Proceedings of Machine Learning Research vol 283:1–13, 2025 7th Annual Conference on Learning for Dynamics and Control

Computing Quasi-Nash Equilibria via Gradient-Response Schemes

Zhuoyu Xiao ZYXIAO@UMICH.EDU
Department of Industrial and Operations Engineering, University of Michigan

Uday V. Shanbhag UDAYBAG@UMICH.EDU
Department of Industrial and Operations Engineering, University of Michigan

Editors: N. Ozay, L. Balzano, D. Panagou, A. Abate

Abstract

We consider a class of smooth static N-player noncooperative games, where player objectives are expectation-valued and potentially nonconvex. In such a setting, we consider the largely open question of efficiently computing a suitably defined *quasi*-Nash equilibrium (QNE) via a stochastic gradient-response framework. First, under a suitably defined quadratic growth property, we prove that both the stochastic synchronous gradient-response (**SSGR**) scheme and its asynchronous counterpart (**SAGR**) are characterized by almost sure convergence to a QNE and a sublinear rate guarantee. Notably, when a potentiality requirement is overlaid under a somewhat stronger pseudomonotonicity condition, this claim can be made for a Nash equilibrium (NE), rather than a QNE. Second, under the weak sharpness property, we show that the deterministic synchronous variant (**SGR**) displays a *linear* rate of convergence sufficiently close to a QNE by leveraging a geometric decay in steplengths. This suggests the development of a two-stage scheme with global non-asymptotic sublinear rates and a local linear rate. We also present applications satisfying the prescribed requirements where preliminary numerics appear promising.

Keywords: Nonconvex games, quasi-Nash equilibrium, stochastic approximation, weak sharpness.

1. Introduction

In the last several decades, the Nash equilibrium (NE) introduced in Nash (1951) has assumed growing relevance in engineered and economic systems, complicated by the presence of competition among a collection of self interested entities (Facchinei and Pang (2009); Lei and Shanbhag (2022b)). Managing such systems has necessitated the need to understand the properties of the associated Nash equilibria, prompting the long-standing interest in studying algorithms for computing a Nash equilibrium of an N-player game (Fudenberg and Levine (1998); Facchinei and Pang (2009)). Specifically, we consider the N-player noncooperative game $\mathcal{G}(\mathbf{f}, X, \boldsymbol{\xi})$, where for any $i \in [N] := \{1, 2, \cdots, N\}$, the ith player solves the following parametrized optimization problem, where $x_{-i} \in X_{-i}$, $X_i \subseteq \mathbb{R}^{n_i}$, $X_{-i} := \prod_{j \neq i} X_j$ and $X := \prod_{i=1}^N X_i$:

$$\min_{x_i \in X_i} f_i(x_i, x_{-i}) := \mathbb{E}[\tilde{f}_i(x_i, x_{-i}, \boldsymbol{\xi})], \tag{1}$$

In addition, the random variable $\boldsymbol{\xi} : \Omega \to \mathbb{R}^m$ is defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, Ξ is defined as $\Xi := \{\boldsymbol{\xi}(\omega) \mid \omega \in \Omega\}$, and $\tilde{f}_i : X \times \Xi \to \mathbb{R}$ is a real-valued function. In the context of continuous-strategy games, much of these developments have necessitated imposing convexity requirements on player objectives and strategy sets, significantly limiting the applicability of such models. In the last decade, there has been the forwarding of a weaker solution concept for equilibrium that aligns with the notion of B-stationarity in optimization problems (cf. Cui and Pang (2021), Definition 6.1.1). Referred to as the quasi-Nash equilibrium (QNE), this solution concept was first suggested by Pang and Scutari (2011). We now proceed to briefly describe some related work.

Related work. Computation of NE in smooth convex continuous-strategy constrained games is tied to resolving variational inequality problems (Facchinei and Pang (2003)). Stochastic generalizations have prompted a study of stochastic gradient-response (Koshal et al. (2013); Yousefian et al. (2016); Lei and Shanbhag (2022a)) as well as best-response schemes (Lei and Shanbhag (2022a)) and their delay-afflicted and asynchronous counterparts (Lei et al. (2020)). In deterministic nonconvex games, QNE computation has leveraged surrogation-based best-response schemes (Cui and Pang (2021); Pang and Razaviyayn (2016); Razaviyayn (2014)). Recall that a QNE of a smooth nonconvex game can be captured by a non-monotone variational inequality problem, a class of problems that has seen some recent study. Table 1 details extragradient-type (EG) schemes or operator extrapolation (OE) schemes for solving non-monotone VIs under either the Minty condition or pseudomonotonicity and its variant, while articulating the distinctions in the current work. Notably, the Minty condition is closely related to pseudomonotonicity (Huang and Zhang (2024)).

work	**scheme**	**stochastic**	**assumption**	**convergence**	**rate**
Iusem et al. (2017)	VR-EG	✓	PM	a.s. limit points	sublinear
Iusem et al. (2019)	DS-SA-EG	✓	PM	a.s. limit points	sublinear
Kannan and Shanbhag (2019)	SEG/MPSA	✓	PM	a.s.; in expectation	sublinear
Dang and Lan (2015)	EG	✗	GM	✓	sublinear
Kotsalis et al. (2022)	OE/VR-OE	✓	GSM	in expectation	sublinear
Vankov et al. (2023)	Popov	✓	Minty	a.s.; in expectation	sublinear
Arefizadeh and Nedić (2024)	EG	✗	Minty	limit points	sublinear
Huang and Zhang (2024)	ARE/PGR/EG	✗	Minty	✓	sublinear

Table 1: A summary of recent schemes for non-monotone VIs (PM: pseudomonotone; GM: generalized monotone; GSM: generalized strongly monotone).

