome, Italy, sagratella@diag.uniroma1.it
[3]Department of Computer, Control and Management Engineering Antonio Ruberti, Sapienza University of Rome, Italy, sasso@diag.uniroma1.it
[4]Institute for Operations Research (IOR), Karlsruhe Institute of Technology (KIT), Germany, stein@kit.edu

October 18, 2023

Abstract

We consider bilevel programs where a single leader interacts with multiple followers who are coupled by a Nash equilibrium problem at the lower level. We generalize the value function reformulation to include multiple followers. This allows us to propose a convergent method based on the sequential convex approximation paradigm, and study the (exact or inexact) iterative solution of the convex subproblems. Since some of our convergence results require a constraint qualification, we give conditions under which it is satisfied. Finally, we propose a novel ESG-oriented multi-portfolio selection model, and test our numerical procedure confirming the theoretical insights.

Keywords: Nash equilibrium problem, bilevel problem, single-leader multi-follower game, sequential convex approximation, multi-portfolio selection

Mathematics Subject Classification: 90C33, 90C25, 90C30, 65K15, 91G10

1 Introduction

We deal with hierarchical programs where, at an upper level, the leader optimizes some criterion taking into account the choices of multiple followers who, at a lower level, aim at selfishly maximizing their utilities. Each follower's objective function, in turn, depends parametrically on both the leader's and the other followers' decisions: the collection of the followers' problems form a parametric (in the leader's decisions) Nash Equilibrium Problem (NEP). We assume the following conditions, which are standard in non-cooperative frameworks, to hold: the followers behave rationally and act simultaneously, and each follower's problem information is shared with the others. Following a common path in the relevant literature, we take the so-called (standard) optimistic point-of-view: we address a bilevel problem where the optimization is carried out with respect to both the leader's decision variables and the followers' ones, which, in turn, are constrained to belong to the lower-level parametric (in the leader's decisions) equilibrium set. We remark that the latter set is not assumed to be a singleton nor do we assume it to be expressed in closed form in terms of the leader's variables.

These structures are profitably used to describe real-world problems, for example in the context of multi-portfolio selection (see Section 7, [5]), energy markets (see e.g. [1]) and mobility-as-a-service (see e.g. [19] and the references therein). But, while the relevant literature has focused mostly on theoretical properties (see e.g. [4] and the references therein) and modelistic aspects, as far as numerical procedures are concerned, to date, few provably convergent approaches have been developed, essentially revolving around the classical Mathematical Program with Equilibrium Constraints (MPEC) formulation of the original problem ([6]). Note that, as highlighted in [7], local solutions of the MPEC formulation, which is what one can hope to compute at best, might not lead to local solution of the original problem.

As a main departure from the MPEC paradigm, we propose a generalized (to account for the presence of multiple followers) optimal-value function formulation, which has some precursors (e.g. [16, 21]) in the simpler framework of single-follower problems: we deal with the difficult-to-treat lower-level equilibrium constraint by introducing as many corresponding value function constraints as the numbers of the followers. We study the properties of the resulting problem under the assumption requiring the followers' objectives to be convex: we remark that convexity is to be intended with respect to *both* the leader's and all the followers' decision variables blocks. Such an assumption turns out to be not too restrictive, as we are able, under mild conditions, to enforce it a posteriori, by suitably modifying the (possibly nonconvex) followers' objectives, resulting in a problem that is nonetheless equivalent to the original one. We show that the followers' optimal-value functions are convex and continuously differentiable, so that we are able to devise a Sequential Convex Approximation (SCA)-like procedure to address the problem at hand.

The algorithm essentially consists in the iterative solution of a sequence of 'well-behaved' convex solvable subproblems. For the procedure to work properly, one needs first to compute a starting equilibrium for the lower-level problem, then one has to iteratively address each follower's individual convex and solvable optimization problem without the need to calculate further equilibria; subsequently, a leader-related strongly convex subproblem has to be solved, where inner convex (local) approximations of the value function constraints appear in lieu of the difficult original ones. We study the convergence properties of the resulting scheme both in the cases where some degree of inexactness is allowed for in the subproblems' solution and whenever they can be solved exactly. Finally, we equip our analysis with thorough numerical tests on real-world data in the multi-portfolio selection context.

Summarizing, as for the main contribution of our work,

- we generalize the value function formulation to account for multiple followers in a lower-level NEP framework. We extend the study on the value function's differentiability properties to account for inexactness in the solution of the followers' problem (see Section 2, and in particular Proposition 2.1);

- focusing on the convergence properties of a SCA-like algorithm tailored to the problem at hand we distinguish two cases. When subproblems are solved inexactly, the procedure is still guaranteed to produce feasible iterates and to achieve descent (see Section 4, and Theorem 4.4 for the behavior of the procedure). We prove in Section 5 convergence to stationary solutions of the original problem when subproblems are solved exactly (see Theorem 5.5).

We are not aware of other solution procedures that can guarantee similar outcomes under the mild conditions we assume. We remark that our procedure requires the computation of just a starting equilibrium, and as for the ensuing iterations, convex solvable subproblems are addressed;

- we study the regularity properties of the constraints mapping of the bilevel problem at hand. More precisely, we show that approaches, that are usually relied upon to recover standard Constraints Qualifications (CQ) when a single follower is present, do not ensure such CQs in the more general framework of multiple followers (see Section 6 for a thorough analysis);

- we present a genuine novel bilevel approach to deal with a nowadays highly sensitive topic: ESG-oriented multi-portfolio selection. We test our numerical procedure using real-world data in Section 7 and obtain results that confirm theoretical insights.

2 The multi-follower game: definitions and main properties

We consider the standard optimistic Multi-Follower ε-inexact Game

$$\begin{array}{cl} \underset{x,y}{\text{minimize}} & F(x,y) \\ \text{s.t.} & x \in X \\ & y \in E_\varepsilon(x), \end{array}$$

where $\varepsilon \in \mathbb{R}^N_+$ and $E_\varepsilon(x)$ is the (lower level) set of ε-inexact equilibria of the parametric (in x) Nash Equilibrium Problem (NEP), whose generic (follower) player ν, for $\nu = 1, \ldots, N$, solves the optimization problem

$$\begin{array}{cl} \underset{y^\nu}{\text{minimize}} & \theta_\nu(x, y^\nu, y^{-\nu}) \\ \text{s.t.} & y^\nu \in Y_\nu. \end{array} \tag{1}$$

We assume that

- the sets $X, Y_1, \ldots, Y_N$ are nonempty, compact and convex subsets of $\mathbb{R}^p, \mathbb{R}^{n_1}, \ldots, \mathbb{R}^{n_N}$, respectively,

- with $n = \sum_{\nu=1}^N n_\nu$ the functions $F, \theta_1, \ldots, \theta_N : \mathbb{R}^p \times \mathbb{R}^n \to \mathbb{R}$ are continuously differentiable and convex on $\mathbb{R}^p \times \mathbb{R}^n$ (we say that θ_ν is fully convex for every ν).

The notation $y = (y^\nu, y^{-\nu})$ with $y^{-\nu} = \left(y^1, \ldots, y^{\nu-1}, y^{\nu+1}, \ldots, y^N\right) \in \mathbb{R}^{n-n_\nu}$ emphasizes the ν-th player's decision variables, but does not reorder the entries of the vector $\left(y^1, \ldots, y^\nu, \ldots, y^N\right)$.

We remark that ε_ν represents inexactness in terms of the optimal value of player ν's problem. In particular, leveraging an optimal-value function reformulation-type approach, and observing that

$$E_\varepsilon(x) \triangleq \left\{ y \in \mathbb{R}^n \mid \theta_\nu(x, y^\nu, y^{-\nu}) - \varphi_\nu(x, y^{-\nu}) \leqslant \varepsilon_\nu, \, y^\nu \in Y_\nu, \, \nu = 1, \ldots, N \right\},$$

where

$$\varphi_\nu(x, y^{-\nu}) \triangleq \min_{y^\nu \in Y_\nu} \theta_\nu(x, y^\nu, y^{-\nu}),$$

one can equivalently replace the constraint $y \in E_\varepsilon(x)$ in the formulation of the standard optimistic Multi-Follower ε-inexact Game with the functional expression in the formula above, yielding the problem

$$
\begin{array}{lll}
\underset{x,y}{\text{minimize}} & F(x,y) & \\
\text{s.t.} & x \in X, \;\; y^\nu \in Y_\nu, \;\; \nu = 1,\ldots,N, & (\text{MFG}_\varepsilon) \\
& \theta_\nu(x,y^\nu,y^{-\nu}) - \varphi_\nu(x,y^{-\nu}) \leqslant \varepsilon_\nu, \; \nu = 1,\ldots,N. &
\end{array}
$$

We define $Y \triangleq \prod_{\nu=1}^N Y_\nu$, and we indicate with W_ε the feasible set of (MFG$_\varepsilon$):

$$W_\varepsilon \triangleq \{(x,y) \in X \times Y \mid \theta_\nu(x,y^\nu,y^{-\nu}) - \varphi_\nu(x,y^{-\nu}) \leqslant \varepsilon_\nu, \; \nu = 1,\ldots,N\}.$$

We also denote, for $\eta \geqslant 0$ and for every ν, by

$$S_\nu^\eta(x,y^{-\nu}) \triangleq \left\{y^\nu \in Y_\nu \mid \theta_\nu(x,y^\nu,y^{-\nu}) - \varphi_\nu(x,y^{-\nu}) \leqslant \eta\right\}$$

the set of η-inexact optimal points for player ν's (parametric) optimization problem (1). Focusing on the solution of follower ν's problem (1), we use η to denote an input level of inexactness allowed for, whereas ε_ν refers to the inexactness guaranteed after different iterations of our proposed algorithmic framework.

For results where a positive level of inexactness is considered, the following further assumptions are called for every ν.

A1 $\begin{pmatrix} \nabla_x \theta_\nu(x,\bullet,y^{-\nu}) \\ \nabla_{y^{-\nu}} \theta_\nu(x,\bullet,y^{-\nu}) \end{pmatrix}$ is uniformly Lipschitz continuous for every x and $y^{-\nu}$, with modulus L independent of x, $y^{-\nu}$ and ν;

A2 $\theta_\nu(x,\bullet,y^{-\nu})$ is uniformly strongly convex for every x and $y^{-\nu}$ with constant c independent of x, $y^{-\nu}$ and ν.

Note that, while, under A2, S_ν^0 reduces to a single-valued mapping for every x and $y^{-\nu}$, this may not be the case for E_0, let alone for E_ε with $\varepsilon > 0$.

The following result gathers important properties of the optimal-value function φ_ν.

Proposition 2.1 *For every ν, φ_ν is convex and continuously differentiable on $\mathbb{R}^p \times \mathbb{R}^{n-n_\nu}$ and the following statements hold*

(i)

$$\nabla \varphi_\nu(x,y^{-\nu}) = \begin{pmatrix} \nabla_x \varphi_\nu(x,y^{-\nu}) \\ \nabla_{y^{-\nu}} \varphi_\nu(x,y^{-\nu}) \end{pmatrix} = \begin{pmatrix} \nabla_x \theta_\nu(x,w^\nu,y^{-\nu}) \\ \nabla_{y^{-\nu}} \theta_\nu(x,w^\nu,y^{-\nu}) \end{pmatrix}, \quad \forall w^\nu \in S_\nu^0(x,y^{-\nu});$$

(ii) assume A1 and A2 to hold and $\eta \geqslant 0$:

$$\left\| \begin{pmatrix} \nabla_x \varphi_\nu(x,y^{-\nu}) \\ \nabla_{y^{-\nu}} \varphi_\nu(x,y^{-\nu}) \end{pmatrix} - \begin{pmatrix} \nabla_x \theta_\nu(x,z^\nu,y^{-\nu}) \\ \nabla_{y^{-\nu}} \theta_\nu(x,z^\nu,y^{-\nu}) \end{pmatrix} \right\| \leqslant \frac{L}{\sqrt{c/2}} \sqrt{\eta}, \quad \forall z^\nu \in S_\nu^\eta(x,y^{-\nu}).$$

Proof. Problem (1) can be equivalently reformulated as an unconstrained program employing the so-called indicator function $\delta_{Y^\nu}(y^\nu)$, where $\delta_{Y^\nu}(y^\nu) = 0$ if $y^\nu \in Y^\nu$ and $\delta_{Y^\nu}(y^\nu) = \infty$ if $y^\nu \notin Y^\nu$. Since, by our assumptions, the function δ_{Y^ν} turns out to be lower semicontinuous and convex, then the convexity of φ_ν follows from [17, Corollary 3.32], since θ_ν is fully convex. In order to show the continuous differentiability of φ_ν and the expression of its gradient stated in point *(i)*, one can rely on, e.g., [6, 18]. However, to maintain the paper self-contained, we prove that, for every $(\overline{x}, \overline{y}^{-\nu})$ and any $w^\nu \in S^0_\nu(\overline{x}, \overline{y}^{-\nu})$, $\nabla_x \theta_\nu(\overline{x}, w^\nu, \overline{y}^{-\nu}) = \overline{u}$, $\nabla_{y^{-\nu}} \theta_\nu(\overline{x}, w^\nu, \overline{y}^{-\nu}) = \overline{v}$ and, as a consequence, $\partial \varphi_\nu(\overline{x}, \overline{y}^{-\nu}) = \{(\nabla_x \theta_\nu(\overline{x}, w^\nu, \overline{y}^{-\nu}), \nabla_{y^{-\nu}} \theta_\nu(\overline{x}, w^\nu, \overline{y}^{-\nu})) : w^\nu \in S^0_\nu(\overline{x}, \overline{y}^{-\nu})\} = \{(\overline{u}, \overline{v})\}$.

