Comput Optim Appl (2014) 59:85–112
DOI 10.1007/s10589-014-9642-3

Non-cooperative games with minmax objectives

Francisco Facchinei · Jong-Shi Pang · Gesualdo Scutari

Received: 20 January 2013 / Published online: 11 March 2014

Abstract We consider noncooperative games where each player minimizes the sum of a smooth function, which depends on the player, and of a possibly nonsmooth function that is the same for all players. For this class of games we consider two approaches: one based on an augmented game that is applicable only to a minmax game and another one derived by a smoothing procedure that is applicable more broadly. In both cases, centralized and, most importantly, distributed algorithms for the computation of Nash equilibria can be derived.

Keywords Nash equilibrium problem · Nondifferentiable objective function · Distributed algorithm · Smoothing

To Masao Fukushima, on the occasion of his 65th birthday, with friendship and admiration.

F. Facchinei (✉)
Department of Computer, Control, and Management Engineering Antonio Ruberti, University of Rome La Sapienza, Via Ariosto 25, 00185 Rome, Italy
e-mail: facchinei@dis.uniroma1.it

J.-S. Pang
Deptartment of Industrial and Systems Engineering, University of Southern California, Los Angeles, CA, USA
e-mail: jongship@usc.edu

G. Scutari
Department of Electrical Engineering, State University of New York (SUNY) at Buffalo, Buffalo, NY, USA
e-mail: gesualdo@buffalo.edu

1 Introduction

It is well known that an optimization problem with a *finite* minmax objective can be converted to an equivalent optimization problem wherein the maximum of the finitely many functions in the objective is replaced by an auxiliary variable that is constrained to be such a maximum. Moreover the conversion retains the convexity of the original optimization problem, if applicable. For a *continuous* minmax optimization problem, a variational inequality (VI) formulation defined by an associated *saddle function* exists that provides a saddle characterization of the original optimization problem; see [11, Section 1.4.1]. In contrast to a single minmax problem, the research on the multi-agent minmax non-cooperative Nash game is largely unexplored. In such a game, there is a finite number of selfish players each of whom solves a finite or continuous minmax optimization problem, leading to a *finite minmax game* or a *continuous minmax game*, respectively.

In general, there are two classifications of a Nash game depending on whether a player's constraints contain or do not contain the rivals' decision variables: in the former case we speak of a *standard game*; in the latter case we have a *generalized* game. For a standard finite minmax game, the conversion mentioned above would convert it into a generalized game. Starting with the pioneering work of Rosen [35], there is now a substantial literature on non-cooperative games of the latter kind [6,7,9,10,15,17, 18,21–28,30–32]; nevertheless, none of these studies have directly addressed the finite minmax game. Although in principle the results in these references could be attempted on the resulting generalized game derived from a standard finite minmax game, these results can be expected to require unnecessarily restrictive assumptions because they do not deal with the special nature of the minmax reformulation; furthermore the possible application of these general results has never been attempted and it is plain that, in any case, it is by no means a trivial task. Moreover, when it comes to distributed algorithms, the state-of-the-art convergence results [12,13,36] for a standard Nash game are not applicable to a finite minmax game because the results require the players' objective functions to be continuously differentiable with Lipschitz gradients. In contrast, the finite minmax game, as stated with the players' objective functions being of the minmax kind, certainly fails this differentiability requirement. The current state of the continuous minmax game is practically in a vacuum, the only case of nondifferentiable game that can be solved in a distributed way being that of potential games; see for example [14] and references therein.

In this paper, we present a first step towards the study of distributed algorithms for nondifferentiable games by assuming that each player minimizes the sum of a nondifferentiable function, which is the same for all players, and of a private, differentiable function that is different for each player. This type of games corresponds to the case in which the players have some common "interest" along with their own private ones; see also Sect. 2.1. Although our main developments are for the minmax case (be it finite or continuous) some of the results are also applicable, in principle, to general nondifferentiable objectives.

We consider two approaches to the solution of minmax games. The first approach is based on a simple observation that permits to transform the original game into a smooth game with an additional player to which the methods in [12,13,36] can be applied.

In the second approach we approximate the original game by a sequence of smooth ones and reduce the problem of developing (distribute) algorithms for the original nondifferentiable game to that of solving a sequence of smooth, convex optimization problems, a task for which distributed algorithms are well studied. The two approaches work under assumptions that share the same flavor, even if they can not be compared directly. The first approach is marginally more limited in scope but is extremely simple and natural. The second approach has the potential of being applicable beyond the case of minmax objectives but is somewhat more complex. Although the solution of minmax games by distributed algorithms is our main goal, the emphasis in this paper is on suitable transformations of a minmax game that permit the applicability of known distributed algorithms rather than on the description of the algorithms themselves. Indeed, the identification of adequate transformations of the original minmax game is the main focus of this paper, since once this has been accomplished the development of distributed algorithms can be accomplished readily.

The paper is organized as follows. In Sect. 2 we introduce in detail the game we study along with the assumptions we use to analyze it. At the end of this section we also give a detailed description of an allocation problem in wireless networks which illustrates and motivates our framework. In Sect. 3 we discuss our first approach, based on a "pulled out" extended game where an additional player is introduced and all players' objective functions are smooth. In Sect. 4 we take an alternative route and explore the idea of smoothing the minmax terms, reducing the solution of the original problem to that of a sequence of smooth games.

2 Problem description and preliminaries

The basic object of our interest is a game $\mathcal{G}$ with N players, where player i makes decisions on the variables $x^i \in \mathbb{R}^{n_i}$ and whose objective function depends also on $x^{-i} \triangleq (x^1, x^2, \ldots, x^{i-1}, x^{i+1}, \ldots, x^N)$. Player i, anticipating x^{-i}, solves the optimization problem:

$$\underset{x^i \in X_i}{\text{minimize}} \left[f_i(x^i, x^{-i}) + \phi(x^i, x^{-i}) \right]; \tag{1}$$

where X_i is a closed convex set in $\mathbb{R}^{n_i}$, and the functions $f_i(x)$ and $\phi(x)$ are continuous on $\Omega \triangleq \prod_{i=1}^{N} \Omega_i$ where each Ω_i is an open convex set containing X_i. In the sequel we will also use the following notation:

$$X \triangleq \prod_{i=1}^{N} X_i, \qquad X_{-i} \triangleq \prod_{j \neq i} X_j, \qquad \Omega_{-i} \triangleq \prod_{j \neq i} \Omega_j.$$

Our aim is the computation of a Nash equilibrium (which we also term, more simply, a solution of the game), that is a point $x^* \in X$ such that

$$f_i(x^{*,i}, x^{*,-i}) + \phi(x^{*,i}, x^{*,-i}) \le f_i(x^i, x^{*,-i}) + \phi(x^i, x^{*,-i}), \qquad \forall x^i \in X_i$$

holds for all players i. A major motivation to study the nonsmooth game $\mathcal{G}$ is to treat the case in which ϕ is a pointwise, continuous or finite maximum function; we will denote the resulting games as $\mathcal{G}^{\mathrm{C}}_{\mathrm{mm}}$ and $\mathcal{G}^{\mathrm{F}}_{\mathrm{mm}}$, respectively. For the game $\mathcal{G}^{\mathrm{C}}_{\mathrm{mm}}$, we have

$$\phi(x) \triangleq \underset{y \in Y}{\text{maximum}}\ g(x, y), \tag{2}$$

where Y is a closed convex set in $\mathbb{R}^m$, and $g(x, y)$ is assumed to be continuous on $\Omega \times \Upsilon$, with Υ being an open convex set containing Y. Furthermore, when dealing with the maximum value function (2) we will also assume that

Assumption 1 The set Y is compact and, for each fixed $x \in \Omega$, $g(x, \bullet)$ is concave on Υ. □

By the compactness of Y, the function ϕ is always well-defined, while the concavity of $g(x, \bullet)$ makes the computation of ϕ a practical task.

For the game $\mathcal{G}^{F}_{\mathrm{mm}}$, we have

$$\phi(x) \triangleq \underset{j \in J}{\text{maximum}}\ g_j(x) \tag{3}$$

with $J \triangleq \{1, 2, \ldots, |J|\}$. Note that a finite minmax problem can always be seen as a particular case of a continuous minmax problem. In fact we can write

$$\max_{j \in J} g_j(x) = \max_{y \in \Delta} \sum_{j \in J} y_j g_j(x),$$

with $\Delta \triangleq \left\{ y \in \mathbb{R}^{|J|} \mid y \ge 0, \sum_{j \in J} y_j = 1 \right\}$ being the unit simplex in $\mathbb{R}^{|J|}$. Nevertheless, the finite minmax game has its special feature that enables for instance the construction of an explicit approximation; see (30).

Returning to the general game $\mathcal{G}$, we introduce a weak convexity property on the functions f_i and ϕ that exploits the following fact: it is the sum function $f_i + \phi$ that we are interested in, and not so much in the individual summands. In turn, this property is built on the following:

Definition 1 Let a function $h : \Omega \to \mathbb{R}$ be given, where $\Omega \subseteq \mathbb{R}^n$ is an open, convex set. We say that h is σ-convex (on Ω) if, for all $x, y \in \Omega$, it holds that

$$h(\beta x + (1-\beta)y) \le \beta h(x) + (1-\beta)h(y) - \frac{\sigma}{2}\beta(1-\beta)\|x - y\|^2,$$

for some real constant σ and for all $\beta \in [0, 1]$. □

Note that if $\sigma = 0$ or $\sigma > 0$ the above definition reduces to the standard definition of convex or strongly convex function respectively. However, if $\sigma < 0$ the above definition is a relaxation of convexity and is sometimes termed as “weak convexity”, although this terminology is not used consistently by researchers. The concept of σ-convexity is not new and has already been used, if sparingly, in the literature, see for example [4,40] and references therein. In the field of algorithms for games a concept of “weak convexity” very close to that of σ-convexity, has been invoked in the analysis of the so called relaxation algorithm, see for example [25,39]. There are also strong links between σ-convexity and lower-C^2 functions, see [34].