Gaps & Questions. (i) Can we develop efficient algorithms with last-iterate convergence guarantees for computing a deterministic or stochastic QNE under conditions that go beyond Minty-type and variants of monotonicity? (ii) Can asynchronous variants of such algorithms be developed while still providing convergence rate guarantees? (iii) Under what conditions can (locally) linear rate be achieved without relying on strong monotonicity? (iv) Are there conditions under which convergence can be strengthened from QNE to NE despite the scourge of nonconvexity?

Main contributions. Motivated by these gaps, after providing some preliminaries in Section 2, we present a.s. convergence and sublinear rate guarantees to a QNE for **SSGR** and **SAGR** in stochastic nonconvex games under a quadratic growth (**QG**) property in Section 3. Notably, under an additional requirement of potentiality and strong pseudomonotonicity (**SP**), convergence can be guaranteed to an NE. In Section 4, under a weak-sharpness (**WS**) property, we prove that for deterministic realization **SGR** scheme, the squared error diminishes at a linear rate sufficiently close to the solution. This allows for developing an asymptotically convergent two-stage scheme that displays local linear convergence. In Section 5, we present applications satisfying the prescribed properties with preliminary numerics displaying promise and conclude in Section 6. We summarize our contributions in Table 2. The missing technical proofs as well as additional schemes can be found in Xiao and Shanbhag (2025).

property	scheme	stochastic	nonconvex	QNE	a.s. convergence	convergence (in mean)
AA	SSGR	✓	✓	✓	subsequential (Thm. 8)	✗
	SAGR	✓	✓	✓	subsequential (Thm. 9)	✗
QG	SSGR	✓	✓	✓	Thm. 10	sublinear (Thm. 10)
	SAGR	✓	✓	✓	Thm. 11	sublinear (Thm. 11)
SP + potential	identical to **QG**			NE	identical to **QG** (Thm. 14)	
WS	SGR	✗	✓	✓	✗	*locally* linear (Thm. 16)

Table 2: A summary of contributions.

Notation. We denote the inner product between vectors x and $y \in \mathbb{R}^n$ by $x^\top y$. We denote the partial derivative map of a smooth function f with respect to x_i by $\nabla_{x_i} f$. $\Pi_X[x]$ denotes the Euclidean projection of x onto set X while $\mathbb{E}[\boldsymbol{\xi}]$ denotes the expectation of a random variable $\boldsymbol{\xi}$. The interior of set X is denoted by $\mathrm{int}(X)$.

2. Preliminaries

We impose the following ground assumption throughout this paper.

Assumption 1 *For any $i \in [N]$, the following hold.* (a) $X_i \subseteq \mathbb{R}^{n_i}$ *is convex and closed.* (b) *Given $x_{-i} \in X_{-i}$, $f_i(\cdot, x_{-i}) = \mathbb{E}[\tilde{f}_i(\cdot, x_{-i}, \boldsymbol{\xi})]$ is C^1 on an open set $\mathcal{O}_i$ such that $X_i \subset \mathcal{O}_i$.*

When for any $i \in [N]$, the ith player-specific objective $f_i(\bullet, x_{-i})$ loses convexity for a given x_{-i} in $\mathcal{G}(\mathbf{f}, X, \Xi)$, both deriving existence as well as computing equilibria becomes challenging. This has led to the weaker solution concept of the *quasi-Nash equilibrium* (QNE), inspired by B-stationarity in potentially nonsmooth optimization problems. Before defining the QNE, recall the notion of B-stationarity (cf. Cui and Pang (2021), Definition 6.1.1). Given an optimization problem $\min_{x \in X} f(x)$, where f is directionally differentiable, we refer to $x^* \in X$ as a B-stationary point of f on X if $f'(x^*; v) \geqslant 0$ for all $v \in \mathcal{T}(x^*; X)$, where $f'(x^*; v)$ represents the directional derivative at x^* along a direction v and $\mathcal{T}(x^*; X)$ denotes the tangent cone to set X at x^* (cf. Cui and Pang (2021), page 199). If f is differentiable and X is convex, B-stationarity of x^* reduces to $\nabla f(x^*)^\top (x - x^*) \geqslant 0$ for all $x \in X$. Inspired by this setup, Pang and Scutari (2011) introduced the QNE, which has been defined next.

Definition 1 (Quasi-Nash equilibrium (cf. Pang and Scutari (2011), Definition 2)) *Consider the N-player game $\mathcal{G}(\mathbf{f}, X, \boldsymbol{\xi})$. For any $i \in [N]$, suppose $f_i(\bullet, x_{-i})$ is C^1 for any $x_{-i} \in X_{-i}$. Then $x^* = (x_i^*)_{i=1}^N$ is a quasi-Nash equilibrium (QNE) if for any $i \in [N]$, we have*

$$\nabla_{x_i} f_i(x_i^*, x_{-i}^*)^\top (x_i - x_i^*) \geqslant 0, \ \forall\, x_i \in X_i.$$

We observe that x^* is a QNE if and only if x^* solves VI (X, F), i.e., x^* satisfies $F(x^*)^\top(x - x^*) \geqslant 0$ for all $x \in X$, where F is expectation-valued, defined as $F(x) := (\nabla_{x_i} f_i(x))_{i=1}^N$. This facilitates the utilization of VI literature (Facchinei and Pang (2003)). We first show the existence guarantee for QNE based on Facchinei and Pang (2003) and Ravat and Shanbhag (2011), extending the classical existence result of an NE (Nash (1951)).