Let us consider a generic $\mathbb{R}^p \times \mathbb{R}^{n-n_\nu} \ni (u, v) \in \partial \varphi_\nu(\overline{x}, \overline{y}^{-\nu})$. By the convexity of φ_ν, we have

$$\begin{aligned} \theta_\nu(x, y^\nu, y^{-\nu}) \geqslant \varphi_\nu(x, y^{-\nu}) &\geqslant \varphi_\nu(\overline{x}, \overline{y}^{-\nu}) + u^T(x - \overline{x}) + v^T(y^{-\nu} - \overline{y}^{-\nu}) \\ &= \theta_\nu(\overline{x}, w^\nu, \overline{y}^{-\nu}) + u^T(x - \overline{x}) + v^T(y^{-\nu} - \overline{y}^{-\nu}), \end{aligned}$$

for every $(x, y^\nu, y^{-\nu}) \in \mathbb{R}^p \times Y_\nu \times \mathbb{R}^{n-n_\nu}$. It follows that

$$(\overline{x}, w^\nu, \overline{y}^{-\nu}) \in \arg\min_{x, y^\nu, y^{-\nu}} \theta_\nu(x, y^\nu, y^{-\nu}) - u^T x - v^T y^{-\nu} \quad \text{s.t.} \quad y^\nu \in Y_\nu,$$

and, in turn, $0 = \nabla_x \theta_\nu(\overline{x}, w^\nu, \overline{y}^{-\nu}) - u$ and $0 = \nabla_{y^{-\nu}} \theta_\nu(\overline{x}, w^\nu, \overline{y}^{-\nu}) - v$. The claim is a consequence of the arbitrariness of (u, v). Finally, the thesis follows by [17, Theorem 9.18, Corollary 9.20].

As for point *(ii)*, for every $(x, y^{-\nu})$ and any $w^\nu \in S^0_\nu(x, y^{-\nu})$,

$$\frac{c}{2}\|z^\nu - w^\nu\|^2 \leqslant \nabla_{y^\nu} \theta_\nu(x, w^\nu, y^{-\nu})^\top (z^\nu - w^\nu) + \frac{c}{2}\|z^\nu - w^\nu\|^2 \leqslant \theta_\nu(x, z^\nu, y^{-\nu}) - \theta_\nu(x, w^\nu, y^{-\nu}) \leqslant \eta,$$

where the first inequality is due to $w^\nu \in S_\nu(x, y^{-\nu})$, the second relation follows from A2, while the last one holds because $z^\nu \in S^\eta_\nu(x, y^{-\nu})$. Thanks to point *(i)*, in view of A1, we have

$$\begin{aligned} \left\| \begin{pmatrix} \nabla_x \varphi_\nu(x, y^{-\nu}) \\ \nabla_{y^{-\nu}} \varphi_\nu(x, y^{-\nu}) \end{pmatrix} - \begin{pmatrix} \nabla_x \theta_\nu(x, z^\nu, y^{-\nu}) \\ \nabla_{y^{-\nu}} \theta_\nu(x, z^\nu, y^{-\nu}) \end{pmatrix} \right\| &= \left\| \begin{pmatrix} \nabla_x \theta_\nu(x, w^\nu, y^{-\nu}) \\ \nabla_{y^{-\nu}} \theta_\nu(x, w^\nu, y^{-\nu}) \end{pmatrix} \right. \\ \left. - \begin{pmatrix} \nabla_x \theta_\nu(x, z^\nu, y^{-\nu}) \\ \nabla_{y^{-\nu}} \theta_\nu(x, z^\nu, y^{-\nu}) \end{pmatrix} \right\| &\leqslant L\|w^\nu - z^\nu\| \leqslant \frac{L}{\sqrt{c/2}}\sqrt{\eta}. \end{aligned}$$

□

Proposition 2.1 implies in particular that the left hand sides $\theta_\nu(x, y^\nu, y^{-\nu}) - \varphi_\nu(x, y^{-\nu})$ of the inequality constraints in (MFG$_\varepsilon$) constitute differences of convex (DC) functions, while the objective function and the appearing sets are convex.

Moreover, Proposition 2.1 indicates that knowing (inexact) solutions to player ν's problem (1) is key to evaluate $\nabla \varphi_\nu$. Clearly, if $\eta > 0$, and thus solutions to (1) are computed inexactly, nonetheless $(\nabla_x \theta_\nu(x, w^\nu, y^{-\nu}), \nabla_{y^{-\nu}} \theta_\nu(x, w^\nu, y^{-\nu}))$ gives an estimate of $\nabla \varphi_\nu(x, y^{-\nu})$ up to an error depending on η and simple problem-related constants.

Note that under our assumptions not only each set $E_\varepsilon(x)$, for every $x \in X$, is nonempty and compact [11, Corollary 2.2.5], but by the continuity of the functions φ_ν, also the graph of the set-valued mapping E_ε on X is nonempty and compact.

We conclude the section with a technical result concerning the outer semicontinuity of the set-valued mapping $W_\bullet$, that is $W : \mathbb{R}^N \rightrightarrows \mathbb{R}^p \times \mathbb{R}^n$.

Proposition 2.2 *W_ε, considered as a set-valued mapping with respect to ε, is outer semi-continuous on $\mathbb{R}^n_+$, relative to $\mathbb{R}^n_+$.*

Proof. See Proposition 2.1 and [2, Theorem 3.1.1]. □

In the forthcoming developements we assume that ∇F is Lipschitz continuous on the compact set $X \times Y$ with modulus $L_{\nabla F}$, and we define the following quantities:

$$H \triangleq \max_{(x,y)\in X\times Y} \|\nabla F(x,y)\|, \qquad D \triangleq \max_{u,\, v \in X\times Y} \|u - v\|.$$

3 An algorithmic procedure

We rely on the Sequential Convex Approximation (SCA) paradigm (see [13] and the references therein for some context about this class of algorithmic procedures) to address (MFG$_\varepsilon$): given a base point (x^k, y^k, z^k), where for every ν, $z^{\nu,k} \in S^\eta_\nu(x^k, y^{-\nu,k})$, the defining subproblem for the algorithm's inner iterations results from a regularization of the objective function, linearizations of the concave parts of the DC constraint functions, and the introduction of iteration dependent feasibility tolerances. It reads as follows:

$$\begin{array}{ll} \underset{x,y}{\text{minimize}} & F(x,y) + \frac{\tau}{2}\|(x,y) - (x^k, y^k)\|_2^2 \\ \text{s.t.} & x \in X,\ \ y^\nu \in Y_\nu,\ \nu = 1,\dots,N, \\ & \theta_\nu(x, y^\nu, y^{-\nu}) - \theta_\nu(x^k, z^{\nu,k}, y^{-\nu,k}) - \nabla_x \theta_\nu(x^k, z^{\nu,k}, y^{-\nu,k})^T(x - x^k) \\ & \quad -\nabla_{y^{-\nu}}\theta_\nu(x^k, z^{\nu,k}, y^{-\nu,k})^T(y^{-\nu} - y^{-\nu,k}) - (\zeta^k_\nu + \overline{\zeta}_\nu) \leqslant 0,\ \nu = 1,\dots,N, \end{array} \tag{P$_{\zeta^k}(x^k,y^k,z^k)$}$$

where $\overline{\zeta} \in \mathbb{R}^N_+$ and $\zeta^k \in \mathbb{R}^N$ are two vectors aiming at relaxing the convexified versions of the N value function constraints $\theta_\nu(x, y^\nu, y^{-\nu}) - \varphi_\nu(x, y^{-\nu}) \leqslant 0$ such that

$$\zeta^k_\nu \triangleq \min\left\{\max\left\{\theta_\nu(x^k, y^k) - \theta_\nu(x^k, z^{\nu,k}, y^{-\nu,k}), 0\right\}, k\left(\eta + \frac{LD}{\sqrt{c/2}}\sqrt{\eta} + \eta\overline{\zeta}_\nu\right)\right\},\ \nu = 1,\dots,N. \tag{2}$$

The term $(\zeta^k_\nu + \overline{\zeta}_\nu)$ is a given tolerance, at iteration k, for the fulfilment of the convexified version of the ν-th value function constraint. In fact, ζ^k_ν is instrumental to compensate the errors yielded by η in the evaluation of $\varphi_\nu(x^k, y^{-\nu,k})$ and $\nabla\varphi_\nu(x^k, y^{-\nu,k})$ (see Proposition 4.1 (i)). Clearly, if $\eta = 0$, and thus the estimates of $\varphi_\nu(x^k, y^{-\nu,k})$ and $\nabla\varphi_\nu(x^k, y^{-\nu,k})$ are exact, then $\zeta^k_\nu = 0$ for every ν. As for $\overline{\zeta}_\nu$, it is related to the fulfilment of constraint qualifications, that play an important role for numerical reasons (see Proposition 4.1 (ii) and the results in Section 6). Notice that, under the initial assumptions, the sequence $\{\zeta^k\}$ is bounded since the difference $\theta_\nu(x^k, y^k) - \theta_\nu(x^k, z^{\nu,k}, y^{-\nu,k})$ in (2) is bounded for every ν.

We denote by

$$\begin{aligned} \mathbb{R}^p \times \mathbb{R}^n \supseteq\ & C_{\zeta^k}(x^k, y^k, z^k) \\ \triangleq\ & \{x \in X,\ \ y^\nu \in Y_\nu,\ \nu = 1,\dots,N \mid \\ & \theta_\nu(x, y^\nu, y^{-\nu}) - \theta_\nu(x^k, z^{\nu,k}, y^{-\nu,k}) - \nabla_x\theta_\nu(x^k, z^{\nu,k}, y^{-\nu,k})^T(x - x^k) \\ & -\nabla_{y^{-\nu}}\theta_\nu(x^k, z^{\nu,k}, y^{-\nu,k})^T(y^{-\nu} - y^{-\nu,k}) - (\zeta^k_\nu + \overline{\zeta}_\nu) \leqslant 0,\ \nu = 1,\dots,N\} \end{aligned}$$

the convex feasible set of problem $(\mathrm{P}_{\zeta^k}(x^k, y^k, z^k))$.

From now on, we assume computational solvability for

- Lower level NEP: equilibria can be computed up to a tolerance $\breve{\eta} \in \mathbb{R}^N_+$, that is, given $x^0 \in X$, $y^0 \in E_{\breve{\eta}}(x^0)$ can be calculated;
- Lower level problems: followers' problems are solvable within a tolerance $\eta \geqslant 0$, thus, points $z^{\nu,k}$ in $S^{\eta}_{\nu}(x^k, y^{-\nu,k})$ can be computed;
- Upper level subproblems: letting $\overline{\eta}^k \geqslant \widehat{\eta} \geqslant 0$, points (x^{k+1}, y^{k+1}) satisfying the $\overline{\eta}^k$-approximate minimum principle for $(\mathrm{P}_{\zeta^k}(x^k, y^k, z^k))$ can be computed, i.e.

$$\begin{aligned}&(x^{k+1}, y^{k+1}) \in C_{\zeta^k}(x^k, y^k, z^k),\\ &\left[\nabla F(x^{k+1}, y^{k+1}) + \tau \begin{pmatrix} x^{k+1} - x^k \\ y^{k+1} - y^k \end{pmatrix}\right]^\top \begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \end{pmatrix} \geqslant -\overline{\eta}^k, \quad \forall\, (x, y) \in C_{\zeta^k}(x^k, y^k, z^k).\end{aligned} \tag{3}$$

  From a practical point of view, it is reasonable to assume $\widehat{\eta} \geqslant \eta$.

The following scheme summarizes the SCA-like procedure described above.