It follows immediately from the definition of σ-convexity that if h_1 and h_2 are σ_1- and σ_2-convex, respectively, then $h_1 + h_2$ is $(\sigma_1 + \sigma_2)$-convex. Furthermore, if h is σ-convex then it is also straightforward to see that $h(\cdot, x^{-i})$ is still σ-convex for any fixed x^{-i}. Taking into account the simple identity

$$\beta\|x\|^2 + (1-\beta)\|y\|^2 - \|\beta x + (1-\beta)y\|^2 = \beta(1-\beta)\|x-y\|^2,$$

it is readily checked that h is σ-convex if and only if $h(x) - \frac{\sigma}{2}\|x\|^2$ is convex, thus implying that a σ-convex function is locally Lipschitz continuous. Furthermore, if h is continuously differentiable we deduce, from classical results about convexity, that h is σ-convex if and only if

$$(\nabla h(x) - \nabla h(y))^T(x-y) \geq \sigma\|x-y\|^2, \qquad \forall x, y \in \Omega. \tag{4}$$

If in addition h is twice continuously differentiable then h is σ-convex on Ω if and only if

$$d^T\nabla^2 h(x)d \geq \sigma\|d\|^2, \qquad \forall x \in \Omega,\ \forall d \in \mathbb{R}^n. \tag{5}$$

Once again we recall that σ can be a nonnegative or a negative number. The characterization (5) of σ-convexity shows that every quadratic function must be σ-convex, where σ is the smallest eigenvalue of the quadratic form. The characterization (4) clearly shows that in general if ∇h is Lipschitz continuous on Ω with constant L, then h is σ-convex on the same set, with $\sigma = -L$, thus indicating that the class of σ-convex functions is wide and certainly encompasses much more than convex functions. The use of this class of functions will become clear shortly. In connection to the above definitions and, in particular, to (4), it will also be useful to define the concept of σ-monotonicity for a vector-valued function. Specifically, given a map $H : \Omega \to \mathbb{R}^n$ where $\Omega \subseteq \mathbb{R}^n$ is an open convex set, we say that H is σ-monotone on Ω, with σ a real constant if

$$(H(x) - H(y))^T(x-y) \geq \sigma\|x-y\|^2, \qquad \forall x, y \in \Omega. \tag{6}$$

If $\sigma = 0$ or $\sigma > 0$ this is the classical definition of monotone or strongly monotone function. It is immediate to verify that H is σ-monotone on Ω if and only if $H(x) - \sigma x$ is monotone on Ω. From this fact we can then derive that if H is continuously differentiable with Jacobian matrix $JH(x)$, then H is σ-monotone on Ω if and only if

$$d^T JH(x)d \geq \sigma \|d\|^2, \qquad \forall x \in \Omega, \ \forall d \in \mathbb{R}^n. \tag{7}$$

If ϕ is the pointwise continuous max function (2), then ϕ is σ^g-convex if for each $y \in Y$, $g(\bullet, y)$ is σ^g-convex. A similar statement can be made for the finite max function (3). We are now ready to give a key assumption on game $\mathcal{G}$; the first part (a) of the assumption is stated in terms of a general function ϕ that covers both the pointwise continuous and finite max function.

Assumption 2 (a) For each fixed $x^{-i} \in \Omega_{-i}$, $f_i(\bullet, x^{-i})$ is σ_i^f-convex on Ω_i; ϕ is σ^ϕ-convex on Ω and $\sigma_i^f + \sigma^\phi \geq 0$ for all i.
(b) Each f_i is continuously differentiable in x on Ω and the partial gradient $\nabla_{x^i} f_j(x)$ is Lipschitz continuous on Ω_i with constant L_{ij}.

□

We remark that the σ^ϕ-convexity on Ω of $\phi(x)$ implies the σ-convexity of $\phi(\bullet, x^{-i})$ on Ω_i for each $x^{-i} \in \Omega_{-i}$. Thus, $f_i(\bullet, x^{-i}) + \phi(\bullet, x^{-i})$ is convex. It is clear that for the sum of two functions to be convex it is not necessary that the two individual functions be both convex, but it could happen that the convexity of one of the two "makes up" for the lack of convexity of the other. The introduction of the constants σ_i^f and σ^ϕ goes in the direction of making this concept of one function compensating for the lack of convexity of the other more quantitative and precise, and will pay dividends later, when it will allow us to easily deal with nonconvex "parts" of the game $\mathcal{G}$.

We already mentioned in the Sect. 1 that the main difficulty in the development of distributed algorithms for the game $\mathcal{G}$ is the presence of the term ϕ. In order to cope with the nondifferentiability of ϕ we propose below two strategies. Applicable to the continuous and finite minmax games and presented in the next section, the first strategy circumvents the difficulty of the nondifferentiable minmax functions by relating the original games $\mathcal{G}_{\text{mm}}^{\text{C,F}}$ to a smooth game with an additional player whose solution can be carried out in a distributed way. Applicable to an abstract nonsmooth function ϕ and discussed in Sect. 4, the second strategy approximates the term ϕ by a smooth function, leading a sequence of smooth(ed) games that form the basis for the development of distributed algorithms for solving the original game $\mathcal{G}$. Before presenting these details, we conclude this section by describing in the next section a concrete example of an application of a minmax game.

2.1 A fair tlc minmax game

Consider a resource allocation problem in wireless *ad-hoc* networks, namely the power control problem over parallel Gaussian Interference Channels (ICs). This problem is well studied, see [13,36] as entry points to the literature on this subject. Here, differently from what usually done, we introduce a degree of fairness in the model.

The problem discussed in this section exemplifies well the point of view in which a "protocol" is being designed. Indeed, the main setting we have in mind in which the results we discuss in this paper might be useful is that of "designing games" or, in the terminology of Aumann, that of "game engineering".

In an IC system, there are N transmitter-receivers (players) that want to exchange messages over a set of C independent noisy channels. For simplicity and without loss of generality we consider transmissions over ICs in the frequency domain, termed as frequency-selective ICs. We associate with each of the N users a nonnegative vector variable $p^i \triangleq (p_n^i)_{n=1}^C \geq 0$, representing the power allocated over the C channels by the transmission-receiver pair i. As such, these variables satisfy some bound constraints $0 \leq p_n^i \leq p_n^{i,\max}$, $i = 1, \ldots, N$, $n = 1, \ldots, C$, where $p_n^{i,\max}$ are given upper bounds imposed by the regulator to limit the amount of power radiated by user i over licensed bands. Furthermore, each transmitter has a power budget limit denoted by P^i, so that the power vector p^i is constrained to satisfy $\sum_{n=1}^{C} p_n^i \leq P^i$. The set of power constraints of each user i is thus defined as

$$\mathcal{P}^i \triangleq \left\{ p^i \in \mathbb{R}^C \, : \, \sum_{n=1}^{C} p_n^i \leq P^i, \quad 0 \leq p^i \leq p^{i,\max} \triangleq \left(p_n^{i,\max} \right)_{n=1}^{C} \right\}. \tag{8}$$

The IC is used to model practical multiuser systems that do not have any infrastructure, meaning that there is neither a centralized authority scheduling the transmissions in the network nor coordination among the users. It follows that the communications of the N pairs may occur simultaneously; this implies that, in addition to the desired signal, each user receives also the signal transmitted by the other $N-1$ pairs, which is an undesired signal, termed as Multi-User Interference (MUI). Stated in mathematical terms, the quality of the transmission of each pair i over the channel n is measured by the Signal-to-Noise-plus-Interference ratio (SINR):

$$\mathrm{SINR}_n^i(p_n^i, p_n^{-i}) \triangleq \frac{|H_{ii}(n)|^2 \, p_n^i}{(\sigma_n^i)^2 + \sum_{j \neq i} |H_{ij}(n)|^2 \, p_n^j}, \tag{9}$$

where $|H_{ii}(n)| > 0$ is the channel gain of pair i over the frequency band n, and $|H_{ij}(n)| \geq 0$ is the (cross-)channel gain between the transmitter j and the receiver i; ${\sigma_n^i}^2$ is the power spectral density (PSD) of the noise at receiver i over the band n; and $p_n^{-i} \triangleq \left(p_n^1, \ldots, p_n^{i-1}, p_n^{i+1}, \ldots, p_n^Q \right)$ is the set of all the users power allocations over the channel n, except the ith one. The useful power signal of pair i over the channel n is thus $|H_{ii}(n)|^2 \, p_n^i$, whereas $\sum_{j \neq i} |H_{ij}(n)|^2 p_n^j$ is the PSD of MUI measured by the receiver i over the channel n. The overall performance of each transmission i is usually measured in terms of the maximum achievable information rate $r_i(p^i, p^{-i})$ over the set of the C parallel channels, which depends on the power allocation of all the users (p^i, p^{-i}). Under basic information theoretical assumptions and given the users' power allocation profile $p^1, \ldots, p^N$, this rate is given by

$$r_i(p^i, p^{-i}) \triangleq \sum_{n=1}^{C} \log\left(1 + \mathrm{SINR}_n^i(p_n^i, p_n^{-i})\right). \tag{10}$$

Corresponding to the above player's problem, we can formulate the system design as a game: the aim of each player i, given the strategy profile p^{-i} of the others, is to choose a feasible power allocation p^i that maximizes the rate $r_i(p^i, p^{-i})$, i.e.,

$$\underset{p^i \in \mathcal{P}^i}{\text{maximize}}\; r_i(p^i, p^{-i}) \tag{11}$$

for all $i = 1, \ldots, N$, where $\mathcal{P}^i$ and $r_i(p^i, p^{-i})$ are defined in (8) and (10), respectively.