Theorem 2 (Existence of a QNE) *Consider the N-player game $\mathcal{G}(\mathbf{f}, X, \boldsymbol{\xi})$. Suppose Assumption 1 holds and for any x, ξ, $\tilde{F}(x, \xi) = (\nabla_{x_i} \tilde{f}_i(x, \xi))_{i=1}^N$. Then a QNE exists if* (i) *or* (ii) *hold:* (i) *If X is bounded;* (ii) *If there exists $x^{\text{ref}} \in X$ such that $\liminf\limits_{\|x\| \to \infty, x \in X} \tilde{F}(x, \xi)^\top (x - x^{\text{ref}}) \geqslant 0$ a.s.*

We now turn to the computation of QNE where strategy updates are simultaneous or asynchronous; the former results in a stochastic synchronous gradient-response (**SSGR**) scheme:

$$x_i^{k+1} = \Pi_{X_i}[x_i^k - \gamma^k \nabla_{x_i} \tilde{f}_i(x_i^k, x_{-i}^k, \xi_i^k)], \ \forall i \in [N], \tag{\textbf{SSGR}}$$

where the history of **SSGR** is defined as $\mathcal{F}_k = \sigma\{x^0, \cup_{t=0}^{k-1} \{\nabla_{x_i} \tilde{f}_i(x^t, \xi_i^t)\}_{i=1}^N\}$. Asynchronous strategy updates results in a stochastic synchronous gradient-response (**SAGR**) scheme:

$$x_{i(k)}^{k+1} = \Pi_{X_{i(k)}}[x_{i(k)}^k - \gamma_{i(k)}^k \nabla_{x_{i(k)}} \tilde{f}_{i(k)}(x_{i(k)}^k, x_{-i(k)}^k, \xi_{i(k)}^k)], \tag{\textbf{SAGR}}$$

where $i(k)$ is the selected player at iteration k with probability $p_{i(k)}$ such that $\sum_{i=1}^N p_i = 1$, and the histories $\mathcal{F}_k$ and $\mathcal{F}_{k+1/2}$ of **SAGR** are defined as $\mathcal{F}_k = \sigma\{x^0, \cup_{t=0}^{k-1} \{i(t), \nabla_{x_{i(t)}} \tilde{f}_{i(t)}(x^t, \xi_{i(t)}^t)\}\}$ and $\mathcal{F}_{k+1/2} = \mathcal{F}_k \cup \{i(k)\}$, respectively. In Sections 3-4, we derive convergence and rate guarantees for **SSGR** and **SAGR** schemes under four properties: (i) acute angle (**AA**); (ii) quadratic growth (**QG**); (iii) weak sharpness (**WS**); and (iv) strong pseudomonotonicity (**SP**). We present their definitions below where X^* denotes the solution set of VI (X, F).

Definition 3 (Four properties) *We say that VI (X, F) satisfies*

(i) **Acute Angle (AA)** *if we have*

$$(x - x^*)^\top F(x) > 0, \ \forall x \in X \backslash X^* \text{ and } x^* \in X^*. \tag{\textbf{AA}}$$

(ii) **Quadratic Growth (QG)** *if there exists $\alpha > 0$ such that*

$$(x - x^*)^\top F(x) \geqslant \alpha \|x - x^*\|^2, \ \forall x \in X \backslash X^* \text{ and } x^* \in X^*. \tag{\textbf{QG}}$$

(iii) **Strong Pseudomonotonicity (SP)** *if there exists $\eta > 0$ such that*

$$(x - y)^\top F(y) \geqslant 0 \implies (x - y)^\top F(x) \geqslant \eta \|x - y\|^2, \ \forall x, y \in X. \tag{\textbf{SP}}$$

(iv) **Weak Sharpness (WS)** *if there exists* $\beta > 0$ *such that*

$$(x - x^*)^\top F(x^*) \geqslant \beta \|x - x^*\|, \forall x \in X \text{ and } x^* \in X^*. \tag{WS}$$

Remark 4 *The* (**QG**) *property may be more suitable than a similar property introduced by Kotsalis et al. (2022) when VI* (X, F) *admits multiple solutions;* F *satisfies* μ*-generalized strong monotonicity on* X *if* $(x - x^*)^\top F(x) \geqslant \mu\|x - x^*\|^2$ *holds for any* $x \in X$*. However, such a definition does not exclude* x *from the solution set* X^* *as in* (**QG**)*. Suppose, we have two distinct solutions* $x^* \neq \hat{x}$*. Then the left-hand side* $(\hat{x} - x^*)^\top F(\hat{x})$ *is nonpositive (from* $\hat{x}$ *being a solution) while the right-hand side* $\mu\|\hat{x} - x^*\|^2$ *is strictly positive.*

Several implications follow from Defition 3 and we formalize them in the following proposition.

Proposition 5 *Consider VI* (X, F) *with a continuous mapping* $F : X \subseteq \mathbb{R}^n \to \mathbb{R}^n$*. Suppose* F *satisfies* (**SP**) *with paramter* $\eta > 0$*. Then the follow hold: (i) The solution set* X^* *is a singleton; (ii)* F *satisfies* (**QG**) *with parameter* $\eta > 0$ *hence* F *also satisfies* (**AA**)*.*

We now turn to the weak sharpness property (**WS**). Weak sharp minima were first defined by Burke and Ferris (1993), where the minimization of a function f over a set X has a weak sharp minimum on solution set X^*, if there exists $\beta > 0$ such that $f(x) - f^* \geqslant \beta \operatorname{dist}(x, X^*)$ for any $x \in X$. In fact, Burke and Ferris (1993) showed that the above primal requirement is equivalent to the geometric inclusion requirement when f is closed, proper, and convex:

$$-\nabla f(x^*) \in \operatorname{int}\left(\bigcap_{y \in X^*} (\mathcal{T}_X(y) \cap \mathcal{N}_{X^*}(y))^\circ\right), \ \forall x^* \in X^*, \tag{2}$$

where Y° denotes the polar of the set $Y \subset \mathbb{R}^n$ (cf. Cui and Pang (2021), page 69). An analogous result (cf. Marcotte and Zhu (1998), Theorem 4.1) was provided for VI (X, F) via the dual gap function, an extended real-valued function defined as $G(x) \triangleq \sup_{y \in X} F(y)^\top (x - y)$. When F is continuous and pseudomonotone plus (cf. Facchinei and Pang (2003), Definition 2.3.9) over X and X is compact, we have

$$\begin{aligned} &(\mathbf{I}) \left[\, G(x) \ \geqslant \ \beta \operatorname{dist}(x, X^*) \text{ for any } x \in X \,\right] \\ \Longleftrightarrow\ &(\mathbf{II}) \left[-G(x^*) \in \operatorname{int}\left(\bigcap_{y \in X^*} (\mathcal{T}_X(y) \cap \mathcal{N}_{X^*}(y))^\circ\right), \ \forall x^* \in X^* \right]. \end{aligned}$$