Algorithm 1: Alternating optimization

Data: $\overline{\zeta} \geqslant 0$, $\eta \geqslant 0$, $\{\overline{\eta}^k\} \geqslant \widehat{\eta} \geqslant 0$, $\breve{\eta} \geqslant 0$, $\tau \geqslant 0$, $x^0 \in X$;
(S.0) Compute $y^0 \in E_{\breve{\eta}}(x^0)$;
for $k = 0, 1, \ldots$ **do**
(S.1) | **for** $\nu = 1, \ldots, N$ **do**
| | Compute $z^{\nu,k} \in S^{\eta}_{\nu}(x^k, y^{-\nu,k})$;
| **end**
(S.2) | Update ζ^k according to (2) and compute (x^{k+1}, y^{k+1}) satisfying (3);
end

Step (S.0) requires the computation of an appoximate equilibrium of the lower level NEP, given x^0: one can rely on many provably convergent methods that are available in the relevant literature. For example, whenever $[\nabla_{y^\nu} \theta_\nu(x^0, y)]_{\nu=1}^N$ is monotone on Y, efficient solution procedures are readily at hand [8, 9, 10, 12]. Performing this preliminary calculation is instrumental to make (x^0, y^0) feasible for $(\mathrm{MFG}_\varepsilon)$, where $\varepsilon = \breve{\eta}$, and, in turn, to maintain feasibility throughout the iterations (see Propositions 4.1 and 5.3).

Concerning (S.1), for each k, one has to solve, up to an accuracy η, N separate followers-related optimization problems (1), where parameters $(x, y^{-\nu}) = (x^k, y^{-\nu,k})$ are fixed. Such N optimization problems are solvable and convex: a host of approaches can be relied upon to address them, possibly in a finite number of iterations when $\eta > 0$. The computed points $z^{\nu,k}$ are, then, employed to obtain the estimates $\theta_\nu(x^k, z^{\nu,k}, y^{-\nu,k})$ for $\varphi_\nu(x^k, y^{-\nu,k})$ and $(\nabla_x \theta_\nu(x^k, z^{\nu,k}, y^{-\nu,k}), \nabla_{y^{-\nu}} \theta_\nu(x^k, z^{\nu,k}, y^{-\nu,k}))$ for $\nabla \varphi_\nu(x^k, y^{-\nu,k})$, that are used to build subproblem $\mathrm{P}_{\zeta^k}(x^k, y^k, z^k)$.

Finally, in (S.2), (x^{k+1}, y^{k+1}) is calculated as an inexact solution to the approximate first order necessary and sufficient optimality conditions (3) for the strongly convex subproblem $\mathrm{P}_{\zeta^k}(x^k, y^k, z^k)$. Under some assumptions, the latter problem is shown to satisfy the Slater constraint qualification. In order to have an insight on inexactness in variational contexts, we refer the reader to [3].

In the forthcoming sections, we analyze the convergence properties of Algorithm 1 when inexactness in the solution of followers' problems is accounted for ($\eta > 0$), and in the exact case ($\eta = 0$).

4 The case when the followers' subproblems are solved inexactly ($\boldsymbol{\eta} > \mathbf{0}$)

With the following proposition, we state two important properties: in point *(i)* we show that any feasible point for any subproblem is feasible for the original multi-follower game considering a suitable inexactness value. Point *(ii)* states that the subproblems are feasible and even more, they enjoy constraint qualifications, making them numerically tractable. For this to be true, the initial point y^0 must be in $E_{\breve{\eta}}(x^0)$.

Proposition 4.1 *Let $\eta > 0$. Under Assumptions A1 and A2:*

(i) if a point (x, y) is feasible for $(\mathrm{P}_{\zeta^k}(x^k, y^k, z^k))$, with any arbitrarily chosen $z^{\nu,k} \in S_\nu^\eta(x^k, y^{-\nu,k})$ $\forall \nu$, then (x, y) is feasible also for $(\mathrm{MFG}_{\varepsilon^k})$, with $\varepsilon_\nu^k \triangleq \eta + \frac{LD}{\sqrt{c/2}}\sqrt{\eta} + \zeta_\nu^k + \overline{\zeta}_\nu, \quad \nu = 1, \ldots, N$;

(ii) if $\overline{\zeta} \geqslant \breve{\eta}$, the pair (x^k, y^k) is feasible for $(\mathrm{P}_{\zeta^k}(x^k, y^k, z^k))$ for any $z^{\nu,k} \in S_\nu^\eta(x^k, y^{-\nu,k})$ $\forall \nu$. If $\overline{\zeta} > \breve{\eta}$, the pair $(x^k, y^k) \in X \times Y$ is a Slater point for $C_{\zeta^k}(x^k, y^k, z^k)$.

Proof. (i) From the feasibility of (x, y) for *$(P_{\zeta^k}(x^k, y^k, z^k))$* we have

$$\begin{aligned} \theta_\nu(x, y^\nu, y^{-\nu}) &\leqslant \theta_\nu(x^k, z^{\nu,k}, y^{-\nu,k}) + \nabla_x \theta_\nu(x^k, z^{\nu,k}, y^{-\nu,k})^T (x - x^k) \\ &\quad + \nabla_{y^{-\nu}} \theta_\nu(x^k, z^{\nu,k}, y^{-\nu,k})^T (y^{-\nu} - y^{-\nu,k}) + \zeta_\nu^k + \overline{\zeta}_\nu, \ \nu = 1, \ldots, N. \end{aligned} \tag{4}$$

Proposition 2.1 entails with any $w^{\nu,k} \in S_\nu^0(x^k, y^{-\nu,k})$

$$\begin{aligned} &\theta_\nu(x^k, z^{\nu,k}, y^{-\nu,k}) + \nabla_x \theta_\nu(x^k, z^{\nu,k}, y^{-\nu,k})^T (x - x^k) + \nabla_{y^{-\nu}} \theta_\nu(x^k, z^{\nu,k}, y^{-\nu,k})^T (y^{-\nu} - y^{-\nu,k}) \\ &= \theta_\nu(x^k, z^{\nu,k}, y^{-\nu,k}) + \nabla_x \theta_\nu(x^k, w^{\nu,k}, y^{-\nu,k})^T (x - x^k) + \nabla_{y^{-\nu}} \theta_\nu(x^k, w^{\nu,k}, y^{-\nu,k})^T (y^{-\nu} - y^{-\nu,k}) \\ &+ [\nabla_x \theta_\nu(x^k, z^{\nu,k}, y^{-\nu,k}) - \nabla_x \theta_\nu(x^k, w^{\nu,k}, y^{-\nu,k})]^T (x - x^k) \\ &+ [\nabla_{y^{-\nu}} \theta_\nu(x^k, z^{\nu,k}, y^{-\nu,k}) - \nabla_{y^{-\nu}} \theta_\nu(x^k, w^{\nu,k}, y^{-\nu,k})]^T (y^{-\nu} - y^{-\nu,k}) \\ &\leqslant \theta_\nu(x^k, z^{\nu,k}, y^{-\nu,k}) + \nabla_x \theta_\nu(x^k, w^{\nu,k}, y^{-\nu,k})^T (x - x^k) + \nabla_{y^{-\nu}} \theta_\nu(x^k, w^{\nu,k}, y^{-\nu,k})^T (y^{-\nu} - y^{-\nu,k}) \\ &+ \left\| \begin{pmatrix} \nabla_x \theta_\nu(x^k, z^{\nu,k}, y^{-\nu,k}) - \nabla_x \theta_\nu(x^k, w^{\nu,k}, y^{-\nu,k}) \\ \nabla_{y^{-\nu}} \theta_\nu(x^k, z^{\nu,k}, y^{-\nu,k}) - \nabla_{y^{-\nu}} \theta_\nu(x^k, w^{\nu,k}, y^{-\nu,k}) \end{pmatrix} \right\| \left\| \begin{pmatrix} x - x^k \\ y^{-\nu} - y^{-\nu,k} \end{pmatrix} \right\| \\ &\leqslant \varphi_\nu(x^k, y^{-\nu,k}) + \eta + \nabla_x \varphi_\nu(x^k, y^{-\nu,k})^T (x - x^k) + \nabla_{y^{-\nu}} \varphi_\nu(x^k, y^{-\nu,k})^T (y^{-\nu} - y^{-\nu,k}) + \frac{LD}{\sqrt{c/2}}\sqrt{\eta} \\ &\leqslant \varphi_\nu(x, y^{-\nu}) + \eta + \frac{LD}{\sqrt{c/2}}\sqrt{\eta}. \end{aligned}$$

In turn, by (4) we obtain

$$\theta_\nu(x, y^\nu, y^{-\nu}) \leqslant \varphi_\nu(x, y^{-\nu}) + \eta + \frac{LD}{\sqrt{c/2}}\sqrt{\eta} + \zeta_\nu^k + \overline{\zeta}_\nu,\ \nu = 1, \ldots, N.$$

The claim follows observing that $x \in X$ and $y^\nu \in Y_\nu$, $\forall \nu$.

(ii) Let $\overline{\zeta} \geqslant \breve{\eta}$. We distinguish two cases. If $\zeta_\nu^k = \max\{\theta_\nu(x^k, y^k) - \theta_\nu(x^k, z^{\nu,k}, y^{-\nu,k}), 0\}$, then

$$\begin{aligned}\theta_\nu(x^k, y^k) - \theta_\nu(x^k, z^{\nu,k}, y^{-\nu,k}) - \nabla_x \theta_\nu(x^k, z^{\nu,k}, y^{-\nu,k})^T(x^k - x^k) \\ - \nabla_{y^{-\nu}} \theta_\nu(x^k, z^{\nu,k}, y^{-\nu,k})^T(y^{-\nu,k} - y^{-\nu,k}) \leqslant \zeta_\nu^k \leqslant \zeta_\nu^k + \overline{\zeta}_\nu.\end{aligned}$$

Otherwise, if $k = 0$

$$\begin{aligned}&\theta_\nu(x^k, y^k) - \theta_\nu(x^k, z^{\nu,k}, y^{-\nu,k}) - \nabla_x \theta_\nu(x^k, z^{\nu,k}, y^{-\nu,k})^T(x^k - x^k) \\ &\quad - \nabla_{y^{-\nu}} \theta_\nu(x^k, z^{\nu,k}, y^{-\nu,k})^T(y^{-\nu,k} - y^{-\nu,k}) \\ &\quad \leqslant \theta_\nu(x^k, y^k) - \varphi_\nu(x^k, y^{-\nu,k}) \leqslant \breve{\eta}_\nu \leqslant \overline{\zeta}_\nu = \zeta_\nu^k + \overline{\zeta}_\nu\end{aligned}$$

where the second relation is due to $y^0 \in E_{\breve{\eta}}(x^0)$, and the last equation to (2). If $k \geqslant 1$

$$\begin{aligned}&\theta_\nu(x^k, y^k) - \theta_\nu(x^k, z^{\nu,k}, y^{-\nu,k}) - \nabla_x \theta_\nu(x^k, z^{\nu,k}, y^{-\nu,k})^T(x^k - x^k) \\ &\quad - \nabla_{y^{-\nu}} \theta_\nu(x^k, z^{\nu,k}, y^{-\nu,k})^T(y^{-\nu,k} - y^{-\nu,k}) \\ &\quad \leqslant \theta_\nu(x^k, y^k) - \varphi_\nu(x^k, y^{-\nu,k}) \leqslant \eta + \frac{LD}{\sqrt{c/2}}\sqrt{\eta} + \zeta_\nu^{k-1} + \overline{\zeta}_\nu \\ &\quad \leqslant k\left(\eta + \frac{LD}{\sqrt{c/2}}\sqrt{\eta} + \eta\overline{\zeta}_\nu\right) + (1-\eta)\overline{\zeta}_\nu = \zeta_\nu^k + (1-\eta)\overline{\zeta}_\nu \leqslant \zeta_\nu^k + \overline{\zeta}_\nu, \qquad (5)\end{aligned}$$

where the second relation is due to point (i) and $(x^k, y^k) \in C_{\zeta^{k-1}}(x^{k-1}, y^{k-1}, z^{k-1})$; the third relation follows from (2) for ζ_ν^{k-1}, and we have the first claim.

If $\overline{\zeta} > \breve{\eta}$, the last inequality in chains of point *(ii)* holds strictly. □

The following technical lemmas are instrumental to prove the results in the convergence Theorem 4.4. The claim in Lemma 4.2 deals with the continuity properties of the point-to-set mapping C: while the outer semicontinuity of C can be easily proven resorting to standard reasonings, some clarifications are in order as for its inner semicontinuity.