A problem with the above game-theoretic formulation is that it could result in some players having a very low transmission rate (i.e. objective function value) while other players could have a much higher one. This is, in many situations undesirable. To add to the model some kind of "fairness" we can modify the above formulation in the following way. It is clear that the main source deterioration of the transmission rate is the MUI: all other things being fixed, the higher the MUI the lower the transmission rate. A measure of the MUI received by one player is given by $g_i(p^i, p^{-i}) \triangleq \sum_{n=1}^{C}\left[\sum_{j\neq i}|H_{ij}(n)|^2 p_n^j\right]$. We set

$$\psi(p) \triangleq \max_{1\leq i\leq N} g_i(p).$$

The function $\psi(p^i, p^{-i})$ is therefore a measure of the maximum MUI a player can experience. Thus there is a mutual interest of players to keep the value of this function low. It is then sensible to modify player i's objective function and solve the following optimization problem

$$\underset{p^i \in \mathcal{P}^i}{\text{maximize}}\; \left[w_i\, r_i(p^i, p^{-i}) - \psi(p^i, p^{-i})\right], \tag{12}$$

where w_i are nonnegative scalars that weight the "interest" of player i in minimizing the worst case MUI for all players. Since

$$\begin{aligned} &w_i\, r_i(p^i, p^{-i}) - \psi(p^i, p^{-i}) \\ &\quad = w_i \log\left(1 + \frac{|H_{ii}(n)|^2\, p_n^i}{(\sigma_n^i)^2 + \sum_{j\neq i}|H_{ij}(n)|^2\, p_n^j}\right) - \max_{1\leq \ell\leq N}\sum_{n=1}^{C}\left[\sum_{j\neq \ell}|H_{\ell j}(n)|^2 p_n^j\right], \end{aligned}$$

it can be seen that the problem (12) has the structure of (1), falls perfectly in the framework described in the first part of this section and, in particular, satisfies Assumption 2.

3 The pulled-out version of the games $\mathcal{G}_{\mathrm{mm}}^{\mathrm{C,F}}$

In this section we describe a smooth reformulation of the minmax games. We first consider the continuous version. The key to this approach is the following result establishing a connection between this game and a related game $\widehat{\mathcal{G}}_{\mathrm{mm}}^{\mathrm{C}}$ with $N+1$ players, wherein player $i = 1, \ldots, N$, anticipating $x^{-i} \in X_{-i}$ and $y \in Y$, solves the following optimization problem:

$$\underset{x^i \in X_i}{\text{minimize}} \left[f_i(x^i, x^{-i}) + g(x^i, x^{-i}, y) \right], \tag{13}$$

and the $(N+1)$-st player, anticipating x, solves the optimization problem:

$$\underset{y \in Y}{\text{maximize}} \; g(x, y). \tag{14}$$

The result below does not require any assumption.

Theorem 1 *If (x^*, y^*) is a Nash equilibrium of the game $\widehat{\mathcal{G}}_{\mathrm{mm}}^{\mathrm{C}}$, then x^* is a Nash equilibrium of the game $\mathcal{G}_{\mathrm{mm}}^{\mathrm{C}}$.*

Proof It suffices to show that for each $i = 1, \ldots, N$,

$$x^{*,i} \in \underset{x^i \in X_i}{\operatorname{argmin}} \left[f_i(x^i, x^{*,-i}) + \max_{y \in Y} g(x^i, x^{*,-i}, y) \right].$$

Let $x^i \in X_i$ be arbitrary. We have

$$\begin{aligned} f_i(x^i, x^{*,-i}) + \max_{y \in Y} g(x^i, x^{*,-i}, y) &\geq f_i(x^i, x^{*,-i}) + g(x^i, x^{*,-i}, y^*) \\ &\geq f_i(x^*) + g_i(x^*, y^*) \\ &= f_i(x^{*,i}, x^{*,-i}) + \max_{y \in Y} g(x^*, y). \end{aligned}$$

This completes the proof of the theorem. □

Remark 1 Although we presented the above result for the continuous minmax game wherein the function $g(x, y)$ and the set Y are the same for all players, the pulled-out game is actually applicable to the situation where the pair (g, Y) is different among the players. More accurately, consider a game where the ith player's problem is

$$\underset{x^i \in X_i}{\text{minimize}} \left[f_i(x^i, x^{-i}) + \max_{y^i \in Y_i} \; g_i(x, y^i) \right], \tag{15}$$

with each max function $\max_{y^i \in Y_i} g_i(x, y^i)$ satisfying Assumption 1. Following the same lines as in the proof of Theorem 1, it is easy to see that each solution of an extended game with $2N$ players wherein player $i = 1, \ldots, N$, anticipating $x^{-i} \in X_{-i}$ and $y^i \in Y_i$, solves the following optimization problem:

$$\underset{x^i \in X_i}{\text{minimize}} \left[f_i(x^i, x^{-i}) + g_i(x^i, x^{-i}, y^i) \right],$$

and each of the last N players, anticipating x, solves the optimization problems:

$$\underset{y^i \in Y_i}{\text{maximize}}\ g_i(x, y^i),$$

is also a solution of the original game. Unfortunately, the extended game is not easily amenable to solution by distributed algorithms because a key norm condition commonly needed for the convergence of such an algorithm [12] does not readily hold for this game. Therefore at this time we do not pursue further the study of the extended game in this more general setting. $\square$

Remark 2 Formally, it is not difficult to extend the player-dependent min-(single)-max game $\mathcal{G}_{\text{mm}}^{\text{C}}$ to a min-multi-max game in which each player i's objective is the composite of several pointwise maximum functions. Specifically, suppose that player i's optimization problem is:

$$\begin{aligned} \underset{x_i \in X_i}{\text{minimize}} \Big[& f_i(x^i, x^{-i}) \\ & + \psi_i \left(\max_{y^{i1} \in Y_{i1}} g_{i1}(x^i, x^{-i}, y^{i1}), \ \ldots, \ \max_{y^{iK_i} \in Y_{iK_i}} g_{iK_i}(x^i, x^{-i}, y^{iK_i}) \right) \Big]; \end{aligned}$$

for some positive integer K_i. One can easily derive an extended game with $\sum_{i=1}^{N} K_i$ additional players whose Nash equilibria will be Nash equilibria of the given game, provided that the functions ψ_i are *isotone* in its arguments, i.e., $\psi_i(t_1, \ldots, t_{K_i}) \geq \psi_i(t'_1, \ldots, t'_{K_i})$ whenever $(t_1, \ldots, t_{K_i}) \geq (t'_1, \ldots, t'_{K_i})$. Since the resulting extended game is still not easily amenable to distributed computation, we omit further discussion of this extension too. $\square$

The result in Theorem 1 goes only in one direction and does not state that every solution of $\mathcal{G}$ is a solution, together with a suitable y, of $\widehat{\mathcal{G}}$. This reverse implication is not actually true in general, as shown by the following example.

Example 1 Consider a game with two players, each controlling one variable, x_1 and x_2 respectively.

$$\begin{aligned} &\underset{x_1 \in \mathbb{R}}{\text{minimize}}\ x_1 + \max_{y \in \Delta} \left[y_1(x_1 - x_2) + y_2(-x_1 + x_2) \right], \\ &\underset{x_2 \in \mathbb{R}}{\text{minimize}}\ x_2 + \max_{y \in \Delta} \left[y_1(x_1 - x_2) + y_2(-x_1 + x_2) \right]. \end{aligned}$$

Note that

$$\max_{y\in\Delta}\,[\,y_1(x_1-x_2)+y_2(-x_1+x_2)]\;=\;\max\{x_1-x_2,-x_1+x_2\},$$

an observation that facilitates the calculation of partial subdifferentials in what follows. It is easy to check that $(0,0)$ is a solution of this game, since in both cases the subdifferential of the objective functions is $1+[-1,1]$ and contains 0. If we look at the extended game we must add a third player, controlling $y=(y_1,y_2)\in\Delta$:

$$\begin{aligned}&\underset{x_1}{\text{minimize}}\;\;[\,x_1+y_1(x_1-x_2)+y_2(-x_1+x_2)\,],\\&\underset{x_2}{\text{minimize}}\;\;[\,x_2+y_1(x_1-x_2)+y_2(-x_1+x_2)\,],\\&\underset{y\in\Delta}{\text{minimize}}\;\;[\,y_1(x_1-x_2)+y_2(-x_1+x_2)\,].\end{aligned}$$

It is easy to see that there is no $y^*\in\Delta$ such that $(0,0,y^*)$ is a Nash equilibrium of this game. In fact for 0 to be an optimal solution of the first player's problem we must have $y^*=(0,1)$, while for 0 to be an optimal solution of the second player's problem we must have $y^*=(1,0)$. □

The point is that in a solution of $\mathcal{G}^{\mathrm{C}}_{\mathrm{mm}}$ each player could "use" a different maximizing y, a problem that can arise only if at an equilibrium x^* the maximum of $g(x^*,y)$ with respect to y is not uniquely attained. It should then be reasonable to expect that in case the maximum is actually reached for only one y, for example if $g(x,\cdot)$ is strictly concave, the original game and the extended one should be equivalent. This conjecture turns out to be true; the following proposition goes a step further in this direction and gives a more relaxed condition (labeled (d) in the result below) under which the game $\mathcal{G}^{\mathrm{C}}_{\mathrm{mm}}$ is equivalent to its pulled-out version $\widehat{\mathcal{G}}^{\mathrm{C}}_{mm}$.

Proposition 1 *Under the following assumptions:*

(a) *the function g is continuously differentiable on $\Omega\times\Upsilon$, and Assumption 1 holds;*
(b) *for all $i=1,\ldots,N$ and all $x^{-i}\in X_{-i}$, the function $f_i(\bullet,x^{-i})$ is continuously differentiable on Ω_i;*
(c) *for all $i=1,\ldots,N$, $x^{-i}\in X_{-i}$, and $y\in Y$, the function $f_i(\bullet,x^{-i})+g(\bullet,x^{-i},y)$ is convex on X_i;*
(d) *$\nabla_x g(x^*,\bullet)$ is constant on $\underset{y\in Y}{\operatorname{argmax}}\; g(x^*,y)$;*

if x^* is a Nash equilibrium of the game $\mathcal{G}^{\mathrm{C}}_{\mathrm{mm}}$, then (x^*,y^*) is a Nash equilibrium of the game $\widehat{\mathcal{G}}^{\mathrm{C}}_{\mathrm{mm}}$ for any $y^*\in\underset{y\in Y}{\operatorname{argmax}}\; g(x^*,y)$.