It follows from the definition of the dual gap function that (**WS**) property implies (**I**), i.e.

$$\begin{aligned} &(\mathbf{WS}) \left[F(x^*)^\top (x - x^*) \geqslant \beta\|x - x^*\| \text{ for any } x \in X \text{ and any } x^* \in X^*\right] \\ \Longrightarrow\ &(\mathbf{WS}^2) \left[F(x^*)^\top (x - x^*) \geqslant \beta \operatorname{dist}(x, X^*) \text{ for any } x \in X\right] \implies (\mathbf{I}). \end{aligned}$$

If X^* is a singleton, then (**WS**) and (**WS**^{2}) are equivalent for any $\beta > 0$, motivating the usage of the (**WS**) requirement on F.

3. QNE computation under (AA), (QG) and (SP)

In this section, we derive asymptotics and convergence rates for **SSGR** and **SAGR** schemes under (**AA**) and (**QG**). We then show that the potentiality property allows for convergence to an NE under (**SP**), rather than just a QNE. We impose Assumption 2 throughout Section 3.

Assumption 2 *For any $k \geqslant 0$ and any $i, i(k) \in [N]$, the following hold.*
(a) **Unbiasedness.** *We have* $\mathbb{E}[w^k \mid \mathcal{F}_k] = 0$ (SSGR) *and* $\mathbb{E}[w^k_{i(k)} \mid \mathcal{F}_{k+1/2}] = 0$ (SAGR) *, where errors w^k and $w^k_{i(k)}$ are defined as $w^k = (w^k_i)_{i=1}^N$ with $w^k_i = \nabla_{x_i} \tilde{f}_i(x^k, \xi^k_i) - \nabla_{x_i}\mathbb{E}[\tilde{f}_i(x^k, \boldsymbol{\xi})]$ and $w^k_{i(k)} = \nabla_{x_{i(k)}} \tilde{f}_{i(k)}(x^k, \xi^k_{i(k)}) - \nabla_{x_{i(k)}}\mathbb{E}[\tilde{f}_{i(k)}(x^k, \boldsymbol{\xi}) \mid \mathcal{F}_{k+1/2}]$.*
(b) **Moment bounds.** (SSGR) *There exist $M_1 > 0$ such that $\mathbb{E}[\|w^k\|^2 \mid \mathcal{F}_k] \leqslant M_1$;* (SAGR) *There exists $M_{1,i(k)} > 0$ such that $\mathbb{E}[\|w^k_{i(k)}\|^2 \mid \mathcal{F}_{k+1/2}] \leqslant M_{1,i(k)}$.*
(c) **Boundedness.** *There exists $M_{2,i} > 0$ and $M_2 > 0$ such that $\|\nabla_{x_i} f_i(x_i, x_{-i})\|^2 \leqslant M_{2,i}$ and $\sum_{i=1}^N \|\nabla_{x_i} f_i(x_i, x_{-i})\|^2 \leqslant M_2 := \sum_{i=1}^N M_{2,i}$.*

3.1. Two key recursions

We first derive two key recursions *without* imposing any properties from Definition 3. The **SAGR** recursion is somewhat more complicated than **SSGR** recursion, given randomness in the stepsize.

Lemma 6 (SSGR recursion) *Consider the N-player game $\mathcal{G}(\mathbf{f}, X, \boldsymbol{\xi})$. Suppose that Assumptions 1 and 2 hold. Let x^* be any QNE. Consider the sequence of iterates $\{x^k\}_{k=0}^\infty$ generated by the **SSGR** scheme and suppose the stepsize $\{\gamma^k\}_{k=0}^\infty$ satisfies $\sum_{k=0}^\infty \gamma^k = \infty$ and $\sum_{k=0}^\infty (\gamma^k)^2 < \infty$. If $M_1, M_2 > 0$ are defined in Assumption 2, then we have for any $k \geqslant 0$:*

$$\mathbb{E}[\|x^{k+1} - x^*\|^2 \mid \mathcal{F}_k] \leqslant \|x^k - x^*\|^2 - 2\gamma^k (x^k - x^*)^\top F(x^k) + 2(\gamma^k)^2 (M_1 + M_2).$$

Now we consider the **SAGR** scheme. Naturally, deterministic steplengths are harder to prescribe, leading to random steplengths as presented by Koshal et al. (2013) and Nedić (2011). To this end, we define the stepsize γ^k_i at any $k \geqslant 0$ as follows:

$$\gamma^k_i := \begin{cases} 1/\Gamma_k(i), & \text{if } \Gamma_k(i) \neq 0, \\ 0, & \text{if } \Gamma_k(i) = 0, \end{cases} \tag{3}$$

where $\Gamma_k(i(k))$ denotes the number of updates that player $i(k)$ (the player chosen at time k) has performed until and including the kth iteration. This leads to an additional source of uncertainty, complicating the analysis.