Lemma 4.2 *Let $\eta > 0$ and $\overline{\zeta} > 0$. Let (x^k, y^k, z^k, ζ^k) be the sequence generated by Algorithm 1. The point-to-set mapping $C : \mathbb{R}^p \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^N \rightrightarrows \mathbb{R}^p \times \mathbb{R}^n$ is continuous at any limit point $(\widehat{x}, \widehat{y}, \widehat{z}, \widehat{\zeta})$ of the sequence, relative to $X \times Y \times Y \times \mathbb{R}_+$.*

Proof. The outer semicontinuity can be easily proven resorting to standard reasoning [2, Theorem 3.1.1]. We only need to prove the inner semicontinuity. The following properties that, by [2, Theorem 3.1.6], are sufficient for the claim to be true, hold:

- sets X and Y are nonempty and convex;

- Since $\eta > 0$, (2) eventually yields $\zeta_\nu^k = \max\{\theta_\nu(x^k, y^k) - \theta_\nu(x^k, z^{\nu,k}, y^{-\nu,k}), 0\}$, and therefore $\widehat{\zeta}_\nu = \max\{\theta_\nu(\widehat{x}, \widehat{y}) - \theta_\nu(\widehat{x}, \widehat{z}^\nu, \widehat{y}^{-\nu}), 0\}$. Then, for all ν, we have

  $$\theta_\nu(\widehat{x}, \widehat{y}) - \theta_\nu(\widehat{x}, \widehat{z}^\nu, \widehat{y}^{-\nu}) - \nabla_x \theta_\nu(\widehat{x}, \widehat{z}^\nu, \widehat{y}^{-\nu})^T(\widehat{x} - \widehat{x}) - \nabla_{y^{-\nu}} \theta_\nu(\widehat{x}, \widehat{z}^\nu, \widehat{y}^{-\nu})^T(\widehat{y}^{-\nu} - \widehat{y}^{-\nu})$$
  $$\leqslant \widehat{\zeta}_\nu < \widehat{\zeta}_\nu + \overline{\zeta}_\nu,$$

  that is, $(\widehat{x}, \widehat{y}, \widehat{z}, \widehat{\zeta})$ is a Slater point for $C_{\widehat{\zeta}}(\widehat{x}, \widehat{y}, \widehat{z})$;

- for every ν, the function $\theta_\nu(x, y^\nu, y^{-\nu}) - \theta_\nu(\widehat{x}, \widehat{z}^\nu, \widehat{y}^{-\nu}) - \nabla_x \theta_\nu(\widehat{x}, \widehat{z}^\nu, \widehat{y}^{-\nu})^T(x - \widehat{x}) - \nabla_{y^{-\nu}} \theta_\nu(\widehat{x}, \widehat{z}^\nu, \widehat{y}^{-\nu})^T(y^{-\nu} - \widehat{y}^{-\nu}) - (\widehat{\zeta}_\nu + \overline{\zeta}_\nu)$ is continuous with respect to $(x, y, \widehat{x}, \widehat{y}, \widehat{z}, \widehat{\zeta})$ and convex with respect to (x, y).

□

Lemma 4.3 *Let (x^k, y^k, z^k, ζ^k) and $\overline{\eta}^k$ be the sequences generated by Algorithm 1. If*

$$\|(x^{k+1}, y^{k+1}) - (x^k, y^k)\|_2 \leqslant \sigma^k, \tag{6}$$

where $\sigma^k > 0$, then

$$\nabla F(x^k, y^k)^\top \begin{pmatrix} x - x^k \\ y - y^k \end{pmatrix} \geqslant -\sigma^k(\tau D + H + L_{\nabla F} D) - \overline{\eta}^k, \quad \forall (x, y) \in C_{\zeta^k}(x^k, y^k, z^k). \tag{7}$$

Proof. In view of (6), the claim follows from the following chain of inequalities that holds for every k and (x, y) in $C_{\zeta^k}(x^k, y^k, z^k)$:

$$\begin{aligned}
-\sigma^k(\tau D + H + L_{\nabla F} D) - \overline{\eta}^k &\leqslant -\sigma^k(\tau D + H + L_{\nabla F} D) \\
&\quad + \left[\nabla F(x^{k+1}, y^{k+1}) + \tau \begin{pmatrix} x^{k+1} - x^k \\ y^{k+1} - y^k \end{pmatrix}\right]^\top \begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \end{pmatrix} \\
&\leqslant -\sigma^k(\tau D + H + L_{\nabla F} D) + \nabla F(x^{k+1}, y^{k+1})^\top \begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \end{pmatrix} \\
&\quad + \tau \left\| \begin{matrix} x^{k+1} - x^k \\ y^{k+1} - y^k \end{matrix} \right\|_2 \left\| \begin{matrix} x - x^{k+1} \\ y - y^{k+1} \end{matrix} \right\|_2 \\
&\leqslant -\sigma^k(H + L_{\nabla F} D) + \nabla F(x^{k+1}, y^{k+1})^\top \begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \end{pmatrix} \\
&\leqslant -\sigma^k(H + L_{\nabla F} D) + \nabla F(x^{k+1}, y^{k+1})^\top \begin{pmatrix} x - x^k \\ y - y^k \end{pmatrix} \\
&\quad + \|\nabla F(x^{k+1}, y^{k+1})\|_2 \left\| \begin{matrix} x^k - x^{k+1} \\ y^k - y^{k+1} \end{matrix} \right\|_2 \\
&\leqslant -\sigma^k L_{\nabla F} D + \nabla F(x^{k+1}, y^{k+1})^\top \begin{pmatrix} x - x^k \\ y - y^k \end{pmatrix} \\
&\leqslant -\sigma^k L_{\nabla F} D + \nabla F(x^k, y^k)^\top \begin{pmatrix} x - x^k \\ y - y^k \end{pmatrix} \\
&\quad + \|\nabla F(x^{k+1}, y^{k+1}) - \nabla F(x^k, y^k)\|_2 \left\| \begin{matrix} x - x^k \\ y - y^k \end{matrix} \right\|_2 \\
&\leqslant \nabla F(x^k, y^k)^\top \begin{pmatrix} x - x^k \\ y - y^k \end{pmatrix},
\end{aligned}$$

where the first relation holds because (x^{k+1}, y^{k+1}) verifies the $\overline{\eta}^k$-approximate minimum principle (3) for problem $(\mathrm{P}_{\zeta^k}(x^k, y^k, z^k))$ while the last inequality is due to the Lipschitz continuity of ∇F over $X \times Y$. □

In Theorem 4.4 we state the convergence properties of Algorithm 1 in the case of $\eta > 0$.

Theorem 4.4 *Let $\eta > 0$ and $\overline{\zeta} \geqslant \breve{\eta}$. Under Assumptions A1 and A2, let (x^k, y^k, z^k, ζ^k) be the sequence generated by Algorithm 1. The following statements hold.*

(i) (x^k, y^k) is feasible for $(\mathrm{MFG}_{\varepsilon^{k-1}})$ where $\varepsilon_\nu^{k-1} \triangleq \eta + \frac{LD}{\sqrt{c/2}}\sqrt{\eta} + \zeta_\nu^{k-1} + \overline{\zeta}_\nu$, $\nu = 1, \ldots, N$, and any limit point $(\widetilde{x}, \widetilde{y}, \widetilde{z}, \widetilde{\zeta})$ is such that $(\widetilde{x}, \widetilde{y})$ is feasible for $(\mathrm{MFG}_\varepsilon)$, where $\varepsilon_\nu = \eta + \frac{LD}{\sqrt{c/2}}\sqrt{\eta} + \widetilde{\zeta}_\nu + \overline{\zeta}_\nu$, $\nu = 1, \ldots, N$. Moreover, for all k, we have:

$$F(x^{k+1}, y^{k+1}) - F(x^k, y^k) \leqslant \overline{\eta}^k - \tau \left\| \begin{matrix} x^k - x^{k+1} \\ y^k - y^{k+1} \end{matrix} \right\|_2^2 . \tag{8}$$

(ii) Assume $\overline{\zeta} > \breve{\eta}$, $\tau > 0$ and $\overline{\eta}^k \downarrow \widehat{\eta} \geqslant 0$.

(a) A convergent subsequence $\mathcal{K}$ of (x^k, y^k, z^k, ζ^k) exists such that

$$\lim_{k \in \mathcal{K}} -\tau \|(x^{k+1}, y^{k+1}) - (x^k, y^k)\|_2^2 + \overline{\eta}^k \geqslant 0. \tag{9}$$

(b) The limit point $(\widehat{x}, \widehat{y}, \widehat{z}, \widehat{\zeta})$ of the subsequence defined by $\mathcal{K}$ verifies $(\widehat{x}, \widehat{y}) \in C_{\widehat{\zeta}}(\widehat{x}, \widehat{y}, \widehat{z})$ and

$$\nabla F(\widehat{x}, \widehat{y})^\top \begin{pmatrix} x - \widehat{x} \\ y - \widehat{y} \end{pmatrix} \geqslant - \left[\widehat{\eta} + \sqrt{\frac{\widehat{\eta}}{\tau}} (\tau D + H + L_{\nabla F} D) \right], \quad \forall (x, y) \in C_{\widehat{\zeta}}(\widehat{x}, \widehat{y}, \widehat{z}). \tag{10}$$

Proof. (i) The claim follows observing that (x^k, y^k) is feasible for $(\mathrm{P}_{\zeta^{k-1}}(x^{k-1}, y^{k-1}, z^{k-1}))$ and, in turn, it is also feasible for $(\mathrm{MFG}_{\varepsilon^{k-1}})$ in view of (i) in Proposition 4.1. Any limit point turns out to be feasible for $(\mathrm{MFG}_{\varepsilon})$, thanks to the outer semicontinuity of the point-to-set mapping $W_\bullet$ (see Proposition 2.2).

Since (x^{k+1}, y^{k+1}) satisfies the $\overline{\eta}^k$-approximate minimum principle (3) for $(\mathrm{P}_{\zeta^k}(x^k, y^k, z^k))$ and $(x^k, y^k) \in C_{\zeta^k}(x^k, y^k, z^k)$ (see (ii) in Proposition 4.1), in view of the convexity of F, we have

$$\begin{aligned} -\overline{\eta}^k &\leqslant \nabla F(x^{k+1}, y^{k+1})^\top \begin{pmatrix} x^k - x^{k+1} \\ y^k - y^{k+1} \end{pmatrix} - \tau \left\| \begin{matrix} x^k - x^{k+1} \\ y^k - y^{k+1} \end{matrix} \right\|_2^2 \\ &\leqslant F(x^k, y^k) - F(x^{k+1}, y^{k+1}) - \tau \left\| \begin{matrix} x^k - x^{k+1} \\ y^k - y^{k+1} \end{matrix} \right\|_2^2, \end{aligned}$$

and, hence, (8) holds.

(ii) (a) From (8) we conclude that $\limsup_k \overline{\eta}^k - \tau \|(x^{k+1}, y^{k+1}) - (x^k, y^k)\|_2^2 \geqslant 0$, since otherwise the sequence of values $F(x^k, y^k)$ would tend to $-\infty$, contradicting the boundedness of F on $X \times Y$. Hence $\mathcal{K}$ exists such that (9) holds.

The convergence of the subsequence defined by $\mathcal{K}$ is due to the compactness of X and Y and because the sequence $\{\zeta^k\}$ is bounded.

(ii) (b) Thanks to (ii) in Proposition 4.1 and Lemma 4.2, we have $(\widehat{x}, \widehat{y}) \in C_{\widehat{\zeta}}(\widehat{x}, \widehat{y}, \widehat{z})$.

For every $k \in \mathcal{K}$,

$$\|(x^{k+1}, y^{k+1}) - (x^k, y^k)\|_2^2 \leqslant \frac{\overline{\eta}^k}{\tau} + \mu^k,$$

where $\mu^k \downarrow 0$. Hence, in view of Lemma 4.3, relation (7) holds with $k \in \mathcal{K}$ and

$$\sigma^k = \sqrt{\frac{\overline{\eta}^k}{\tau} + \mu^k}.$$

Taking the limit in (7) for $k \in \mathcal{K}$, in view of the continuity of C (see Lemma 4.2) we obtain the claim. □

Relation (10) leads to the following inexact suboptimality-like condition that is true by the convexity of F.

Corollary 4.5 *Let* $\eta > 0$, $\overline{\zeta} > \breve{\eta}$ *and* (x^k, y^k, z^k, ζ^k) *be the sequence generated by Algorithm 1. Under Assumptions A1 and A2, letting* $\overline{\eta}^k \downarrow \widehat{\eta} \geqslant 0$ *and* $\tau > 0$*, a limit point* $(\widehat{x}, \widehat{y}, \widehat{z}, \widehat{\zeta})$ *exists that verifies*

$$(\widehat{x}, \widehat{y}) \in W_\varepsilon, \qquad F(\widehat{x}, \widehat{y}) \leqslant F(x, y) + \left[\widehat{\eta} + \sqrt{\frac{\widehat{\eta}}{\tau}}(\tau D + H + L_{\nabla F} D)\right], \quad \forall (x, y) \in C_{\widehat{\zeta}}(\widehat{x}, \widehat{y}, \widehat{z}), \tag{11}$$

where $\varepsilon_\nu = \eta + \frac{LD}{\sqrt{c/2}}\sqrt{\eta} + \widehat{\zeta}_\nu + \overline{\zeta}_\nu$ *for every* ν.

Some comments are in order.