Proof Set for simplicity $\theta_i(x)\triangleq f_i(x)+\max_{y\in Y} g(x,y)$. It follows from the well-known Danskin Theorem, see e.g. [11, Theorem 10.2.1], that the function θ_i is differentiable at x^* and

$$\nabla\theta_i(x^*)\;=\;\nabla f_i(x^*)+\nabla_x g(x^*,y^*).$$

Since $x^{*,i} \in \underset{x^i \in X_i}{\operatorname{argmin}}\ \theta_i(x^i, x^{*,-i})$, it follows that

$$(x^i - x^{*,i})^T \left[\nabla_{x^i} f_i(x^*) + \nabla_{x^i} g_i(x^*, y^{*,i}) \right] \geq 0, \quad \forall\, x^i \in X_i,$$

which shows that $x^{*,i} \in \underset{x^i \in X_i}{\operatorname{argmin}} \left[f_i(x^i, x^{*,-i}) + g_i(x^i, x^{*,-i}, y^{*,i}) \right]$. Thus, the pair (x^*, y^*) is a Nash equilibrium of the game $\widehat{\mathcal{G}}^{\mathrm{C}}_{\mathrm{mm}}$. □

In spite of its conceptual interest, condition (d) introduced in the previous proposition implies the differentiability of the max function ϕ at a solution, a fact that shows that when ϕ is genuinely nondifferentiable we should not expect $\mathcal{G}^{\mathrm{C}}_{\mathrm{mm}}$ and $\widehat{\mathcal{G}}^{\mathrm{C}}_{\mathrm{mm}}$ to be equivalent. The advantage of the reduction of the game $\mathcal{G}^{\mathrm{C}}_{\mathrm{mm}}$ to $\widehat{\mathcal{G}}^{\mathrm{C}}_{\mathrm{mm}}$ is that the latter, under appropriate assumptions, is a monotone game, meaning that its first-order formulation is a monotone VI that can therefore be solved by suitable distributed methods. To explain this, we assume that all the functions f_i and g are twice continuously differentiable and define the following vector function

$$F(x, y) \triangleq \begin{pmatrix} (\nabla_{x^i} f_i(x))_{i=1}^N + \nabla_x g(x, y) \\ -\nabla_y g(x, y) \end{pmatrix}$$

It is clear that VI(S, F), where $S \triangleq \left(\prod_{i=1}^N X_i \right) \times Y$, represents the first-order conditions of the game $\widehat{\mathcal{G}}^{\mathrm{C}}_{\mathrm{mm}}$, i.e., the concatenation of the first-order optimality conditions of each of the $(N+1)$-player's optimization problems. To check the monotonicity of F we examine the positive semidefiniteness of its Jacobian which is given by

$$JF(x, y) = \begin{bmatrix} A(x, y) & J_y(\nabla_x g(x, y)) \\ -J_x(\nabla_y g(x, y)) & 0 \end{bmatrix},$$

where

$$A(x, y) \triangleq \begin{bmatrix} \nabla^2_{x^1 x^1} f_1(x) & \nabla^2_{x^1 x^2} f_1(x) & \cdots & \nabla^2_{x^1 x^N} f_1(x) \\ \vdots & & \ddots & \vdots \\ \nabla^2_{x^N x^1} f_N(x) & \nabla^2_{x^N x^2} f_N(x) & \cdots & \nabla^2_{x^N x^N} f_N(x) \end{bmatrix} + \nabla^2_{xx} g(x, y),$$

Since $J_y(\nabla_x g(x, y)) = \big(J_x(\nabla_y g(x, y))\big)^T$, it is clear that $JF(x, y)$ is *bisymmetric*, and therefore F is monotone if and only if $A(x, y)$ is positive semidefinite for every $(x, y) \in X \times Y$. The following theorem does not require a proof.

Theorem 2 *Suppose all f_i and g are twice continuously differentiable and scalars σ^f and σ^ϕ exist such that the mapping $(\nabla_{x^i} f_i(x))_{i=1}^N$ is σ^f-monotone and the mapping $\nabla_x g(\bullet, y)$ is σ^g-monotone on X for all $y \in Y$. If $\gamma \triangleq \sigma^g + \sigma^f \geq 0$, then F is monotone on S.* □

If $g(x, y) \triangleq 0$ and we have a smooth Nash equilibrium problem, the key condition about the mapping $(\nabla_{x^i} f_i(x))_{i=1}^N$ simply becomes the requirement that this function be monotone, i.e. a setting under which the game can be solved by distributed methods via regularization [36] (and, essentially, also by centralized algorithms). So, the theorem above neatly extends this condition to our minmax setting. Note that in order for the positive semidefiniteness of $A(x, y)$ to hold, it is not necessary for either of the two summands in the definition of $A(x, y)$ to be semidefinite. In particular $\nabla^2_{xx} g(x, y)$ need not be positive semidefinite, i.e., $g(x, y)$ need not be convex in x.

In case the function ϕ is a finite minimax, we can use its representation as a continuous minmax over the unit simplex Δ. The function F becomes

$$F(x, y) \triangleq \begin{pmatrix} \left(\nabla_{x^i} f_i(x) + \sum_{j \in J} y_j \nabla_{x^i} g_j(x) \right)_{i=1}^N \\ -(g_j(x))_{j \in J} \end{pmatrix}$$

and

$$\begin{aligned} A(x, y) \triangleq & \begin{bmatrix} \nabla^2_{x^1 x^1} f_1(x) & \nabla^2_{x^1 x^2} f_1(x) & \cdots & \nabla^2_{x^1 x^N} f_1(x) \\ \vdots & & \ddots & \vdots \\ \nabla^2_{x^N x^1} f_N(x) & \nabla^2_{x^N x^2} f_N(x) & \cdots & \nabla^2_{x^N x^N} f_N(x) \end{bmatrix} \\ & + \sum_{j \in J} y_j \begin{bmatrix} \nabla^2_{x^1 x^1} g_j(x) & \nabla^2_{x^1 x^2} g_j(x) & \cdots & \nabla^2_{x^1 x^N} g_j(x) \\ \vdots & & \ddots & \vdots \\ \nabla^2_{x^N x^1} g_j(x) & \nabla^2_{x^N x^2} g_j(x) & \cdots & \nabla^2_{x^N x^N} g_j(x) \end{bmatrix}, \end{aligned}$$

which very neatly shows the interplay between the f_i and the g_j and from which the desired positive semidefiniteness of $A(x, y)$ follows readily.

Under the conditions stated in Theorem 2 we can solve the games $\widehat{\mathcal{G}}^{\mathrm{C,F}}_{\mathrm{mm}}$, and therefore $\mathcal{G}^{\mathrm{C,F}}_{\mathrm{mm}}$, by solving the monotone VI(S, F). Plenty of centralized algorithms are available to accomplish this task; see [11]. The development of distributed algorithms is less immediate though, as can be seen from the work [36] that we briefly summarize here for completeness. In order to apply these methods we need F to be strongly monotone; additionally, we need a way to "compensate" for the cross-influences between players. The idea is then to use the proximal-point method to solve VI(S, F); see [11]. By this approach we reduce the solution of VI(S, F) to that of a sequence of strongly monotone VIs. More precisely, given an iterate $z^\nu \triangleq (x^\nu, y^\nu)$, we generate the new iterate $z^{\nu+1}$ by solving (possibly approximatively) the strongly monotone VI$(S, F + \rho(\bullet - z^\nu))$, where ρ is a positive parameter that can be chosen to be arbitrary large, in principle. It can then be shown that, by choosing ρ appropriately, the latter VI can be solved in a distributed manner using the techniques described in [36], thus yielding the desired distributed method for the solution of the games $\mathcal{G}^{\mathrm{C,F}}_{\mathrm{mm}}$.

Example 2 We consider a game of the type

$$\underset{x^i \in X_i}{\text{minimize}} \left[f_i(x^i, x^{-i}) + c\|x\|_N \right]; \tag{16}$$

i.e. we take $\phi = c\| \cdot \|_N$, where c is a positive constant and $\| \cdot \|_N$ denotes a given norm. For example, if we consider the ℓ_1-norm, $\|x\|_1 \triangleq \sum_{i=1}^{n} |x_i|$, this might be viewed as an attempt *for each player* to find a *sparse* response i.e., an (optimal) response with as many zeros as possible to a given x^{-i}. If a different norm is used, this example is mostly relevant in the case in which we are "designing the game", as discussed at the beginning of Sect. 2.1.

In general, consider the dual norm $\| \bullet \|_D$, defined by

$$\|x\|_D \triangleq \max_{\|y\|_N \le 1} x^T y.$$

It is well known that the dual of the dual norm is again the original norm; thus we can also write

$$\|x\|_N = \max_{\|y\|_D \le 1} x^T y.$$

Therefore we see that the game (16) falls in our general framework and the extended game $\widehat{\mathcal{G}}$ in this case reads

$$\underset{x^i \in X_i}{\text{minimize}} \left[f_i(x^i, x^{-i}) + c\,(x^i)^T y, \right], \quad i = 1, \ldots, N; \qquad \underset{y:\|y\|_D \le 1}{\text{maximize}}\ x^T y.$$

In many practical cases the $(N+1)$st problem can be solved analytically. For example, if $\| \cdot \|_N$ is the ℓ_1-norm the dual norm is the ℓ_∞-norm, so that the maximum of $x^T y$ in the $(N+1)$st problem is attained by y being the vector of all zeros except for the component corresponding to (one of) the largest component(s), in absolute value, of x. More precisely if the largest component of x is positive, the corresponding component in y is 1 otherwise it is -1. Another case of interest is when the norm $\| \bullet \|_N$ is the Euclidean norm. Since the Euclidean norm is self-dual, the $(N+1)$st problem becomes $\underset{y:\|y\|_2 \le 1}{\text{maximize}}\ x^T y$ whose solution, in the non-trivial case in which $x \neq 0$, is easily seen to be $-x/\|x\|_2$. □

4 A smoothing approach

We consider now a different approach, based on smoothing. Smoothing is a conceptually simple idea that is often used in optimization, usually to very good numerical effect. Recently smoothing has been the focus of renewed interest, see for example [3,5] and the references therein, and has also been investigated in the broader of approximations of Nash equilibria in general; see for example [16,20]. Our approach differs from previous ones in that, on the one hand it uses stronger assumptions than in [16,20], but on the other hand it permits the development of practical diagonalization numerical schemes that solve optimization subproblems (instead of subgames as in the references). We assume that the nonsmooth function $\phi(x)$ can be approximated by a smooth function $\phi(x; s)$, where $s \in S \subseteq \mathbb{R}^s$ is a smoothing parameter, with S a set containing the origin. Informally, we expect that $\phi(x; 0)$ coincides with $\phi(x)$ and that the closer s is to 0, the better $\phi(x; s)$ approximates $\phi(x)$. As in [16,20], the preliminary idea underlying the smoothing approach is then to solve a sequence of games $\mathcal{G}(s_\nu)$, where player i, anticipating x^{-i}, solves the optimization problem:

$$\underset{x^i \in X_i}{\text{minimize}} \left[f_i(x^i, x^{-i}) + \phi(x^i, x^{-i}; s_\nu) \right]. \tag{17}$$

Provided that the convergence of the approximation $\phi(x, s_\nu)$ to $\phi(x)$ as $s_\nu \downarrow 0$ is "uniform", e.g., as implied by Assumption 3, then the results in [16,20] will ensure the convergence of the Nash equilibria of the family of games $\{\mathcal{G}(s_\nu)\}$ to a Nash equilibrium of the original game $\mathcal{G}$. While this is a reasonable convergence result in theory and provides the foundation for our discussion below, this sequential smoothed subgame method is at best a conceptual step toward the distributed resolution of the game $\mathcal{G}$. In what follows, we present a *semi-diagonalization approach*, wherein we avoid solving such sub-games and instead solve convex subprograms whose generation is in sync with the change of the approximation scalar s_ν. The adjective "semi" signifies that we only "diagonalize" the given smooth functions $f_i(x)$, leaving the artificially smoothed function $\phi(x; s)$ intact in forming the convex subprograms. This is in contrast to a full diagonalization wherein we also diagonalize $\phi(x; s)$. As it turns out, we can establish the convergence of the semi-diagonalization scheme whereas it seems difficult to establish the convergence of the full-diagonalization scheme under realistic assumptions.