Lemma 7 (SAGR recursion) *Consider the N-player game $\mathcal{G}(\mathbf{f}, X, \boldsymbol{\xi})$. Suppose that Assumptions 1 and 2 hold. Let x^* be any QNE. Consider the sequence $\{x^k\}_{k=0}^\infty$ generated by the **SAGR** scheme and suppose that $\{\gamma^k_{i(k)}\}_{k=0}^\infty$ is defined as in (3) for any $k \geqslant 0$. If $M_{1,i}, M_{2,i} > 0$ are defined in Assumption 2 for $i \in [N]$, then we have for any $k \geqslant 0$:*

$$\mathbb{E}[\|x^{k+1} - x^*\|^2 \mid \mathcal{F}_k] \leqslant (1 + \max_i p_i |\gamma^k_i - \tfrac{1}{kp_i}|)\|x^k - x^*\|^2 + 2\sum_{i=1}^N p_i (\gamma^k_i)^2 M_{1,i}$$
$$+ \sum_{i=1}^N p_i (2(\gamma^k_i)^2 + |\gamma^k_i - \tfrac{1}{kp_i}|) M_{2,i} - \frac{2}{k}(x^k - x^*)^\top F(x^k),$$

where p_i is the probability that the ith player updates the strategy such that $\sum_{i=1}^N p_i = 1$.

3.2. SSGR and SAGR under (AA) and (QG)

Theorem 8 *Consider the N-player game $\mathcal{G}(\mathbf{f}, X, \boldsymbol{\xi})$. Suppose Assumptions 1 and 2 hold. Consider the sequence $\{x^k\}_{k=0}^{\infty}$ generated by the **SSGR** scheme under the (**AA**) property. Suppose the stepsize $\{\gamma^k\}_{k=0}^{\infty}$ satisfies $\sum_{k=0}^{\infty} \gamma^k = \infty$ and $\sum_{k=0}^{\infty} (\gamma^k)^2 < \infty$. If the solution set X^* is compact, then some subsequence of iterates $\{x^k\}_{k=0}^{\infty}$ converges a.s. to a QNE.*

Theorem 9 *Consider the N-player game $\mathcal{G}(\mathbf{f}, X, \boldsymbol{\xi})$. Suppose that Assumptions 1 and 2 hold. Consider the sequence $\{x^k\}_{k=0}^{\infty}$ generated by the **SAGR** scheme under the (**AA**) property where $\gamma_{i(k)}^k$ is defined as in (3) for any $k \geqslant 0$. If the solution set X^* is compact, then some subsequence of iterates $\{x^k\}_{k=0}^{\infty}$ converges a.s. to a QNE.*

Only almost sure *subsequential* convergence is available for **SSGR** and **SAGR** schemes under (**AA**). The Robbins-Siegmund lemma (cf. Robbins and Siegmund (1971), Theorem 1) plays a crucial role in the proofs of the two above results. The detailed proofs can be found in Xiao and Shanbhag (2025). Next, we impose stronger properties (**QG**) to recover a.s. convergence.

Theorem 10 *Consider the N-player game $\mathcal{G}(\mathbf{f}, X, \boldsymbol{\xi})$. Suppose Assumptions 1 and 2 hold. Let $\{x^k\}_{k=0}^{\infty}$ be generated by the **SSGR** scheme. Suppose that (**QG**) holds and X^* is a singleton. Then the following two statements hold.*

(a) (diminishing stepsize) *Suppose $\gamma^k = \gamma^0/k$, where $\gamma^0 > 1/(2\alpha)$. Let $Q(\gamma^0)$ be defied as $Q(\gamma^0) := \max\{\frac{2(\gamma^0)^2(M_1+M_2)}{2\alpha\gamma^0-1}, \mathbb{E}[\|x^1 - x^*\|^2]\}$, where M_1 and M_2 are defined in Assumption 2. Then we have* (i) $\lim_{k\to\infty} x^k = x^*$ *a.s.;* (ii) $\mathbb{E}[\|x^k - x^*\|^2] \leqslant \frac{Q(\gamma^0)}{k}$ *holds for $k \geqslant 1$.*

(b) (constant stepsize) *Suppose $\gamma^k = \delta$ such that $q := 1 - 2\alpha\delta < 1$. Then we have $\mathbb{E}[\|x^k - x^*\|^2] \leqslant \mathcal{O}(\delta)$ after $\mathcal{O}(\lceil \frac{1}{\delta} \ln(\frac{1}{\delta})\rceil)$ steps.*

Proof (a) By Lemma 6 and the (**QG**) property, i.e., $(x - x^*)^\top F(x) \geqslant \alpha\|x - x^*\|^2$, it follows that

$$\mathbb{E}[\|x^{k+1} - x^*\|^2 | \mathcal{F}^k] \leqslant (1 - 2\alpha\gamma^k)\|x^k - x^*\|^2 + 2(\gamma^k)^2 (M_1 + M_2) . \tag{4}$$

When k is sufficiently large, we have $0 \leqslant 2\alpha\gamma^k \leqslant 1$. By Lemma 2.2.10 in Polyak (1987), we may claim a.s. convergence (i). Taking unconditional expectations on both sides of (4), we obtain

$$\mathbb{E}[\|x^{k+1} - x^*\|^2] \leqslant (1 - 2\alpha\gamma^k)\mathbb{E}[\|x^k - x^*\|^2] + 2(\gamma^k)^2 (M_1 + M_2) .$$

By invoking Lemma 5 in Xiao and Shanbhag (2025), we obtain $\mathbb{E}[\|x^k - x^*\|^2] \leqslant \frac{Q(\gamma^0)}{k}$ for $k \geqslant 1$.
(b) Suppose $v_k = \mathbb{E}[\|x^k - x^*\|^2]$ is such that $v_{k+1} \leqslant (1 - 2\alpha\gamma^k)v_k + (\gamma^k)^2 C$, where $C = 2(M_1 + M_2)$. For every k, let $\gamma_k = \delta$ such that $q = 1 - 2\alpha\delta < 1$, implying

$$\begin{aligned} v_{k+1} &\leqslant q v_k + \delta^2 C \leqslant q^2 v_{k-1} + q\delta^2 C + \delta^2 C \\ &\leqslant q^{k+1} v_0 + \delta^2 C(1 + q + q^2 + \cdots + q^k) \leqslant q^{k+1} v_0 + \delta^2 C \tfrac{1}{1-q} = q^{k+1} v_0 + \tfrac{\delta C}{2\alpha}. \end{aligned}$$