- Point (i) of Theorem 4.4 is the most practically relevant, since it states feasibility of each iteration (x^k, y^k) with respect to the single-leader multi-follower problem, with a degree of inexactness ε^{k-1} depending on ζ^{k-1}. Moreover, relation (8) ensures decreasing values for the leader's objective F, whenever $\left\|(x^k, y^k) - (x^{k+1}, y^{k+1})\right\|_2^2 > \frac{\overline{\eta}^k}{\tau}$, and therefore whenever the movement achieved by Step (S.2) is not too small. In this perspective, a practical stopping criterion could be $\left\|(x^k, y^k) - (x^{k+1}, y^{k+1})\right\|_2^2 \leqslant \frac{\overline{\eta}^k}{\tau}$ or directly checking the decrease of F.

  Notice that the parameter ζ_ν^k, which relaxes player ν's convexified value function constraint, could also decrease for a given iteration. For example, assuming $\zeta_\nu^{k-1} > 0$, if $(x^k, y^k) \in W_{\overline{\varepsilon}}$ with $\overline{\varepsilon}_\nu = \overline{\zeta}_\nu$, then $\zeta_\nu^k = 0$.

- Relation (11) is a suboptimality-like condition that takes into account the value-function constraints, as in general $X \times Y \nsubseteq C_{\widehat{\zeta}}(\widehat{x}, \widehat{y}, \widehat{z})$, where $\widehat{\zeta}$ is the limit inexactness vector.

5 The case when the subproblems can be solved exactly ($\boldsymbol{\eta} = \mathbf{0}$)

In the specific case of $\eta = 0$, whenever one is able to solve exactly the followers' problems (e.g. in the quadratic box-constrained case), $\zeta^k = 0$ and lower-level inexactness-related assumptions A1 and A2 are no more needed and sharper results can be obtained at the price of requiring constraint qualifications to hold. On the other hand, Slater's constraint qualifications are not guaranteed to hold (see Proposition 4.1 (ii)) for the subproblems ($\text{P}_0(x^k, y^k, z^k)$). However, requiring the Mangasarian-Fromovitz Constraint Qualification to hold for W_ε, with $\varepsilon = \overline{\zeta}$, at a point, ensures constraint qualifications for the subproblems with the same base point.

Proposition 5.1 *Let* $(x, y) \in W_\varepsilon$*, with* $\varepsilon = \overline{\zeta}$*. The Mangasarian-Fromovitz Constraint Qualification (MFCQ) for* W_ε *holds at* (x, y) *if and only if the following relations imply* $\lambda_\nu = 0$, $\nu = 1, \ldots, N$:

$$\begin{aligned}
&0 \in \textstyle\sum_{\mu=1}^N \lambda_\mu (\nabla_x \theta_\mu(x, y^\mu, y^{-\mu}) - \nabla_x \theta_\mu(x, w^\mu, y^{-\mu})) + N_X(x) \\
&\textit{and, } \forall \nu = 1, \ldots, N, \\
&0 \in \lambda_\nu \nabla_{y^\nu} \theta_\nu(x, y^\nu, y^{-\nu}) + \sum_{\mu \neq \nu} \lambda_\mu (\nabla_{y^\nu} \theta_\mu(x, y^\mu, y^{-\mu}) - \nabla_{y^\nu} \theta_\mu(x, w^\mu, y^{-\mu})) + N_{Y_\nu}(y^\nu) \\
&\lambda_\nu \in N_{\mathbb{R}_-}(\theta_\nu(x, y^\nu, y^{-\nu}) - \theta_\nu(x, w^\nu, y^{-\nu}) - \overline{\zeta}_\nu),
\end{aligned} \tag{12}$$

where w^ν *is any arbitrarily chosen point in* $S_\nu^0(x, y^{-\nu})$*, and* $N_A(a)$ *denotes the normal cone to the closed convex set* A *at* a.

Proof. Recalling the definition of MFCQ (see, e.g. [17, Statement of Theorem 6.14]), the claim is a consequence of Proposition 2.1. □

We observe that a necessary condition for the MFCQ to hold for W_ε, with $\varepsilon = \overline{\zeta}$, at any (x, y) is $\overline{\zeta} > 0$. If this is not the case, it is easy to prove that the MFCQ is not satisfied at any feasible point. With the next result, we state the relation between the constraint qualifications of the original problem (MFG$_\varepsilon$) and the subproblems ($\mathrm{P}_{\zeta^k}(x^k, y^k, z^k)$), where $z^k = w^k$ and every $w^{\nu,k}$ is any arbitrarily chosen point in $S^0_\nu(x^k, y^{-\nu,k})$.

Proposition 5.2 *Let $(x, y) \in W_\varepsilon$, with $\varepsilon = \overline{\zeta}$, and $w^\nu \in S^0_\nu(x, y^{-\nu})$ for every ν. Then, (x, y) satisfies the MFCQ for W_ε if and only if the Slater CQ holds for $C_0(x, y, w)$.*

Proof. Suffice it to observe that the MFCQ for $C_0(x, y, w)$ at (x, y) is exactly the same as conditions (12). The claim follows recalling that, in view of the convexity of $C_0(x, y, w)$, the MFCQ at (x, y) is equivalent to the Slater CQ for $C_0(x, y, w)$. □

The following results are similar to the ones in the inexact case (Section 4), and will lead to the convergence Theorem 5.5.

Proposition 5.3 *Letting $\eta = 0$,*

(i) if a point (x, y) is feasible for ($\mathrm{P}_0(x^k, y^k, z^k)$), with any arbitrarily chosen $z^{\nu,k} \in S^0_\nu(x^k, y^{-\nu,k})$ $\forall \nu$, then (x, y) is feasible also for (MFG$_\varepsilon$), with $\varepsilon = \overline{\zeta}$;

(ii) if $\overline{\zeta} \geqslant \breve{\eta}$, (x^k, y^k) for every k is feasible for ($\mathrm{P}_0(x^k, y^k, z^k)$) for any $z^{\nu,k} \in S^0_\nu(x^k, y^{-\nu,k})$ $\forall \nu$.

Proof. The claims can be shown similarly to the proof of Proposition 4.1. □

Lemma 5.4 *Let $\eta = 0$ and $(x^k, y^k, z^k, 0)$ be the sequence generated by Algorithm 1. The point-to-set mapping $C : \mathbb{R}^p \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^N \rightrightarrows \mathbb{R}^p \times \mathbb{R}^n$ is, relative to $X \times Y \times Y \times \mathbb{R}_+$,*

- *outer semicontinuous at any limit point of the sequence;*
- *inner semicontinuous at any limit point $(\tilde{x}, \tilde{y}, \tilde{z}, 0)$ of the sequence such that the MFCQ holds for W_ε, with $\varepsilon = \overline{\zeta}$.*

Proof. The outer semicontinuity can be easily proven resorting to standard reasoning [2, Theorem 3.1.1]. The inner semicontinuity follows from the same line of reasoning as in the proof of Lemma 4.2, by simply observing that the Slater CQ holds for $C_0(\tilde{x}, \tilde{y}, \tilde{z})$ in view of Proposition 5.2, due to the assumed MFCQ for W_ε at $(\tilde{x}, \tilde{y})$. □

In Theorem 5.5 we state the convergence properties of Algorithm 1 in case of $\eta = 0$.

Theorem 5.5 *Let $\eta = 0$, $\overline{\zeta} \geqslant \breve{\eta}$ and $(x^k, y^k, z^k, 0)$ be the sequence generated by Algorithm 1. The following statements hold.*

(i) (x^k, y^k) for any k, and all its limit points are feasible for (MFG$_\varepsilon$), with $\varepsilon = \overline{\zeta}$, and (8) holds for every k.

(ii) *Let* $\overline{\eta}^k \downarrow \widehat{\eta} \geqslant 0$ *and* $\tau > 0$.

(a) *A convergent subsequence* $\mathcal{K}$ *of* $(x^k, y^k, z^k, 0)$ *exists such that* (9) *holds.*

(b) *The limit point* $(\widehat{x}, \widehat{y}, \widehat{z}, 0)$ *of the subsequence defined by* $\mathcal{K}$ *verifies* $(\widehat{x}, \widehat{y}) \in C_0(\widehat{x}, \widehat{y}, \widehat{z}) \subseteq W_\varepsilon$*, with* $\varepsilon = \overline{\zeta}$*, and*

(1) *if the MFCQ holds for* W_ε *at* $(\widehat{x}, \widehat{y})$*, then*

$$\nabla F(\widehat{x}, \widehat{y})^\top \begin{pmatrix} x - \widehat{x} \\ y - \widehat{y} \end{pmatrix} \geqslant - \left[\widehat{\eta} + \sqrt{\frac{\widehat{\eta}}{\tau}} (\tau D + H + L_{\nabla F} D) \right], \quad \forall (x, y) \in C_0(\widehat{x}, \widehat{y}, \widehat{z}); \tag{13}$$

(2) *else,* $(\widehat{x}, \widehat{y})$ *satisfies the Fritz-John necessary conditions for* (MFG_ε).

(iii) *Let* $\overline{\eta}^k = 0$ *for all* k *and* $\tau > 0$.

(a) *We have*

$$\lim_k \|(x^{k+1}, y^{k+1}) - (x^k, y^k)\|_2^2 = 0.$$

(b) *For every limit point* $(\widetilde{x}, \widetilde{y}, \widetilde{z}, 0)$*, it holds that* $(\widetilde{x}, \widetilde{y}) \in C_0(\widetilde{x}, \widetilde{y}, \widetilde{z}) \subseteq W_\varepsilon$*, with* $\varepsilon = \overline{\zeta}$*, and*

(1) *if the MFCQ holds for* W_ε *at* $(\widetilde{x}, \widetilde{y})$*, then*

$$\nabla F(\widetilde{x}, \widetilde{y})^\top d \geqslant 0 \quad \forall d \in T_{W_\varepsilon}(\widetilde{x}, \widetilde{y}), \tag{14}$$

where $T_{W_\varepsilon}(\widetilde{x}, \widetilde{y})$ *denotes the tangent cone to* W_ε *at* $(\widetilde{x}, \widetilde{y})$*;*

(2) *else,* $(\widetilde{x}, \widetilde{y})$ *satisfies the Fritz-John necessary conditions for* (MFG_ε).

Proof. The proof follows the same line of reasoning as in the one of Theorem 4.4, along with the following considerations:

- point (*ii*)(a) is the same as point (*ii*)(a) of Theorem 4.4. Point (*iii*)(a) comes from (8) because $\overline{\eta}^k = 0$, by the same line of reasoning as point (*ii*)(a) of Theorem 4.4.
- the inner semicontinuity of C_0 needed in points (*ii*)(*b*)(1) and (*iii*)(*b*)(1) comes from Lemma 5.4;
- relation (13) comes immediately from (10);
- Lemma 4.3 holds for every subsequence thanks to point (*iii*)(a), then (14) is due to (10), recalling that the tangent cone $T_{C_0(\widetilde{x}, \widetilde{y}, \widetilde{z})}(\widetilde{x}, \widetilde{y})$ to the convex set $C_0(\widetilde{x}, \widetilde{y}, \widetilde{z})$ at $(\widetilde{x}, \widetilde{y})$ is given by the closure of the set of feasible directions for $C_0(\widetilde{x}, \widetilde{y}, \widetilde{z})$ at $(\widetilde{x}, \widetilde{y})$. In turn, the Slater CQ holds for $C_0(\widetilde{x}, \widetilde{y}, \widetilde{z})$ thanks to Proposition 5.2, thus, due to [17, Theorem 6.31] and in view of Proposition 2.1,

$$T_{C_0(\widetilde{x}, \widetilde{y}, \widetilde{z})}(\widetilde{x}, \widetilde{y}) = \left\{ s \in T_{X \times Y}(\widetilde{x}, \widetilde{y}) \;\middle|\; \begin{bmatrix} \nabla_x \theta_\nu(\widetilde{x}, \widetilde{y}) - \nabla_x \varphi_\nu(\widetilde{x}, \widetilde{y}^{-\nu}) \\ \nabla_{y^\nu} \theta_\nu(\widetilde{x}, \widetilde{y}) \\ \nabla_{y^{-\nu}} \theta_\nu(\widetilde{x}, \widetilde{y}) - \nabla_{y^{-\nu}} \varphi_\nu(\widetilde{x}, \widetilde{y}^{-\nu}) \end{bmatrix}^\top s \leqslant 0, \ \nu = 1, \ldots, N \right\},$$

which is easily seen to be nothing else but $T_{W_\varepsilon}(\widetilde{x}, \widetilde{y})$, with $\varepsilon = \overline{\zeta}$;

- as for points $(ii)(b)(2)$ and $(iii)(b)(2)$, the Fritz-John conditions hold by definition whenever MFCQ fails.

□

Relations (13) and (14) lead to the following inexact suboptimality-like condition that is true by the convexity of F.