To lay the foundation for the semi-diagonalization idea, assume that a point $z \in X$ is given and consider the following optimization problem:

$$\underset{x \in X}{\text{minimize}} \quad \phi(x; s) + \sum_{i=1}^{N} f_i(x^i, z^{-i}). \tag{18}$$

Let $T_s : X \to X$ be the multifunction that maps z to set of optimal solutions of (18). A fixed point of T_s is a vector z such that

$$\phi(z; s) + \sum_{i=1}^{N} f_i(z^i, z^{-i}) \leq \phi(x; s) + \sum_{i=1}^{N} f_i(x^i, z^{-i}), \quad \forall x \in X.$$

The result below establishes that a fixed point of T_s is a solution of the game $\mathcal{G}(s)$ wherein each player i, anticipating x^{-i}, solves the optimization problem (17) with s_ν replaced by s.

Proposition 2 *Any fixed point of T_s is a solution of the game $\mathcal{G}(s)$.*

Proof Suppose by contradiction that z is a fixed point of T_s but not a Nash equilibrium of (17). Then there exist an i and $y^i \in X_i$ such that

$$f_i(y^i, z^{-i}) + \phi(y^i, z^{-i}; s) \; < \; f_i(z^i, z^{-i}) + \phi(z^i, z^{-i}; s).$$

But then, since $(y^i, z^{-i}) \in X$, we have

$$\phi(y^i, z^{-i}, s) + f_i(y^i, z^{-i}) + \sum_{j \neq i} f_j(z^j, z^{-j}) \; < \; \phi(z, s) + \sum_{j=1}^{N} f_j(z^j, z^{-j}),$$

contradicting the fact that z is a fixed point of T_s. □

We remark that the result above holds also for $s = 0$, that is for the map $T : X \to X$, where for $z \in X$, $T(z)$ is the set of minimizers of

$$\operatorname*{minimize}_{x \in X} \quad \phi(x) + \sum_{i=1}^{N} f_i(x^i, z^{-i}). \tag{19}$$

The basic idea is then to solve a sequence of optimization problems (18) for s tending to zero. Note also that, parallel to a similar result for the extended game approach, the reverse of Proposition 2 does not hold in general, as shown by the following example.

Example 3 Consider a game where with two players, one controlling x_1 and the other x_2. Assume that $\phi(x_1, x_2) \triangleq (x_1)^2 + (x_2)^2 + |x_1 - x_2|$, $f_1 \equiv 0 \equiv f_2$ and $X_1 = \mathbb{R} = X_2$. Note that ϕ is strongly convex jointly in both variables and has the origin as its unique minimum. However, the corresponding game, where the two players solve $\operatorname*{minimize}_{x_1} \phi(x_1, x_2)$ and $\operatorname*{minimize}_{x_2} \phi(x_1, x_2)$ respectively, has a continuum of Nash equilibria, given by $x_1 = \alpha = x_2$ with $\alpha \in [-1/2, 1/2]$. □

At the time of this writing, we do not know whether fixed points of T_s can be viewed as a "refinement" of the Nash equilibria of game $\mathcal{G}$; we believe that the answer is negative in general, i.e. there are cases where $\mathcal{G}$ has solutions but T_s has no fixed point. We refer the interested reader to [28] for more discussion of the notion of "refinement" of game equilibria.

By Proposition 2, showing that a fixed point of T_s is an equilibrium of the game $\mathcal{G}(s)$, it is natural to conjecture that a fixed-point iteration, where the smoothing parameter s is driven to 0, could provide, under adequate assumptions, a convergent scheme for the solution of (1). The resulting (centralized) algorithm is described next.

Algorithm 1 Let a sequence $\{s_\nu\} \subset S \setminus \{0\}$ converging to 0 be given. Starting with $x^0 \in X$, define a sequence $\{x^\nu\}$ by letting $x^{\nu+1}$ be a solution of

$$\operatorname*{minimize}_{x \in X} \left[\phi(x, s_\nu) + \sum_{i=1}^{N} f_i(x^i, x^{\nu, -i}) \right]. \tag{20}$$

In practical implementation, the iteration is terminated under a prescribed termination criterion. □

In the classical case in which $\phi \equiv 0$ this algorithm calls, in (20), for the solution of N decoupled convex optimization problems and reduces to the standard best-response approach for the solution of the (smooth) game wherein player i minimizes $f_i(\bullet, x^{-i})$ over X_i; see for example [12,36] and the references therein. Such a best-response method is today well understood. In contrast to (20), the full-diagonalization scheme would solve the the following problem:

$$\underset{x \in X}{\text{minimize}} \sum_{i=1}^{N} \left[\phi(x^i, x^{\nu,-i}, s_\nu) + f_i(x^i, x^{\nu,-i}) \right], \tag{21}$$

which completely separates into N independent subproblems, with the ith one being $\underset{x^i \in X_i}{\text{minimize}} \left[\phi(x^i, x^{\nu,-i}, s_\nu) + f_i(x^i, x^{\nu,-i}) \right]$. The different level of separability leads to a major operational difference between the two diagonalization schemes, with the semi-scheme solving subproblems of the kind (20) while the full-scheme solving (21). Theoretically, the full-diagonalization scheme turns out to be overly simplistic versus the former method of semi-diagonalization. For one thing, we are not able to establish the convergence of the so-generated sequence of iterates in the full-diagonalization scheme under reasonable assumptions, such as Assumption 3 below. In contrast, while (20) is a non-separable, albeit convex optimization problem, its convergence can be established (see Theorem 3) under the following:

Assumption 3 For every $s \in S \setminus \{0\}$, $\phi(\bullet; s)$ is continuously differentiable; moreover the following two properties hold:

(a) $\lim_{s \downarrow 0} [\phi(x) - \phi(x; s)] = 0$ uniformly in x;
(b) a positive constant σ_a exists such that $\phi(\bullet; s)$ is σ_a-convex on Ω for every $s \in S \setminus \{0\}$. □

Assumption 3(a) simply says that $\phi(x; s)$ is a uniform approximation of $\phi(x)$ as s approaches zero. A sufficient condition for this assumption to hold is that there exist positive constants C and ς such that $\|\phi(x) - \phi(x; s)\| \leq C\|s\|^\varsigma$ for all $x \in \Omega$ and all $s \in S \setminus \{0\}$. This stronger requirement will later be used to verify the satisfaction of Assumption 3(a) for some specific approximation functions $\phi(x; s)$. Assumption 3(b) stipulates that the approximation must preserve the σ-convexity property of the approximated function $\phi(\bullet; s)$, with a possibly different constant. A simple continuity argument shows that under Assumption 3(b), ϕ itself is σ_a-convex. Bearing this in mind, condition (22) below is easily seen to be just a slight reinforcement of Assumption 2(a).

Proposition 3 *Suppose that for each $i = 1, \ldots, N$ and each fixed $x^{-i} \in \Omega_{-i}$, $f_i(\bullet, x^{-i})$ is σ_i^f-convex on Ω_i and Assumption 3 holds. Assume further that*

$$\gamma \triangleq \sigma_a + \min_i \sigma_i^f > 0. \tag{22}$$

The following two statements are valid for every $x^\nu \in X$:

(a) *Both problems below have unique solutions, which we denote by $x^\nu(s)$ and $x^\nu(0)$, respectively:*

$$\underset{x \in X}{\text{minimize}} \;\; \phi(x; s) + \sum_{i=1}^{N} f_i(x^i, x^{\nu,-i}) \tag{23}$$

and

$$\underset{x \in X}{\text{minimize}} \;\; \phi(x) + \sum_{i=1}^{N} f_i(x^i, x^{\nu,-i}). \tag{24}$$

(b) $\lim_{s \downarrow 0} \left[x^\nu(0) - x^\nu(s) \right] = 0.$

If in addition there exist positive constants C and ς such that for all $x \in X$, $\|\phi(x) - \phi(x; s)\| \le C\|s\|^\varsigma$, then

$$\| x^\nu(0) - x^\nu(s) \| \;\le\; 2\sqrt{\frac{C}{\gamma}}\sqrt{\|s\|^\varsigma}\,. \tag{25}$$

Proof Observe first that by (22), $\phi(\bullet; s) + \sum_{i=1}^{N} f_i(\bullet, x^{\nu,-i})$ is a strongly convex function with constant $\gamma > 0$ and therefore (23) has a unique solution. By the same token, since, as observed just before the theorem, $\phi(x)$ is σ_a-convex, also (24) has a unique solution. Since $x^\nu(s)$ is an optimal solution of (23), we have, by strong convexity,

$$\begin{aligned} &\phi(x^\nu(0); s) + \sum_{i=1}^{N} f_i(x^{\nu,i}(0), x^{\nu,-i}) \\ &\quad \ge \phi(x^\nu(s), s) + \sum_{i=1}^{N} f_i(x^{\nu,i}(s), x^{\nu,-i}) + \frac{\gamma}{2} \| x^\nu(0) - x^\nu(s) \|^2. \end{aligned}$$