Let $N = \frac{1}{\delta}$ and $k = \lceil N\tilde{K} \rceil$ for some $\tilde{K}$, where we observe that q^k can be bounded as

$$q^k = (1 - 2\alpha\delta)^k \leqslant \left((1 - 2\alpha\delta)^N\right)^{\tilde{K}} = \left((1 - 2\alpha\delta)^{\frac{1}{\delta}}\right)^{\tilde{K}} .$$

We know that for $|u| < n$, $(1 - u/n)^n \leqslant e^{-u}$ holds, implying that

$$q^k \leqslant \left((1 - 2\alpha\delta)^{\frac{1}{\delta}} \right)^{\tilde{K}} \leqslant \left(e^{-2\alpha} \right)^{\tilde{K}} \leqslant \delta, \text{ if } \tilde{K} \geqslant \tfrac{1}{2\alpha} \ln\left(\tfrac{1}{\delta}\right).$$

Therefore, after $k = N\tilde{K} = \lceil \mathcal{O}(\frac{1}{\delta} \ln(\frac{1}{\delta})) \rceil$ steps, $v_k \leqslant \mathcal{O}(\delta)$, implying sublinear convergence. ■

We conclude by deriving a non-asymptotic sublinear rate for **SAGR** under **QG**.

Theorem 11 *Consider the N-player game $\mathcal{G}(\mathbf{f}, X, \boldsymbol{\xi})$. Suppose Assumptions 1 and 2 hold. Let $\{x^k\}_{k=0}^{\infty}$ be generated by the **SAGR** scheme. Suppose that (**QG**) holds. Suppose X^* is a singleton. Then we have:* (a) $\lim_{k\to\infty} x^k = x$ *a.s.;* (b) *There exists sufficiently large K and constants $\tilde{M}_1, \tilde{M}_2 > 0$ such that for $k \geqslant K$:*

$$\mathbb{E}[\|x^k - x^*\|^2] \begin{cases} \leqslant \frac{2+8N^2(\tilde{M}_1+\tilde{M}_2)+2\tilde{M}_2}{(2\beta-1/2+q)k^{1/2-q}} + o(k^{-1/2+q}), & \textit{if } 2\alpha > 1/2 - q, \\ = \mathcal{O}(k^{-2\beta} \log k), & \textit{if } 1/2 - q = 2\alpha, \\ = \mathcal{O}(k^{-2\beta}), & \textit{if } 1/2 - q > 2\alpha. \end{cases}$$

Proof We prove part (b). By Lemma 7 and the (**QG**) property, after taking unconditional expectations, we obtain the following recursion where the randomness in $\{\gamma_i^k\}$ bears reminding:

$$\mathbb{E}[\|x^{k+1} - x^*\|^2] \leqslant \mathbb{E}[(1 - \tfrac{2\alpha}{k} + \max_i p_i|\gamma_i^k - \tfrac{1}{kp_i}|)\|x^k - x^*\|^2] + 2\sum_{i=1}^{N} p_i \mathbb{E}[(\gamma_i^k)^2] M_{1,i}$$

$$+ \sum_{i=1}^{N} p_i \mathbb{E}[(2(\gamma_i^k)^2 + |\gamma_i^k - \tfrac{1}{kp_i}|)] M_{2,i}, \text{ where } \tilde{M}_1 := \sum_{i=1}^{N} M_{1,i} \text{ and } \tilde{M}_2 := \sum_{i=1}^{N} M_{2,i}.$$

By invoking Lemmas 4 and 6 in Xiao and Shanbhag (2025), we obtain the desired result. ■

The singleton assumption on X^* plays a crucial role in the proof of both Theorems 10 and 11. The following corollary is immediate from Proposition 5, where the latter shows that the (**SP**) property suffices for claiming that X^* is a singleton. Note that there could well be other settings where such a uniqueness property emerges, albeit locally.

Corollary 12 *Consider the N-player game $\mathcal{G}(\mathbf{f}, X, \boldsymbol{\xi})$ and suppose Assumptions 1 and 2 hold. Let $\{x^k\}_{k=0}^{\infty}$ be generated by **SSGR** or **SAGR**. Suppose (**SP**) holds. Then the claims of Theorem 10 or Theorem 11 hold under the same assumptions.*

3.3. Computing NE for nonconvex potential games

In this subsection, we show that under the pseudoconvexity property (cf. Karamardian and Schaible (1990), Definition 3.2) of the potential function, convergence and rate guarantees for the computation of NE can be provided, rather than just QNE, despite the presence of nonconvexity.

Potential games emerge widely in economic and engineered systems. Recall that we say an N-player game is said to be a potential game (Monderer and Shapley (1996)) if there exists a function $\mathcal{P} : X \to \mathbb{R}$ such that for any $i \in [N]$ and any $x_{-i} \in X_{-i}$, we have $f_i(x_i, x_{-i}) - f_i(y_i, x_{-i}) = \mathcal{P}(x_i, x_{-i}) - \mathcal{P}(y_i, x_{-i})$ for any $x_i, y_i \in X_i$.

Proposition 13 *Consider the N-player game $\mathcal{G}(\mathbf{f}, X, \boldsymbol{\xi})$ with potential function $\mathcal{P}$. The following implications hold if $\mathcal{P}$ is smooth and pseudoconvex:*

$$x^* \in QNE \implies x^* \text{is a B-stationary point of } \mathcal{P} \text{ w.r.t. } X \implies x^* \in NE.$$

Consequently, we may provide the following rate and complexity guarantees for the computation of a Nash equilibrium via **SSGR** and **SAGR**.