Corollary 5.6 *Let $\eta = 0$, $\varepsilon = \overline{\zeta}$ and $(x^k, y^k, z^k, 0)$ be the sequence generated by Algorithm 1. The following statements hold:*

(i) letting $\overline{\eta}^k \downarrow \widehat{\eta} \geqslant 0$ and $\tau > 0$, a limit point $(\widehat{x}, \widehat{y}, \widehat{z}, 0)$ exists that verifies

$$(\widehat{x}, \widehat{y}) \in W_\varepsilon, \qquad F(\widehat{x}, \widehat{y}) \leqslant F(x, y) + \left[\widehat{\eta} + \sqrt{\frac{\widehat{\eta}}{\tau}}(\tau D + H + L_{\nabla F} D)\right], \quad \forall (x, y) \in C_0(\widehat{x}, \widehat{y}, \widehat{z});$$

(ii) if $\overline{\eta}^k = 0$ for all k, any limit point $(\widetilde{x}, \widetilde{y}, \widetilde{z}, 0)$ is such that

$$(\widetilde{x}, \widetilde{y}) \in W_\varepsilon, \qquad F(\widetilde{x}, \widetilde{y}) \leqslant F(x, y), \quad \forall (x, y) \in C_0(\widetilde{x}, \widetilde{y}, \widetilde{z}).$$

Some comments are in order.

- The comments regarding the feasibility of the sequence (x^k, y^k), and the decrease of $F(x^k, y^k)$ at the end of Section 4 apply also for this section, recalling that $\varepsilon^k = \overline{\zeta}$.
- About the convergence analysis of Theorem 5.5, we remark that point (ii), where the upper-level subproblems are solved inexactly ($\overline{\eta}^k > 0$), holds for a single limit point, while point (iii), where $\overline{\eta}^k = 0$, holds for every limit point.

  Whenever the limit $\widehat{\eta} = 0$, in both the frameworks described in points (ii) and (iii), we are able to prove that Algorithm 1 either converges to a Fritz-John point, or to a point satisfying standard stationarity conditions for (MFG$_\varepsilon$). We want to stress that such an outcome is what one can hope to achieve when dealing with a nonconvex program such as (MFG$_\varepsilon$), through a descent algorithm. Algorithm 1 is the first numerical method in the literature to achieve convergence in the single-leader multi-follower game, under the conditions of our framework, which cover many applications (see Section 7).

6 On constraint qualifications for the multi-follower game

As shown in the results of Section 5, the MFCQ turns out to be essential to prove that Algorithm 1 enjoys enhanced convergence properties (see Theorem 5.5, Corollary 5.6). Therefore, it is crucial to understand if it is sensible to assume the MFCQ for (MFG$_\varepsilon$), with $\varepsilon = \overline{\zeta}$. From now on, we assume $\eta = 0$ and $\varepsilon = \overline{\zeta}$ in this section.

It is easy to show that, whenever $\overline{\zeta} = 0$, at any point $(x, y) \in W_\varepsilon$, the MFCQ for (MFG$_\varepsilon$) is violated: it suffices to consider $y^\nu = w^\nu$ for all ν in (12).

However, for every $\overline{\zeta} > 0$, the MFCQ holds, e.g., at each $(x, y) \in W_\varepsilon$ with $y \in E_{\breve{\eta}}(x)$ and $\breve{\eta} < \overline{\zeta}$, since all inequality constraints in the description of W_ε are inactive, and $\lambda_\nu = 0$ follows from the complementarity condition in (MFG$_\varepsilon$) for each $\nu \in \{1, \ldots, N\}$.

Also, the MFCQ holds at any point $(x, y) \in W_\varepsilon$ at which exactly one inequality constraint is active. In fact, if only player ν's inequality is active, then the complementarity conditions in (12) yield $\lambda_\mu = 0$ for all $\mu \neq \nu$, and the second line in (12) collapses to the condition $0 \in \lambda_\nu \nabla_{y^\nu} \theta_\nu(x, y^\nu, y^{-\nu}) + N_{Y_\nu}(y^\nu)$. Assuming $\lambda_\nu > 0$ thus results in $0 \in \nabla_{y^\nu} \theta_\nu(x, y^\nu, y^{-\nu}) + N_{Y_\nu}(y^\nu)$ and, by the convexity of player ν's optimization problem, in the optimality of y^ν for player ν. This, however, contradicts the activity of player ν's inequality constraint, that is, $\theta_\nu(x, y^\nu, y^{-\nu}) = \varphi_\nu(x, y^{-\nu}) + \bar{\zeta}_\nu$. Hence also λ_ν must vanish, and MFCQ holds at (x, y). Clearly, the case of a single follower is an instance of this specific situation.

The following example shows that under an additional structural assumption, the MFCQ even holds at every point of W_ε.

Example 6.1 For each $\nu \in \{1, \ldots, N\}$ let θ_ν be a convex quadratic function on $\mathbb{R}^p \times \mathbb{R}^n$ that is strongly convex with respect to y^ν. This means that there exist a suitably partitioned symmetric positive semidefinite block matrix $A^\nu = (A^\nu_{ij})_{i,j \in \{0,1,\ldots,N\}}$ with $A^\nu_{\nu,\nu} \succ 0$, $\nu = 1, \ldots, N$, a corresponding vector $b^\nu = (b^\nu_i)_{i \in \{0,1,\ldots,N\}}$ and a scalar c^ν such that

$$\theta_\nu(x, y) = \frac{1}{2} \begin{pmatrix} x \\ y \end{pmatrix}^\top A^\nu \begin{pmatrix} x \\ y \end{pmatrix} + (b^\nu)^\top \begin{pmatrix} x \\ y \end{pmatrix} + c^\nu.$$

Note that each player ν's problem (1) then possesses a unique optimal point $w^\nu = w^\nu(x, y^{-\nu})$ for all choices of $(x, y^{-\nu}) \in X \times Y_{-\nu}$, where $Y_{-\nu} \triangleq \prod_{\mu \neq \nu} Y_\mu$.

Furthermore, let us assume that $-(A^\nu_{\nu,\nu})^{-1}(A^\nu_{\nu,0}\, X + A^\nu_{\nu,-\nu} Y_{-\nu} + b^\nu_\nu) \subseteq Y_\nu$ holds for each ν, where we put $A^\nu_{\nu,-\nu} = (A^\nu_{\nu,\mu})_{\mu \neq \nu}$. Player ν's first order optimality condition then is

$$0 = A^\nu_{\nu,0}\, x + A^\nu_{\nu,\nu} y^\nu + A^\nu_{\nu,-\nu} y^{-\nu} + b^\nu_\nu,$$

and convexity of the optimization problem entails the explicit description of the unique optimal point

$$w^\nu(x, y^{-\nu}) = -(A^\nu_{\nu,\nu})^{-1} \left(A^\nu_{\nu,0}\, x + A^\nu_{\nu,-\nu} y^{-\nu} + b^\nu_\nu \right)$$

for any $(x, y^{-\nu}) \in X \times Y_{-\nu}$.

Consequently, for each ν the corresponding inequality from the description of the set W_ε may be written as

$$\begin{aligned} \bar{\zeta}_\nu &\geqslant \theta_\nu(x, y^\nu, y^{-\nu}) - \varphi_\nu(x, y^{-\nu}) = \theta_\nu(x, y^\nu, y^{-\nu}) - \theta_\nu(x, w^\nu(x, y^{-\nu}), y^{-\nu}) \\ &= \frac{1}{2}(y^\nu)^\top A^\nu_{\nu,\nu} y^\nu - \frac{1}{2}(w^\nu)^\top A^\nu_{\nu,\nu} w^\nu + (A^\nu_{\nu,0} x + A^\nu_{\nu,-\nu} y^{-\nu} + b^\nu_\nu)^\top (y^\nu - w^\nu) \\ &= \frac{1}{2}(y^\nu - w^\nu)^\top A^\nu_{\nu,\nu} (y^\nu - w^\nu). \end{aligned}$$

Hence, although the function $-\varphi_\nu$ in general is concave, the linearity of the function $y^\nu - w^\nu(x, y^{-\nu})$ in $(x, y^\nu, y^{-\nu})$ implies that $\theta_\nu(x, y^\nu, y^{-\nu}) - \varphi_\nu(x, y^{-\nu})$ is convex.

Finally choose some $\bar{x} \in X$ and some Nash equilibrium $\bar{y} \in E_{\check{\eta}}(\bar{x})$, with $\check{\eta} = 0$. Then $(\bar{x}, \bar{y})$ is a Slater point with regard to the inequalities describing W_ε. Since all of these inequalities are smooth and convex, MFCQ holds everywhere in W_ε.

On the other hand, the next example shows that under our general assumptions MFCQ may be violated in W_ε at points with more than one active constraint, and, thus, whenever different followers are present.

Example 6.2 For a nonempty compact and convex set $X \subseteq \mathbb{R}^p$, $N = 2$, $n_1 = n_2 = 1$, and continuously differentiable convex functions ψ_1, ψ_2 on X let

$$\begin{aligned}\theta_1(x, y_1, y_2) &= \psi_1(x) + (y_1^2 + 5)e^{-2y_2}, \\ \theta_2(x, y_1, y_2) &= \psi_2(x) + (y_2^2 + 5)e^{-2y_1}\end{aligned}$$

as well as $Y_1 = Y_2 = [-2, 2]$. It is not hard to see that both θ_1 and θ_2 are continuously differentiable and fully convex on $X \times Y_1 \times Y_2$, so that our general assumptions are satisfied.

As optimal points for player 1 and 2 one computes $w_1(x, y_2) = w_2(x, y_1) = 0$ for any $(x, y_1, y_2) \in X \times Y_1 \times Y_2$, so that the corresponding optimal-value functions are

$$\begin{aligned}\varphi_1(x, y_2) &= \psi_1(x) + 5e^{-2y_2}, \\ \varphi_2(x, y_1) &= \psi_2(x) + 5e^{-2y_1}\end{aligned}$$

and, in turn,

$$\begin{aligned}\theta_1(x, y_1, y_2) - \varphi_1(x, y_2) &= y_1^2 e^{-2y_2}, \\ \theta_2(x, y_1, y_2) - \varphi_2(x, y_1) &= y_2^2 e^{-2y_1}.\end{aligned}$$

Note that the latter two functions are not convex on $X \times Y_1 \times Y_2$ and that for any $\overline{\zeta} > 0$ they describe the set

$$W_\varepsilon = \{(x, y_1, y_2) \in X \times Y_1 \times Y_2 | \ y_1^2 e^{-2y_2} \leqslant \overline{\zeta}_1, \ y_2^2 e^{-2y_1} \leqslant \overline{\zeta}_2\}.$$

For any $x \in X$ and the choice $\overline{\zeta}_1 = \overline{\zeta}_2 = e^{-2}$ let us consider the point $(x, 1, 1) \in W_\varepsilon$. Then, both inequality constraints are active at $(x, 1, 1)$ with gradients $2e^{-2}(0, 1, -1)^\top$ and $2e^{-2}(0, -1, 1)^\top$, respectively. Hence, MFCQ is violated at $(x, 1, 1)$.

Note that $\overline{\zeta}_1 = \overline{\zeta}_2 = e^{-2}$ is tailored to yield the degeneracy at $(x, 1, 1)$. On the other hand, MFCQ holds everywhere in W_ε for any other choice of $\overline{\zeta}$.

7 Numerical analysis

7.1 The bilevel multi-portfolio selection model

Inspired by the ESG multi-portfolio selection model of [5], let us consider a model where multiple account owners (followers) can invest in different financial assets, and the investment firm (leader) can set monetary incentives to influence the financial choices of the followers. Specifically, each account ν invests a given budget $b^\nu \in \mathbb{R}_+$ in K financial assets with the aim of optimizing multiple objectives. In this context, the variables $y^\nu \in \mathbb{R}^{n_\nu}$ represent the fraction of follower ν's budget to be invested in each asset. We consider the measure of account ν's portfolio expected return

$$I_\nu(y^\nu) \triangleq b^\nu (\mu^\nu)^T y^\nu,$$

where μ^ν are the expectations of the assets' return, and account ν's portfolio expected risk

$$R_\nu(y^\nu) \triangleq \frac{1}{2}(b^\nu)^2 (y^\nu)^T \Sigma^\nu y^\nu,$$

where Σ^ν is the symmetric and positive semidefinite covariance matrix. We also include in our analysis the transaction cost term

$$TC_\nu(y^1, \ldots, y^N) \triangleq b^\nu (y^\nu)^T \Omega^\nu \sum_{\lambda=1}^{N} b^\lambda (y^\lambda),$$

where $\Omega^\nu \in \mathbb{R}^{K \times K}$ is the positive semidefinite (not necessarily symmetric) market impact matrix, whose entry at position (i, j) gives account ν's expected impact of the liquidity of asset i on the liquidity of asset j (for further information about Ω^ν, see [15] and [20]). We remark that this term takes into account the liquidity component of the transaction costs, and since the trades are pooled and executed simultaneously, this term depends on the aggregated trades of all accounts. Lastly, each account wants to optimize the following sustainability-oriented criterion

$$ESGS_\nu(y^\nu) \triangleq b^\nu ESG^T y^\nu,$$

where $ESG \in \mathbb{R}^K$ are the assets' Environmental, Social and Governance (ESG) scores, that provide a measure of the sustainability of each asset [5].