Therefore

$$\begin{aligned} &\frac{\gamma}{2} \| x^\nu(0) - x^\nu(s) \|^2 \\ &\le \left[\phi(x^\nu(0); s) + \sum_{i=1}^{N} f_i(x^{\nu,i}(0), x^{\nu,-i}) \right] - \left[\phi(x^\nu(s); s) + \sum_{i=1}^{N} f_i(x^{\nu,i}(s), x^{\nu,-i}) \right] \\ &= \left[\phi(x^\nu(0); s) + \sum_{i=1}^{N} f_i(x^{\nu,i}(0), x^{\nu,-i}) \right] - \left[\phi(x^\nu(0)) + \sum_{i=1}^{N} f_i(x^{\nu,i}(0), x^{\nu,-i}) \right] \end{aligned}$$

$$
\begin{aligned}
&+\left[\phi(x^{\nu}(0)) + \sum_{i=1}^{N} f_i(x^{\nu,i}(0), x^{\nu,-i})\right] - \left[\phi(x^{\nu}(s); s) + \sum_{i=1}^{N} f_i(x^{\nu,i}(s), x^{\nu,-i})\right] \\
&= \phi(x^{\nu}(0); s) - \phi(x^{\nu}(0)) \\
&+\left[\phi(x^{\nu}(0)) + \sum_{i=1}^{N} f_i(x^{\nu,i}(0), x^{\nu,-i})\right] - \left[\phi(x^{\nu}(s); s) + \sum_{i=1}^{N} f_i(x^{\nu,i}(s), x^{\nu,-i})\right].
\end{aligned}
$$

Based on the last inequality, we prove the last claim of the proposition and omit the proof of statement (b) as the latter proof is similar.

We have

$$
\left|\, \phi(x^{\nu}(0); s) - \phi(x^{\nu}(0)) \,\right| \le C \,\|\, s \,\|^{\varsigma}. \tag{26}
$$

To bound the term:

$$
\left[\phi(x^{\nu}(0)) + \sum_{i=1}^{N} f_i(x^{\nu,i}(0), x^{\nu,-i})\right] - \left[\phi(x^{\nu}(s); s) + \sum_{i=1}^{N} f_i(x^{\nu,i}(s), x^{\nu,-i})\right],
$$

note that the first term in square brackets is the minimum value of the function $\phi + \sum_{i=1}^{N} f_i(\bullet, x^{\nu,-i})$ over X, while the second term is the minimum value of the function $\phi(\bullet; s) + \sum_{i=1}^{N} f_i(\bullet, x^{\nu,-i})$ over X. Since these two functions differ pointwise by at most $C\|s\|^{\varsigma}$, it follows that

$$
\begin{aligned}
&\left|\left[\phi(x^{\nu}(0)) + \sum_{i=1}^{N} f_i(x^{\nu,i}(0), x^{\nu,-i})\right]\right. \\
&\left. - \left[\phi(x^{\nu}(s), s) + \sum_{i=1}^{N} f_i(x^{\nu,i}(s), x^{\nu,-i})\right]\right| \le C \,\|\, s \,\|^{\varsigma}.
\end{aligned} \tag{27}
$$

The desired bound (25) follows readily from (26) and (27). □

Part (b) of Proposition 3 shows that the (single-valued) map T_s converges pointwise to T on X as s tends to zero; i.e.,

$$
\lim_{s \downarrow 0} \left[\, T_s(z) - T(z) \,\right] = 0, \quad \text{for every } z \in X. \tag{28}
$$

Based on this fact we can now prove convergence of Algorithm 1 in Theorem 3 below. For this purpose, we define a key $N \times N$ matrix:

$$\Gamma \triangleq \begin{bmatrix} 0 & L_{12} & \cdots & L_{1N} \\ L_{21} & 0 & \cdots & L_{2N} \\ \vdots & & \ddots & \vdots \\ L_{N1} & L_{N2} & & 0 \end{bmatrix},$$

where, we recall, L_{ij} are Lipschitz constants defined in Assumption 2. In essence the matrix Γ describes the interaction between players: the higher the influence of the jth player's variables x^j on the i-th player objective function f_i, the higher L_{ij}. We let $\|\Gamma\|$ denote the Euclidean norm of Γ (induced by the ℓ_2-norm of vectors in $\mathbb{R}^N$).

Theorem 3 *Assume the setting of Proposition 3 and that $\|\Gamma\| < \gamma$. Then the sequence $\{x^\nu\}$ produced by Algorithm 1 converges to a solution of the game $\mathcal{G}$.*

Proof We first prove that, for any $\|s\| \in S \setminus \{0\}$, every T_s is a contraction with a common contraction constant given by $\beta \triangleq \|\Gamma\|/\gamma$, which does not depend on s. Let $0 \neq s \in S$ be fixed but arbitrary. By the minimum principle, $x(z) \triangleq T_s(z)$ satisfies

$$(y - x(z))^T \nabla_x \phi(x(z); s) + \sum_{i=1}^N (y^i - x^i(z))^T \nabla_{x^i} f_i(x^i(z), z^{-i}) \geq 0, \quad \forall y \in X.$$

Similarly we can write, for z',

$$(y - x(z'))^T \nabla_x \phi(x(z'); s) + \sum_{i=1}^N (y^i - x^i(z'))^T \nabla_{x^i} f_i(x^i(z'), z'^{,-i}) \geq 0, \quad \forall y \in X.$$

By taking $y = x(z')$ in the first inequality and $y = x(z)$ in the second one and adding them we get

$$\begin{aligned} &(x(z') - x(z))^T \left[\nabla_x \phi(x(z); s) - \nabla_x \phi(x(z'); s) \right] + \\ &\sum_{i=1}^N (x^i(z') - x^i(z))^T \left[\nabla_{x^i} f_i(x^i(z), z^{-i}) - \nabla_{x^i} f_i(x^i(z'), (z')^{-i}) \right] \geq 0. \end{aligned}$$

By recalling the σ_a-convexity assumption and (4), we can then write

$$\begin{aligned} &\sigma_a \|x(z') - x(z)\|^2 \\ &\quad \leq \sum_{i=1}^N (x^i(z') - x^i(z))^T \left[\nabla_{x^i} f_i(x^i(z), z^{-i}) - \nabla_{x^i} f_i(x^i(z'), z'^{,-i}) \right] \\ &\quad + \sum_{i=1}^N (x^i(z') - x^i(z))^T \left[\nabla_{x^i} f_i(x^i(z), z^{-i}) - \nabla_{x^i} f_i(x^i(z'), z'^{,-i}) \right], \end{aligned}$$

from which we deduce, recalling Assumption 2 and the definition of Γ,

$$\gamma \, \left\| x(z') - x(z) \right\| \, \le \, \| \Gamma \| \, \left\| z' - z \right\| .$$

Thus, if $\|\Gamma\| < \gamma$, T_s is a contraction with constant given by $\beta \triangleq \|\Gamma\|/\gamma$. This establishes our claim. Since

$$\| T(x) - T(y) \| \, = \, \lim_{s \downarrow 0} \| T_s(x) - T_s(y) \| \, \le \, \beta \, \| x - y \|,$$

it follows that T is also a contraction, and therefore has a unique fixed point $\bar{x}$. Let $\bar{x}_s$ be the fixed point of T_s (therefore $\bar{x}_s$ is the equilibrium of the game (17)). We have

$$\begin{aligned} \| \bar{x}_s - \bar{x} \| \, &= \, \| T_s(\bar{x}_s) - T(\bar{x}) \| \le \| T_s(\bar{x}_s) - T_s(\bar{x}) \| + \| T_s(\bar{x}) - T(\bar{x}) \| \\ &\le \beta \, \| \bar{x}_s - \bar{x} \| + \| T_s(\bar{x}) - T(\bar{x}) \|, \end{aligned}$$

which yields

$$\| \bar{x}_s - \bar{x} \| \, \le \, \frac{1}{1 - \beta} \, \| T_s(\bar{x}) - T(\bar{x}) \|,$$

showing that

$$\lim_{s \downarrow 0} \| \bar{x}_s - \bar{x} \| \, = \, 0.$$

(See also [1, Theorem 7.9] whose proof requires the uniform convergence of T_s to T; such uniform convergence is not applicable here because we only have the pointwise convergence (28) of T_s to T proved under the uniform approximation Assumption 3(a).) Let $\{x^\nu\}$ be the sequence generated by Algorithm 1. Writing $\bar{x}^\nu \triangleq \bar{x}_{s_\nu}$, we have $\lim\limits_{\nu \to \infty} \bar{x}^\nu = \bar{x}$; moreover,

$$\| x^{\nu+1} - \bar{x}^\nu \| \, \le \, \beta \, \| x^\nu - \bar{x}^\nu \| \, \le \, \beta \, \| x^\nu - \bar{x}^{\nu-1} \| + \beta \, \| \bar{x}^\nu - \bar{x}^{\nu-1} \|.$$

Since $\{\bar{x}^\nu\}$ is a convergent sequence, it follows that $\|\bar{x}^\nu - \bar{x}^{\nu-1}\|$ converges to 0. By [33, unnumbered Corollary to Lemma 3 on page 45], this implies that $\|x^{\nu+1} - \bar{x}^\nu\|$ goes to 0, thus $\{x^\nu\}$ converges to $\bar{x}$ as desired. □

The key assumption in Theorem 3 is the condition $\|\Gamma\| < \gamma$. This condition relates the strong convexity of each player's objective function to the Lipschtiz constants of the gradients of these objective functions. Essentially it prescribes that the strong convexity of the objective functions is "large" if compared to the Lipschitz constants that, we recall, represent the magnitude of the interactions between players. Note that if $\phi \equiv 0$, then the condition $\|\Gamma\| < \gamma$ is equivalent to an analogous condition in [13]. It is also worth pointing out that there are cases where this condition can, to some extent, be enforced; an example of this situation is when we consider the problem of solving an optimization problem (or more generally a hemivariational inequality) over the solution set of a Nash equilibrium problem, see [13, Corollary 1 and ensuing discussion].