Theorem 14 *Consider the N-player game $\mathcal{G}(\mathbf{f}, X, \boldsymbol{\xi})$ and suppose it admits a C^1 and pseudoconvex potential function $\mathcal{P}$. Suppose that Assumptions 1 and 2 hold. Then, all prior convergence and rate guarantees established for computing a QNE reduce to guarantees for computing an NE.*

4. Locally linear rate of SGR under (WS)

In this section, we derive a *locally* linear rate under the (**WS**) property and L-Lipschitz of F, i.e., $\|F(x) - F(y)\| \leqslant L\|x - y\|$ for any x and y, inspired by recent results on deterministic nonconvex optimization (Chen et al. (2022); Davis et al. (2018)). This requires a significant extension to contend with the deterministic game-theoretic regime, but allows for capturing stochastic games over finite sample spaces. The general stochastic extension is not straightforward. Davis et al. (2024) extended their early work (Davis et al. (2018)) to the stochastic setup by using the restart technique, which is essentially different from our one-step scheme here. We leave this for our future research.

In this section, the **SSGR** is specialized to the synchronous gradient response (**SGR**) scheme:

$$x_i^{k+1} = \Pi_{X_i}[x_i^k - \gamma^k \nabla_{x_i} f_i(x_i^k, x_{-i}^k)], \ \forall i \in [N]. \tag{SGR}$$

We still impose Assumption 2-(c) in this section. There are two key ingredients for **SGR** in establishing a locally linear rate under the (**WS**) property: (i) geometrically decaying stepsize: $\gamma^k = \gamma^0 q^k$ for some $\gamma^0 > 0$ and $q \in (0,1)$; (ii) suitable initialization: $\|x^0 - x^*\| \leqslant D$ for some $D > 0$. We begin with a technical lemma.

Lemma 15 *Suppose $e_0 = \frac{(1-\delta)\beta}{NL}$ for some $0 < \delta < 1$ such that $\delta^2\beta^2 < MN$ holds, where β is the (**WS**) parameter, L is the Lipschitz constant of F, and $M := M_2$ is defined in Assumption 2. Suppose $\gamma^0 \in (0, \frac{\sqrt{N}e_0}{2\beta - 2L\sqrt{N}e_0}]$. We choose γ^0 and $q \in (0,1)$ as*

$$(\gamma^0, q) = \begin{cases} \left(\frac{\sqrt{N}e_0}{2\beta - 2L\sqrt{N}e_0}, \frac{\left[(2\beta\sqrt{N} - 2\beta)(2\beta - 2L\sqrt{N}e_0) + M\sqrt{N}\right]^{1/2}}{N^{1/4}(2\beta - 2L\sqrt{N}e_0)} \right), & \frac{\sqrt{N}e_0}{2\beta - 2L\sqrt{N}e_0} < \frac{\beta e_0 - LNe_0^2}{M} \\ \left(\frac{\beta e_0 - LNe_0^2}{M}, \left(1 - \frac{\delta^2\beta^2}{MN}\right)^{1/2} \right). & \frac{\sqrt{N}e_0}{2\beta - 2L\sqrt{N}e_0} \geqslant \frac{\beta e_0 - LNe_0^2}{M} \end{cases}$$

Then the following two claims hold:

$$\text{(i) } \frac{2\beta}{Ne_0} - 2L > 0; \text{ (ii) } 1 - \left(\frac{2\beta}{Ne_0} - 2L\right)\gamma^0 + \frac{M}{Ne_0^2}(\gamma^0)^2 \leqslant q^2.$$

Theorem 16 *Consider the deterministic specialization of the N-player game $\mathcal{G}(\mathbf{f}, X, \boldsymbol{\xi})$. Suppose Assumptions 1 and 2-(c) hold. Consider the sequence $\{x^k\}$ generated by **SGR**. Suppose that the initialization satisfies $x^0 \in \tilde{X} \triangleq \left\{x \in X \mid \|x - x^*\|^2 \leqslant Ne_0^2\right\}$ where x^* is a QNE and the geometric stepsize $\gamma^k = \gamma^0 q^k$ is adopted for any $k \geqslant 0$, where (e_0, γ^0, q) are defined in Lemma 15. Suppose F is L-Lipschitz and the (**WS**) property holds with parameter β on $\tilde{X}$. Then for any $k \geqslant 0$, we have*

$$\|x^k - x^*\|^2 \leqslant Ne_0^2 q^{2k}. \tag{5}$$

Remark 17 (i) *The assumption $\delta^2\beta^2 < MN$ in Lemma 15 is mild since $\delta \in (0,1)$ and generally $M \gg \delta$.* (ii) *We observe that local linear convergence (5) only emerges in a neighborhood of the solution. Naturally, assessing when $\|x - x^*\|^2 \leqslant Ne_0^2$ is difficult since x^* is not available a priori. In fact, a two-stage scheme may hold promise; we maintain the slower sublinear stepsize when $\|x - x^*\|^2 > Ne_0^2$, switching to a geometrically decaying stepsize when $\|x - x^*\|^2 \leqslant Ne_0^2$. Note that this does not necessitate knowing x^* but the non-asymptotic sublinear rate is employed to assess when the required condition is met. We test our two-stage scheme in the numerics section.*

5. Applications

In this section, we consider two different applications: stochastic network congestion problem (Cominetti et al. (2006)) and deterministic nonconvex Nash-Cournot games, satisfing (**QG**) and (**WS**) properties, respectively. We run **SSGR** and **SAGR** on the stochastic network congestion model with three different stepsizes. It is observed that a larger stepsize leads to fewer iterations for the same accuracy requirement, and **SAGR** needs more iterations than **SSGR**. We also compare **SGR** and **Two-Stage SGR** and observe that the **Two-Stage SGR** significantly outperforms **SGR**. The details of the data, the parameter choices, and further numerical experiments may be found in Xiao and Shanbhag (2025).