Overall, each account owner minimizes the objective function

$$\begin{aligned}
\theta_\nu(x, y^\nu, y^{-\nu}) \triangleq \;& -I_\nu(y^\nu) + \rho^\nu R_\nu(y^\nu) + TC_\nu(y^1, \ldots, y^N) - x^\nu ESGS_\nu(y^\nu) \\
= \;& -b^\nu(\mu^\nu)^T y^\nu + \rho^\nu \tfrac{1}{2} (b^\nu)^2 (y^\nu)^T \Sigma^\nu y^\nu \\
& + b^\nu (y^\nu)^T \Omega^\nu [b^\nu(y^\nu) + \textstyle\sum_{\lambda \neq \nu} b^\lambda (y^\lambda)] \\
& - x^\nu b^\nu ESG^T y^\nu,
\end{aligned}$$

where $\rho^\nu \in \mathbb{R}_+$ is account ν's risk-aversion parameter and the positive parameter x^ν weights the impact of the term $ESGS_\nu$. As for the leader, the firm controls the variables x, with $x \triangleq (x^1 \ldots x^N) \in \mathbb{R}_+^N$, and therefore the monetary incentives to be given to each follower in order to influence their investments towards ESG-oriented portfolios. Overall, the firm's objective is to maximize the ESG score of all portfolios, while limiting the monetary incentives provided:

$$F(x, y) \triangleq -ESG^T \left(\sum_{\nu=1}^{N} b^\nu y^\nu \right) + \alpha \left\| x \right\|_2^2,$$

where $\alpha \in \mathbb{R}_+$ weights the firms' willingness to incentivise investors.

Note that, as the main departure from the model in [5], each follower's estimation of the transaction cost matrix Ω^ν is different, and we do not assume the followers' NEP to be reduced to a simpler potential game. Therefore, the problem presented in this section needs to be treated as a single-leader multi-follower bilevel game, and we treat it with this level of generality for the first time in the literature.

As for the convexity assumptions, while F is clearly convex, the followers' problems could be nonconvex with respect to $(x, y^\nu, y^{-\nu})$. However, in the spirit of [16], adding to each θ_ν a sufficiently large artificial quadratic term (in x and $y^{-\nu}$) , the resulting modified function

$$\theta'_\nu(x, y^\nu, y^{-\nu}) \triangleq \theta_\nu(x, y^\nu, y^{-\nu}) + \frac{\beta_\nu}{2}(x^T x + (y^{-\nu})^T y^{-\nu})$$

turns out to have the same minima with respect to y^ν, but it is convex for sufficiently large values of β_ν, as described in the following proposition.

Proposition 7.1 *For each ν, let* $\text{null}(\Sigma^\nu) \cap \text{null}(\Omega^\nu) = \{0\}$*. Then, for any*

$$\beta_\nu \geqslant \frac{\lambda_{\max}(ESGESG^T + \Omega^\nu (\Omega^\nu)^T \sum_{\lambda \neq \nu} (b^\lambda)^2)}{\lambda_{\min}(\rho^\nu \Sigma^\nu + 2sym(\Omega^\nu))},$$

$\arg\min_{y^\nu \in Y_\nu} \theta'_\nu(x, y^\nu, y^{-\nu}) = \arg\min_{y^\nu \in Y_\nu} \theta_\nu(x, y^\nu, y^{-\nu})$*, and θ'_ν is convex with respect to* $(x, y^\nu, y^{-\nu})$*.*

Proof. For the sake of notational simplicity, and without loss of generality, we prove this for player $\nu = N$. The first assertion is due to the fact that θ'_ν and θ_ν only differ by an additive term which is constant in y^ν.

To prove the second assertion, consider the Hessian matrix of θ'_N:

$$\nabla^2 \theta'_N(x, y^\nu, y^{-\nu}) = \begin{pmatrix} \beta_N I_N & 0 & \cdots & & b^N ESG^T \\ 0 & \beta_N I_{n_1} & & & b^1 b^N (\Omega^N)^T \\ & & \ddots & & \\ \vdots & & & & \vdots \\ & & & \beta_N I_{n_{N-1}} & b^1 b^{N-1} (\Omega^N)^T \\ b^N ESG & b^1 b^N \Omega^N & \cdots & b^1 b^{N-1} \Omega^N & (b^N)^2 (\rho^N \Sigma^N + 2\text{sym}(\Omega^N)) \end{pmatrix}.$$

By the Schur complement theorem, for any $\beta_N > 0$, $\nabla^2 \theta'_N(x, y^\nu, y^{-\nu})$ is positive semidefinite if and only if the following matrix is positive semidefinite:

$$(b^N)^2 (\rho^N \Sigma^N + 2\text{sym}(\Omega^N)) - \frac{1}{\beta_N} \left((b^N)^2 ESGESG^T + (b^N)^2 \Omega^N (\Omega^N)^T \sum_{\lambda=1}^{N-1} (b^\lambda)^2 \right),$$

that is, any unit vector z satisfies

$$\beta_N z^T \left((\rho^N \Sigma^N + 2\text{sym}(\Omega^N)) \right) z \geqslant z^T \left((ESGESG^T + \Omega^N (\Omega^N)^T \sum_{\lambda=1}^{N-1} (b^\lambda)^2) \right) z.$$

Due to $\text{null}(\Sigma^N) \cap \text{null}(\Omega^N) = \{0\}$, $\lambda_{\min}(\rho^N \Sigma^N + 2\text{sym}(\Omega^N)) \leqslant z^T (\rho^N \Sigma^N + 2\text{sym}(\Omega^N)) z$ and $\lambda_{\max}(ESGESG^T + \Omega^N (\Omega^N)^T \sum_{\lambda=1}^{N-1} (b^\lambda)^2) \geqslant z^T \left((ESGESG^T + \Omega^N (\Omega^N)^T \sum_{\lambda=1}^{N-1} (b^\lambda)^2) \right) z$, we have the second assertion. □

The previous result ensures that the followers' objectives, although not natively fully convex, can be substituted with θ'_ν without changing the original bilevel program's solutions. The resulting problem with θ'_ν satisfies all the assumptions required for Algorithm 1 to achieve the convergence results of the previous sections. In the following numerical tests, θ'_ν is used for all ν.

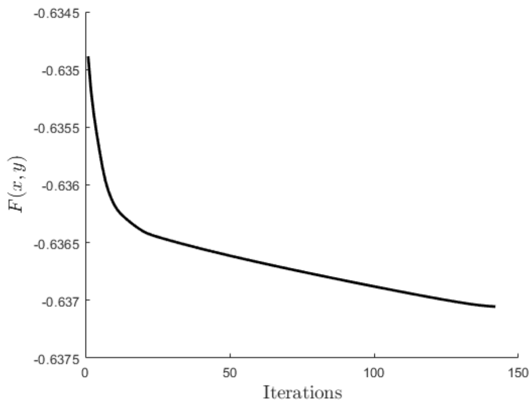
We consider two different datasets: $K_{\text{DIJA}} = 28$ assets from the Dow Jones Industrial Average (DJIA), and $K_{\text{NDX}} = 55$ assets from the NASDAQ 100 (NDX), consisting in daily prices, adjusted for dividends and stock splits, daily traded volumes and daily ESG scores (from 01/01/2019 to 31/12/2020), see [5]. For both datasets we consider $N = 5$ followers, resulting in $N \times (K_{\text{DIJA}} + 1) = 145$ and $N \times (K_{\text{NDX}} + 1) = 280$ total variables. The market impact matrices Ω^ν were computed in the same fashion as [5, 15], and the values for b^ν and ρ^ν were uniformly randomly generated in the intervals $(0, 200)$ and $(0, 0.2)$ respectively. We do not allow for shortselling, so that, for each ν, Y_ν is the standard K-simplex, moreover, we set $X = [0, 2]^N$.

7.2 Algorithmic choices and results

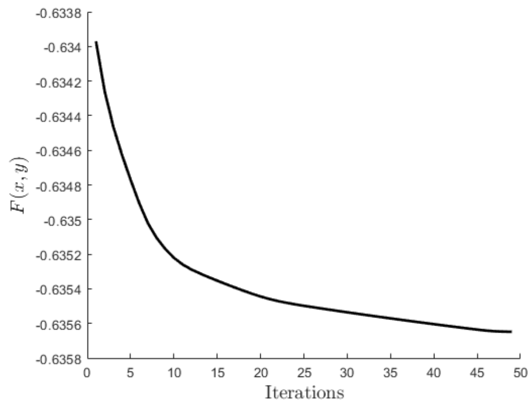
We use two different starting points x_0 for the two datasets, resulting in four test scenarios. The initial equilibrium $y^0 \in E_{\breve{\eta}}(x_0)$ is always computed starting from the equally weighted portfolio via a projected gradient-like method (see [11]), which is convenient since the projection on Y, which is made up by N K-simplexes, can be achieved through a finite-steps procedure (see [14]). Specifically, the stopping criterion used for the computation of the initial equilibrium is $\|y - P_Y(y - [\nabla \theta_\nu(x_0, y)]_{\nu=1}^N)\| \leqslant 1e-05$. Steps (S.1) and (S.2) of the algorithm are done using the `matlab` built-in functions `quadprog` and `fmincon`, respectively, using the default tolerances. The parameter β, used to compute the functions θ'_ν is set to 4, to ensure the theoretical properties needed for convergence, and we set $\alpha = \tau = 1e-03$. In our experiments, the inexactness in solving the followers'problem (η) and the leader's problem ($\overline{\eta}^k$) turn out to be negligible compared to the values we set for $\overline{\zeta} = 1e-04$. Therefore, in the implementation of the algorithm we set $\zeta_k = 0$ for all k, and we will show that the results are consistent with the analysis of Section 5. Moreover, the condition $\overline{\zeta} \geqslant \breve{\eta}$ required in Theorem 5.5 is satisfied in all our test settings. As for the stopping criterion for Algorithm 1, we choose a minimum relative decrease of the leader's function F, i.e. we stop whenever $F(x^k, y^k) - F(x^{k+1}, y^{k+1}) < (1e-06)(1 + |F(x^k, y^k)|)$.

Figure 1 shows the values of $F(x^k, y^k)$ through the iterations until the stopping criterion is met for the DIJA and the NDX datasets, and for both staring points x^0 considered. Note that, in accordance with relation (8) with $\overline{\eta}^k = 0$, $F(x^k, y^k)$ is monotone decreasing. Moreover, the decrease of $F(x^k, y^k)$ shows a desirable behaviour, since it is rather 'smooth' while flattening. Figure 2 shows the opposite of the value of the total ESG scores, i.e., the value of $F(x, y) - \alpha \|x\|_2^2$, in the same four test settings considered. This confirms that the leader's objective of maximizing the account owners' total ESG score is overall successful, and also shows that the term $\alpha \|x\|_2^2$ does not interfere with the leader's main goal. Tables 1-4 gather the values of the components of the followers' objectives and the approximation for inexact optimality of the followers' parametric problems. The results of these tables show that, for the final point, each component of followers' objectives is in the same order of magnitude, including the Transaction Cost term, indicating that the followers actually influence one another through this component of their objective function. Moreover, whenever the final value for x is higher, the followers select investments with higher ESG scores, showing that the leader's variables can actually influence the followers' choices. The values Eps Final $= \theta_\nu(x, y^\nu, y^{-\nu}) - \theta_\nu(x, z^\nu, y^{-\nu})$, where (x, y, z) is the final point computed, indicate the degree of approximation with respect to the optimal solution of each follower's individual optimization problem and they are consistent with the results of Theorem 5.5. We want to stress how these values are close to the value we set for $\overline{\zeta}$, showing that the equilibrium constraint is, in fact, active at the solution computed. Another proof of this is given by the fact that the minimum value for F, disregarding the value function constraints, is -7.328e-01 for the DIJA dataset (while the computed final values of F are -6.370e-01 and -6.356e-01), and -4.959e-01 for the NDX dataset (while the computed final values of F are -3.064e-01 and -3.088e-01), showing that the followers' game actually impacts the choice of the leader. We can conclude that the test settings considered represent complex bilevel multi-follower games, in which we can observe a concrete influence of the followers on one another, as well as between the two levels (leader - followers) in both ways.