Remark 3 Algorithm 1 requires all the players to use the same sequence of smoothing parameters $\{s_\nu\}$. While this poses no problems in the "designing games" scenario, it might be more difficult to rationalize this assumption in a purely competitive scenario, unless a minimum degree of coordination is admitted. It would seem plausible that Algorithm 1 still works when different players use different sequence of smoothing parameters, but we have no definitive answers to this issue at this time and leave this to further research.

Remark 4 Algorithm 1, as described, is a centralized algorithm. However its distributed implementation is easily accomplished. In fact the assumptions used to show convergence, in particular (22), obviously imply that the objective function of the minimization subproblem (20) is strongly convex. Therefore the well-studied synchronous or asynchronous Jacobi or Gauss-Seidel distributed algorithms are readily applicable [2,37]. When applied to (20), such an algorithm would cyclically solve subproblems of the type:

$$\underset{x^i \in X_i}{\text{minimize}} \quad \left[\phi(x^i, z^{-i}, s_\nu) + f_i(x^i, x^{\nu,-i}) \right], \quad i = 1, \ldots, N,$$

where z^{-i} represents the information about the other players' decisions variables available to player i at a particular inner iteration within the Jacobi or Gauss-Seidel loop. □

It is now clear that the key element in solving problem $\mathcal{G}$ by smoothing is the ability to define smoothing approximations satisfying Assumption 3. In the remaining part of this section we therefore turn our attention to this issue. We first consider approximation schemes that are specifically tailored to ϕ with a minmax structure; then we also hint at more general (albeit somewhat abstract) schemes which could be used to define suitable approximation schemes for an arbitrary nonsmooth function ϕ.

4.1 The minmax case

We assume that ϕ has the minmax structure given in (2) which we rewrite here for convenience:

$$\phi(x) = \max_{y \in Y} g(x, y),$$

which is assumed to satisfy Assumption 1 holds. Danskin's theorem (see for example [11]) tells us that if the maximum in the definition of ϕ is always achieved for a single y, then ϕ is continuously differentiable. Since we are supposing that ϕ is not continuously differentiable, it is clear that $g(x, \cdot)$ is not strictly concave. An obvious and well established strategy to achieve continuous differentiability for ϕ is to strongly convexify g by defining a (smoothed) function as

$$\phi(x;s) \triangleq \max_{y \in Y} \left[g(x,y) - \frac{s}{2}\|y\|^2 \right], \tag{29}$$

where s is a positive parameter. As noted above, it is clear that $\phi(\bullet; s)$ is continuously differentiable. This is confirmed by the next proposition where, more interestingly, it is also shown that $\phi(x;s)$ provides a uniform approximation of the original $\phi(x)$ and preserves σ-convexity.

Proposition 4 *Suppose that in addition to Assumption 1 a scalar ϕ^g exists such that $g(\bullet, y)$ is ϕ^g-convex on X for every $y \in Y$. The following statements hold for the function $\phi(x;s)$ defined by (29):*

(a) $\phi(\bullet; s)$ is C^1;

(b) there exist a scalar $\rho > 0$ such that $0 \leq \phi(x;s) - \phi(x) \leq \frac{s}{2}\rho^2$ for all $s > 0$;

(c) $\phi(\bullet; s)$ is σ^ϕ-convex on X for any positive s.

Proof It suffices to prove (b) as the proof of (a) and (c) is easy. To show (b), let ρ be the radius of a sphere centered at the origin, that contains Y, and let $y(x)$ be a maximizer of the problem $\max_{y \in Y} g(x,y)$. We have $\phi(x) = g(x, y(x))$. By definition, $\phi(x;s) \leq \phi(x)$. Moreover,

$$\begin{aligned}\phi(x) - \phi(x;s) &= g(x, y(x)) - \phi(x;s) \\ &\leq g(x, y(x)) - \left[g(x, y(x)) - \frac{s}{2}\|y(x)\|^2 \right] \leq \frac{s}{2}\rho^2. \quad \square\end{aligned}$$

We next consider the case of a finite minmax function defined by (3). As pointed out in Sect. 2, a finite minmax function can be seen as a particular infinite minmax function. The application of the approximation scheme just proposed above would then lead to the following approximation of ϕ:

$$\phi(x;s) = \max_{y \in \Delta} \left[\sum_{j \in J} y_j g_j(x) - \frac{s}{2}\|y\|^2 \right]$$

This would require the solution of a strongly concave, singly constrained quadratic optimization problem for each evaluation of $\phi(x;s)$. Although this task can be performed efficiently, below we give a still simpler alternative that takes further advantage of the structure of a finite minmax. More precisely, we consider the following, rather classical, approximation

$$\phi(x;s) \triangleq s \log \left[\sum_{j \in J} e^{g_j(x)/s} \right], \tag{30}$$

where s is a positive parameter. The following proposition shows that this approximation satisfies Assumption 3.

Proposition 5 *Suppose that every g_j is twice continuously differentiable and σ-convex. For any positive s the following assertions hold for the function ϕ defined by (30):*

(a) $\phi(x; s) \geq \phi(x)$ *and* $\phi(x; s) - \phi(x) \leq s \log |J|$*;*

(b) $\nabla_x \phi(x; s) = JG(x)^T \lambda(x)$*, where* $G(x) \triangleq (g_j(x))_{j\in J}$ *and* $\lambda(x) \in \mathbb{R}^{|J|}$ *with*

$$\lambda_i(x) \triangleq \frac{e^{g_i(x)/s}}{\sum_{j\in J} e^{g_j(x)/s}}$$

(c) $\nabla^2_{xx}\phi(x; s) = \sum_{j\in J} \lambda_j(x)\nabla^2 g_j(x) + \frac{1}{s} JG(x)^T(\Lambda(x) - \lambda(x)\lambda(x)^T)JG(x)$*, where* $\Lambda(x) = \mathit{Diag}(\lambda(x))$*;*

(d) $\phi(\bullet; s)$ *is* σ*-convex.*

Proof Assertion (a) follows from the fact that $\phi(x, s)$ can be rewritten as

$$\phi(x) + s \log \left[\sum_{j\in J} e^{(g_j(x)-\phi(x))/s} \right].$$

The proof of statements (b) and (c) involves straightforward, albeit lengthy calculations. Statement (d) follows from the expression of the Hessian of $\phi(\bullet; s)$ in (c), using (5) and observing that: (i) the $\lambda_j(x)$ are positive numbers that sum up to 1 and (ii) the matrix $\Lambda(x) - \lambda(x)\lambda(x)^T$ is positive semidefinite, since it is symmetric and easily seen to be diagonally dominant by (i). □

4.2 Smoothing by convolution

Although our primary interest is in the case in which the nonsmooth function ϕ is a (finite or infinite) minmax function, it is worth pointing out that convolution gives a general method to suitably smooth a function ϕ. Introduced by [38], convolution has been applied to the minimization of semicontinuous functions [8] and for solving convex nondifferentiable optimization problems [29] and their stochastic variants [41]. In what follows, we apply this method to the nonsmooth game $\mathcal{G}$. First, we briefly recall some basic facts about convolution; see for example [19].

A map $\theta_s : \mathbb{R}^n \to \mathbb{R}_+$ is a convolution kernel with support radius $s > 0$ if s is the radius of the smallest ball with center at the origin that contains the support of θ_s and $\int_{\mathbb{R}^n} \theta_s(x)\, dx = 1$. It is well-known fact that there exist C^∞ convolution kernels of any given support radius. Given a continuous, not necessarily smooth function

$\phi : \mathbb{R}^n \to \mathbb{R}$ we can define the convolution of θ_s and ϕ by

$$\phi(x; s) \triangleq \theta_s * \phi(x) \triangleq \int_{\mathbb{R}^n} \theta_s(y)\phi(x-y)\,dy = \int_{\mathbb{R}^n} \theta_s(x-z)\phi(z)\,dz. \quad (31)$$

It is an established fact that convolution provides a uniform smooth approximation of ϕ on compact sets. More precisely, if $K \subset \mathbb{R}^n$ is compact, $\varepsilon > 0$ is given, and θ_s is C^r, then $\theta_s * \phi(x)$ is C^r and $|\theta_s * \phi(x) - \phi(x)| < \varepsilon$ for any s sufficiently small. What is more interesting is that convolution also preserves σ-convexity and thus defines an approximation satisfying Assumption 3.

Proposition 6 *Let $\phi : \mathbb{R}^n \to \mathbb{R}$ be a σ^ϕ-convex function and let $\theta_s : \mathbb{R}^n \to \mathbb{R}_+$ be a continuously differentiable convolution kernel with support radius s. Suppose that ϕ is globally Lipschitz continuous with constant L on a closed set $K \subset \Omega$. Then $\phi(\bullet; s)$ is continuously differentiable and*

(a) $|\phi(x) - \phi(x; s)| \le L\|s\|$ for all $x \in K$;

(b) $\phi(\bullet; s)$ is σ^ϕ-convex.