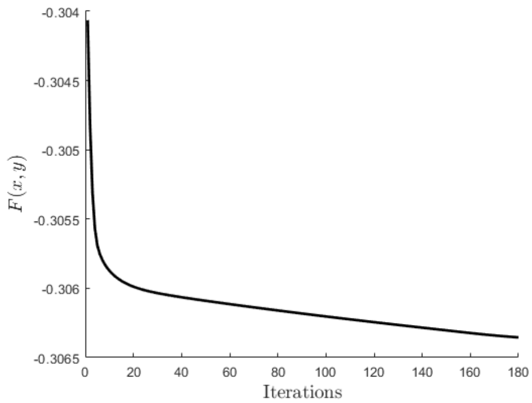
We want to stress that Algorithm 1 is quite efficient, only requiring the solution of (strongly) convex problems, which is a simple task even using the built-in matlab optimization functions employed. Meeting the stopping criterion only takes around 30 seconds for



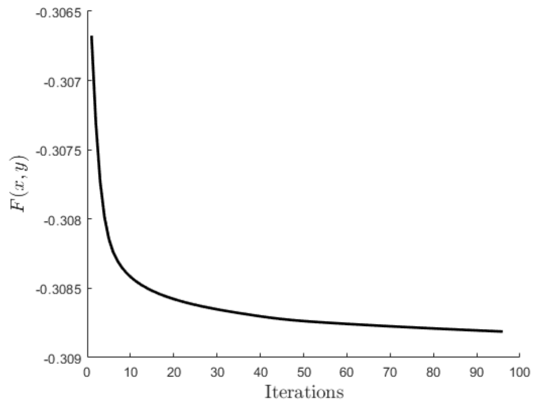
(a) $x^0 = [0.6]_{\nu=1}^5$ for DIJA dataset



(b) $x^0 = [1.2]_{\nu=1}^5$ for DIJA dataset



(c) $x^0 = [0.6]_{\nu=1}^5$ for NDX dataset



(d) $x^0 = [1.2]_{\nu=1}^5$ for NDX dataset

Figure 1: Values of $F(x, y)$ through the iterations for both datasets considered

the DIJA dataset, and around one minute for the NDX dataset. For this specific model, the subproblems are particularly easy (being quadratic programs), and ζ^k is not needed due to the high degree of precision in their solution. The inclusion of the additional inexactness term ζ^k might be needed for subproblems which are computationally harder to solve, and we plan on analysing this in future works.

8 Conclusions

We develop a novel sequential convex approximation method to address single-leader multi-follower games under mild convexity assumptions. We analyze both the exact and inexact iterative solution of the convex subproblems, and study the related convergence properties. We also study the validity of standard constraint qualifications for the complicated framework we deal with. Finally, we propose a novel bilevel ESG-oriented multi-portfolio selection model, so that we are able to test our method numerically, confirming the theoretical insights.

	Accounts				
	1	2	3	4	5
Income	7.397e-02	1.667e-01	1.556e-01	2.323e-01	1.921e-01
Risk	6.519e-02	5.632e-01	5.141e-02	1.225e-01	2.364e-01
Transaction	4.210e-02	1.823e-01	2.198e-03	4.613e-02	2.336e-02
ESG	9.614e-02	1.493e-01	7.739e-02	1.507e-01	1.653e-01
Eps Final	9.984e-05	9.975e-05	9.986e-05	9.979e-05	9.979e-05
x^ν	6.029e-01	6.000e-01	5.981e-01	6.010e-01	6.011e-01

Table 1: Results for final point computed, for the DIJA dataset, with $x^0 = [0.6]_{\nu=1}^5$, where: Income = I_ν, Risk = $\rho^\nu R_\nu$, Transaction = TC_ν, ESG = S_ν Eps final = $\theta_\nu - \varphi_\nu$.

	Accounts				
	1	2	3	4	5
Income	7.057e-02	1.662e-01	1.575e-01	2.311e-01	1.905e-01
Risk	6.590e-02	5.625e-01	5.555e-02	1.225e-01	2.378e-01
Transaction	3.992e-02	1.840e-01	1.711e-03	4.585e-02	2.186e-02
ESG	9.715e-02	1.496e-01	7.882e-02	1.513e-01	1.659e-01
Eps Final	9.983e-05	9.970e-05	9.525e-05	9.974e-05	9.975e-05
x^ν	1.199e+00	1.199e+00	1.198e+00	1.199e+00	1.199e+00

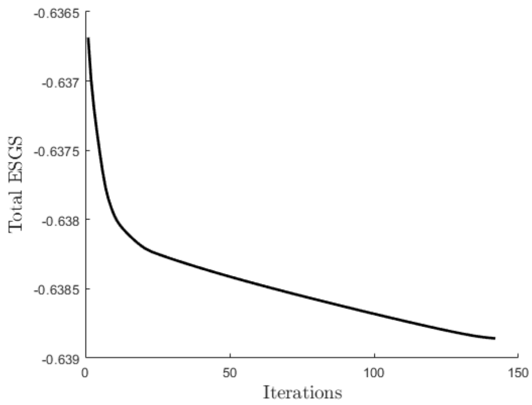
Table 2: Results for final point computed, for the DIJA dataset, with $x^0 = [1.2]_{\nu=1}^5$, where: Income = I_ν, Risk = $\rho^\nu R_\nu$, Transaction = TC_ν, ESG = S_ν, Eps final = $\theta_\nu - \varphi_\nu$.

	Accounts				
	1	2	3	4	5
Income	7.635e-01	3.410e-01	1.595e-01	3.497e-01	2.659e-01
Risk	3.559e-01	3.316e-01	4.299e-02	3.327e-01	1.215e-01
Transaction	6.480e-02	2.611e-02	1.269e-02	4.184e-02	2.180e-02
ESG	1.087e-01	6.352e-02	1.733e-02	8.433e-02	3.308e-02
Eps Final	9.988e-05	9.987e-05	9.936e-05	9.989e-05	9.977e-05
x^ν	6.108e-01	6.091e-01	5.985e-01	6.153e-01	6.000e-01

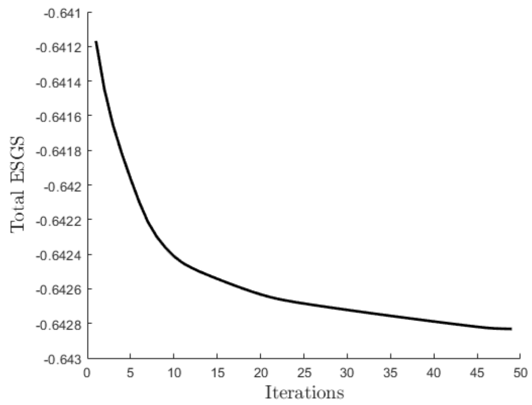
Table 3: Results for final point computed, for the NDX dataset, with $x^0 = [0.6]_1^5$, where: Income = I_ν, Risk = $\rho^\nu R_\nu$, Transaction = TC_ν, ESG = S_ν, Eps final = $\theta_\nu - \varphi_\nu$.

	Accounts				
	1	2	3	4	5
Income	7.640e-01	3.366e-01	1.586e-01	3.435e-01	2.639e-01
Risk	3.599e-01	3.290e-01	4.251e-02	3.294e-01	1.204e-01
Transaction	6.481e-02	2.727e-02	1.275e-02	4.169e-02	2.194e-02
ESG	1.117e-01	6.565e-02	1.755e-02	8.748e-02	3.365e-02
Eps Final	9.988e-05	9.987e-05	9.925e-05	9.985e-05	9.978e-05
x^ν	1.205e+00	1.204e+00	1.198e+00	1.203e+00	1.199e+00

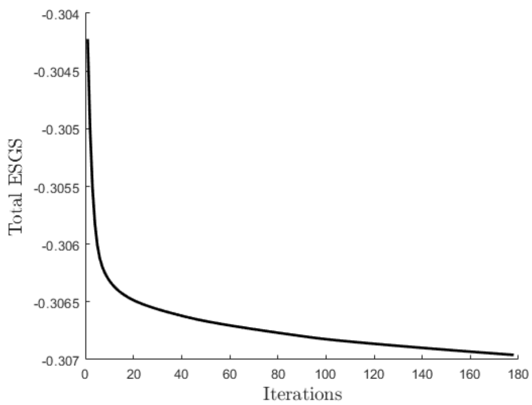
Table 4: Results for final point computed, for the NDX dataset, with $x^0 = [1.2]_1^5$, where: Income = I_ν, Risk = $\rho^\nu R_\nu$, Transaction = TC_ν, ESG = S_ν, Eps final = $\theta_\nu - \varphi_\nu$.



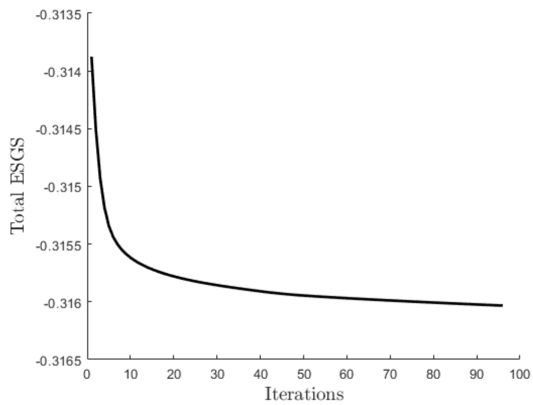
(a) $x^0 = [0.6]_{\nu=1}^5$ for DIJA dataset



(b) $x^0 = [1.2]_{\nu=1}^5$ for DIJA dataset



(c) $x^0 = [0.6]_{\nu=1}^5$ for NDX dataset



(d) $x^0 = [1.2]_{\nu=1}^5$ for NDX dataset

Figure 2: Values of total ESG score, i.e., $F(x, y) - \alpha \|x\|_2^2$, through the iterations for both datasets considered

References

[1] Didier Aussel, Sébastien Lepaul, and Léonard von Niederhäusern. A multi-leader-follower game for energy demand-side management. *Optimization*, 72(2):351–381, 2023.

[2] B. Bank, J. Guddat, D. Klatte, B. Kummer, and K. Tammer. *Non-linear parametric optimization.* Akademie-Verlag Berlin, 1982.

[3] Giancarlo Bigi, Lorenzo Lampariello, Simone Sagratella, and Valerio Giuseppe Sasso. Approximate variational inequalities and equilibria. *Computational Management Science*, 20(1):43, 2023.

[4] Francesco Caruso, M Beatrice Lignola, and Jacqueline Morgan. Regularization and approximation methods in stackelberg games and bilevel optimization. *Bilevel optimization: Advances and next challenges*, pages 77–138, 2020.

[5] Francesco Cesarone, Lorenzo Lampariello, Davide Merolla, Jacopo Maria Ricci, Simone Sagratella, and Valerio Giuseppe Sasso. A bilevel approach to esg multi-portfolio selection. *Computational Management Science*, 20(1):24, 2023.

[6] S. Dempe. *Foundations of bilevel programming.* Springer Science & Business Media, 2002.

[7] S. Dempe and J. Dutta. Is bilevel programming a special case of a mathematical program with complementarity constraints? *Mathematical programming*, 131(1-2):37–48, 2012.

[8] 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.

[9] Axel Dreves and Simone Sagratella. Nonsingularity and stationarity results for quasi-variational inequalities. *Journal of Optimization Theory and Applications*, 185(3):711–743, 2020.

[10] F. Facchinei and L. Lampariello. Partial penalization for the solution of generalized Nash equilibrium problems. *Journal of Global Optimization*, 50(1):39–57, 2011.

[11] F. Facchinei and J.-S. Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems.* Springer, 2003.

[12] F. Facchinei and S. Sagratella. On the computation of all solutions of jointly convex generalized Nash equilibrium problems. *Optimization Letters*, 5(3):531–547, 2011.

[13] Francisco Facchinei, Lorenzo Lampariello, and Gesualdo Scutari. Feasible methods for nonconvex nonsmooth problems with applications in green communications. *Mathematical Programming*, 164(1-2):55–90, 2017.

[14] L. Lampariello, C. Neumann, J.M. Ricci, S. Sagratella, and O. Stein. Equilibrium selection for multi-portfolio optimization. *European Journal of Operational Research*, 295:363–373, 2021.

[15] L. Lampariello, C. Neumann, J.M. Ricci, S. Sagratella, and O. Stein. Equilibrium selection for multi-portfolio optimization. *European Journal of Operational Research*, 295(1):363–373, 2021.

[16] L. Lampariello and S. Sagratella. Numerically tractable optimistic bilevel problems. *Computational Optimization and Applications*, 76:277–303, 2020.

[17] R.T. Rockafellar and J.B. Wets. *Variational Analysis.* Springer, 1998.

[18] T. Tanino and T. Ogawa. An algorithm for solving two-level convex optimization problems. *International journal of systems science*, 15(2):163–174, 1984.

[19] Haoning Xi, Didier Aussel, Wei Liu, S Travis Waller, and David Rey. Single-leader multi-follower games for the regulation of two-sided mobility-as-a-service markets. *European Journal of Operational Research*, 2022.

[20] Y. Yang, F. Rubio, G. Scutari, and D.P. Palomar. Multi-portfolio optimization: A potential game approach. *IEEE Transactions on Signal Processing*, 61(22):5590–5602, 2013.

[21] Jane J Ye, Xiaoming Yuan, Shangzhi Zeng, and Jin Zhang. Difference of convex algorithms for bilevel programs with applications in hyperparameter selection. *Mathematical Programming*, 198(2):1583–1616, 2023.