Proof Integrating and taking into account that $\int_{\mathbb{R}^n} \theta(y)dy = 1$, we get

$$\begin{aligned} |\phi(x; s) - \phi(x)| &= \left| \int_{\mathbb{R}^n} \theta_s(y)(\phi(x-y) - \phi(x))dy \right| \\ &\le \int_{\mathbb{R}^n} \theta_s(y)|(\phi(x-y) - \phi(x))|dy \\ &\le L\|s\| \int_{\mathbb{R}^n} \theta_s(y)dy = L\|s\|. \end{aligned}$$

In a similar way we can prove the second assertion. We can write

$$\begin{aligned} &\phi(\beta x + (1-\beta)z; s) = \int_{\mathbb{R}^n} \theta_s(y)\phi(\beta x + (1-\beta)z - y)dy \\ &= \int_{\mathbb{R}^n} \theta_s(y)\phi(\beta(x-y) + (1-\beta)(z-y))dy \\ &\le \int_{\mathbb{R}^n} \theta_s(y)\left[\beta\phi(x-y) + (1-\beta)\phi(z-y) - \frac{\sigma^\phi}{2}\beta(1-\beta)\|x-z\|^2\right]dy \\ &= \beta\phi(x; s) + (1-\beta)\phi(z; s) - \frac{\sigma^\phi}{2}\beta(1-\beta)\|x-z\|^2, \end{aligned}$$

showing the σ^ϕ-convexity of the smoothed function. $\square$

Remark 5 As we already mentioned, the σ^ϕ-convexity of ϕ implies its locally Lipschitz property. Therefore, the globally Lipschtiz assumption in the previous proposition is automatically satisfied if K is bounded. However, this assumption can certainly hold also for an unbounded K. As an example, it suffices to consider the case in which ϕ is piecewise linear. □

With the above established results, it is clear that in principle we can use Algorithm 1 in conjunction with convolution to compute an equilibrium of game $\mathcal{G}$. The problem with this approach is obviously that the actual computation of the convolution might be difficult in practice (both analytically and numerically). It might be interesting to observe that the definition of $\phi(x; s)$ given in (31) can be viewed as the expectation of the function $\phi(x - \bullet)$ under the probability density function θ_s; that is,

$$\phi(x; s) = \mathbb{E}_{\theta_s}\left[\phi(x - \bullet)\right].$$

Thus one way to evaluate the function $\phi(x; s)$ is via sampling on the distribution given by θ_s. Furthermore it is known that in our setting the gradient of $\phi(\nabla; s)$ is given, component-wise, by $\dfrac{\partial \phi(x; s)}{\partial x_i} = \dfrac{\partial \theta_s(x)}{\partial x_i} * \phi(x)$, so that the same sampling procedure can be used to compute the gradient of $\phi(\bullet; s)$. The details of such a sampling approach and the analysis of its statistical properties are beyond the scope of this paper and are best left to a separate study. Nevertheless the results in this section show that in principle the scope of the smoothing approach is promising. Nevertheless, one should be mindful of the need to evaluate integrals in a convolution-based smoothing method. In most settings, this is challenging; a possible approach is to consider a stochastic approximation scheme as discussed in [41]. The latter reference addresses the solution of Cartesian stochastic VI problems arising from multi-user stochastic optimization problems or stochastic Nash games. Of particular note is the agent-specific choices of smoothing parameters which allow for the development of limited coordination schemes. See the cited reference for more details.

5 Conclusions

We have presented two approaches to the solution of games with nonsmooth objective functions, focussing on nonsmoothness arising form max-type functions and assuming that the nonsmooth part of the objective function of each player is the same for all players. For this class of problems we showed how it is possible, through suitable transformations of the original problem in either a single augmented problem or a sequence of smoothed problems, to recover a solution of the original problem. The advantage of the transformations is that they permit the application of well-known, efficient, centralized solution methods and, more importantly, of distributed solution methods. Our results can be viewed as a first step towards a much more difficult task, the solution, by distributed algorithms, of games where the nonsmooth part of each player is not necessarily the same for all players or where convolution is a natural step to take to smooth out the nonsmoothness. The solution of such games by distributed algorithms is very challenging and is the topic of our current research.

Acknowledgments The authors thank two referees who have offered constructive comments that have improved the presentation of the paper. In particular, one referee has provided additional references, including [38], that are related to our work. The work of the second author was based on research supported by the U.S.A. National Science Foundation Grant CMMI–0969600 while the work of the third author was based on research supported by the U.S.A. National Science Foundation CNS–1218717 and CAREER Grant #1254739

References

1. Berinde, V.: Iterative Approximation of Fixed Points. Lectures Notes in Mathematics, vol. 1912. Springer, Berlin (2007)
2. Bertsekas, D.P.,Tsitsiklis, J.N.: Parallel and Distributed Computation: Numerical Methods. Athena Scientific (1997) (Originally published by Prentice-Hall Inc, in 1989)
3. Burke, J.V., Hoheisel, T., Kanzow, C.: Gradient consistency for integral-convolution smoothing functions. Preprint 309, University of Würzburg, Insitute of Mathematics, Würzburg, Germany (2012)
4. Cannarsa, P., Sinestrari, C.: Semiconcave Functions, Hamilton–Jacobi Equations, and Optimal Control. Birkhäuser, Boston (2004)
5. Chen, X.: Smoothing methods for nonsmooth, nonconvex minimization. Math. Progr. **134**, 71–99 (2012)
6. Dreves, A., Facchinei, F., Kanzow, C., Sagratella, S.: On the solution of the KKT conditions of generalized Nash equilibrium problems. SIAM J. Optim. **21**, 1082–1108 (2011)
7. Eaves, B.C.: Polymatrix games with joint constraints. SlAM J. Appl. Math. **24**, 418–423 (1973)
8. Ermoliev, Y.M., Norkin, V.I., Wets, R.J.B.: The minimization of semicontinuous functions: molifier subgradients. SIAM J. Control Optim. **33**, 149–167 (1995)
9. Facchinei, F., Kanzow, C.: Generalized Nash equilibrium problems. 4OR **5**, 173–210 (2007)
10. Facchinei, F., Kanzow, C.: Penalty methods for the solution of generalized Nash equilibrium problems. SIAM J. Optim. **20**, 2228–2253 (2010)
11. Facchinei, F., Pang, J.S.: Finite-Dimensional Variational Inequalities and Complementarity Problems, vol. I and II. Springer, New York (2003)
12. Facchinei, F., Pang, J.S.: Nash equilibria: the variational approach. In: Eldar, Y., Palomar, D. (eds.) Convex Optimization in Signal Processing and Communications, pp. 443–493. Cambridge University Press, Cambridge (2009)
13. Facchinei, F., Pang, J.-S., Scutari, G., Lampariello, L.: VI-constrained hemivariational inequalities: distributed algorithms and power control in ad-hoc networks. Math. Progr. (2013). doi:10.1007/s10107-013-0640-5. In print. Published electronically
14. Facchinei, F., Piccialli, V., Sciandrone, M.: Decomposition algorithms for generalized potential games. Comput. Optim. Appl. **50**, 237–262 (2011)
15. Fukushima, M.: Restricted generalized Nash equilibria and controlled penalty algorithm. Comput. Manag. Sci. **8**, 201–218 (2011)
16. Gürkan, G., Pang, J.-S.: Approximations of Nash equilibria. Math. Progr. **117**, 223–253 (2009)
17. Haurie, A., Krawczyk, J.B.: Optimal charges on river effluent from lumped and distributed sources. Environ. Model. Assess. **2**, 93–106 (1997)
18. von Heusinger, A., Kanzow, C., Fukushima, M.: Newton's Method for computing a normalized equilibrium in the generalized Nash game through fixed point formulation. Math. Progr. **132**, 99–123 (2012)
19. Hirsch, F.: Differential Topology. Springer, New York (1976)
20. Jofré, A., Wets, R.J.-B.: Variational convergence of bivariate functions: iopsided convergence. Math. Progr. **115**, 275–295 (2009)
21. Kannan, A., Shanbhag, U.V., Kim, H.M.: Addressing supply-side risk in uncertain power markets: stochastic generalized Nash models and scalable algorithms. Optim. Methods Softw. **28**, 1095–1138 (2013)
22. Kannan, A., Shanbhag, U.V., Kim, H.M.: Strategic behavior in power markets under uncertainty. Energy Syst. **2**, 115–141 (2011)
23. Krawczyk, J.B.: Coupled constraint Nash equilibria in environmental games. Resour. Energy Econ. **27**, 157–181 (2005)

24. Krawcyzk, J.B.: Numerical solutions to couple-constraint (or generalized) Nash equilibrium problems. Comput. Manag. Sci. **4**, 183–204 (2007)
25. Krawczyk, J.B., Uryasev, S.: Relaxation algorithms to find Nash equilibria with economic applications. Environ. Model. Assess. **5**, 63–73 (2000)
26. Kubota, K., Fukushima, M.: Gap function approach to the generalized Nash equilibrium problem. J. Optim. Theory Appl. **144**, 511–531 (2010)
27. Kulkarni, A.A., Shanbhag, U.V.: On the variational equilibrium as a refinement of the generalized Nash equilibrium. Automatica **48**, 45–55 (2012)
28. Kulkarni, A.A., Shanbhag, U.V.: Revisiting generalized Nash games and variational inequalities. J. Optim. Theory Appl. **154**, 175–186 (2012)
29. Lakshmanan, H., Farias, D.: Decentralized resource allocation in dynamic networks of agents. SIAM J. Optim. **19**, 911–940 (2008)
30. Nabetani, K., Tseng, P., Fukushima, M.: Parametrized variational inequality approaches to generalized Nash equilibrium problems with shared constraints. Comput. Optim. Appl. **48**, 423–452 (2011)
31. Pang, J.S., Fukushima, M.: Quasi-variational inequalities, generalized Nash equilibria, and multi-leader-follower games. Comput. Manag. Sci. **2**, 21–56 (2005). Erratum, ibid. 6 (2009) 373–375
32. Pang, J.S., Scutari, G.: Nonconvex games with side constraints. SIAM J. Optim. **21**, 1491–1522 (2011)
33. Polyak, B.T.: Introduction to Optimization. Optimization Software, New York (1987)
34. Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis. Springer, Berlin (1998)
35. Rosen, J.: Existence and uniqueness of equilibrium points for concave n-person games. Econometrica **33**, 520–534 (1965)
36. Scutari, G., Facchinei, F., Pang, J.-S., Palomar, D.: Real and complex monotone communication games. IEEE Trans Inform Theory (submitted). Technical report (2013). Available at http://arxiv.org/abs/1212.6235
37. Scutari, G., Facchinei, F., Song, P., Palomar, D., Pang, J.-S.: Decomposition by partial linearization: parallel optimization of multi-agent systems. IEEE Trans. Signal Process. **62**, 641–656 (2014)
38. Steklov, V.A.: Sur les expressions asymptotiques de certaines fonctions definies par les équations differentielles du second ordre et leurs applications au problème du dévelopement d'une fonction arbitraire en séries procédant suivant les diverses fonctions. Commun. Kharkov Math. Soc. **2**, 97–199 (1907)
39. Uryasev, S., Rubinstein, R.Y.: On relaxation algorithms in computation of noncooperative equilibria. IEEE Trans. Autom. Control **39**, 1263–1267 (1994)
40. Vial, J.-P.: Strong and weak convexity of sets and functions. Math. Oper. Res. **8**, 231–259 (1983)
41. Yousefian, F., Nedich, A., Shanbhag, U.V.: On stochastic gradient and subgradient methods with adaptive steplength sequences. Automatica **48**, 56–67 (2012)