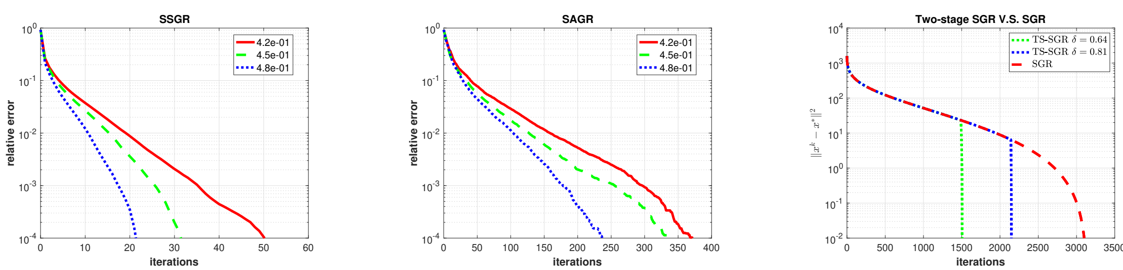

Figure 1: **SSGR** (left) and **SAGR** (center) on a network congestion problem under (**QG**); Two-stage scheme (right) on a nonconvex Nash-Cournot problem under (**WS**).

6. Conclusion

The consideration of nonconvexity in continuous-strategy static noncooperative games is at a relatively nascent stage. Few efficient schemes exist with last-iterate convergence guarantees for contending with games with smooth expectation-valued and potentially nonconvex objectives. To this end, we develop stochastic synchronous and asynchronous gradient response schemes with a.s. convergence and sublinear rate guarantees for computing a QNE under the (**QG**) property. Surprisingly, this claim can be strengthened to computing an NE when overlaying a potentiality and (**SP**). In a deterministic setting, local linear rates can be derived under the (**WS**) property, paving the way for a two-stage asymptotically convergent scheme with fast local convergence. We also provide preliminary numerics to support our theoretical results.

Acknowledgments

We would like to thank anonymous reviewers for their helpful comments. This work was funded in part by the ONR under grants N00014-22-1-2589 and N00014-22-1-2757, and in part by the DOE under grant DE-SC0023303.

References

Sina Arefizadeh and Angelia Nedić. Non-monotone variational inequalities. In *IEEE 60th Annual Allerton Conference on Communication, Control, and Computing*, pages 01–07, 2024.

James V. Burke and Michael C. Ferris. Weak sharp minima in mathematical programming. *SIAM Journal on Control and Optimization*, 31(5):1340–1359, 1993.

Shixiang Chen, Alfredo Garcia, and Shahin Shahrampour. On distributed nonconvex optimization: Projected subgradient method for weakly convex problems in networks. *IEEE Transactions on Automatic Control*, 67(2):662–675, 2022.

Roberto Cominetti, José R. Correa, and Nicolás E. Stier-Moses. Network games with atomic players. In *International colloquium on automata, languages, and programming*, pages 525–536. Springer, 2006.

Ying Cui and Jong-Shi Pang. *Modern nonconvex nondifferentiable optimization*. SIAM, 2021.

Cong D. Dang and Guanghui Lan. On the convergence properties of non-euclidean extragradient methods for variational inequalities with generalized monotone operators. *Computational Optimization and applications*, 60:277–310, 2015.

Damek Davis, Dmitriy Drusvyatskiy, Kellie J. MacPhee, and Courtney Paquette. Subgradient methods for sharp weakly convex functions. *Journal of Optimization Theory and Applications*, 179: 962–982, 2018.

Damek Davis, Dmitriy Drusvyatskiy, and Vasileios Charisopoulos. Stochastic algorithms with geometric step decay converge linearly on sharp functions. *Mathematical Programming*, 207(1): 145–190, 2024.

Francisco Facchinei and Jong-Shi Pang. *Finite-dimensional variational inequalities and complementarity problems. Vol. I.* Springer Series in Operations Research. New York, 2003.

Francisco Facchinei and Jong-Shi Pang. *Nash equilibria: The variational approach*. Convex Optimization in Signal Processing and Communication, Cambridge University Press, 2009.

Drew Fudenberg and David K. Levine. *The theory of learning in games*, volume 2. MIT Press, Cambridge, MA, 1998.

Kevin Huang and Shuzhong Zhang. Beyond monotone variational inequalities: Solution methods and iteration complexities. *Pacific Journal of Optimization*, 20:403–428, 2024.

Alfredo N. Iusem, Alejandro Jofré, Roberto I. Oliveira, and Philip Thompson. Extragradient method with variance reduction for stochastic variational inequalities. *SIAM Journal on Optimization*, 27 (2):686–724, 2017.

Alfredo N. Iusem, Alejandro Jofré, Roberto I. Oliveira, and Philip Thompson. Variance-based extragradient methods with line search for stochastic variational inequalities. *SIAM Journal on Optimization*, 29(1):175–206, 2019.

Aswin Kannan and Uday V. Shanbhag. Optimal stochastic extragradient schemes for pseudomonotone stochastic variational inequality problems and their variants. *Computational Optimization and Applications*, 74(3):779–820, 2019.

Stepan Karamardian and Siegfried Schaible. Seven kinds of monotone maps. *Journal of Optimization Theory and Applications*, 66:37–46, 1990.

Jayash Koshal, Angelia Nedić, and Uday V. Shanbhag. Regularized iterative stochastic approximation methods for stochastic variational inequality problems. *IEEE Transactions on Automatic Control*, 58(3):594–609, 2013.

Georgios Kotsalis, Guanghui Lan, and Tianjiao Li. Simple and optimal methods for stochastic variational inequalities, i: operator extrapolation. *SIAM Journal on Optimization*, 32(3):2041–2073, 2022.

Jinlong Lei and Uday V. Shanbhag. Distributed variable sample-size gradient-response and best-response schemes for stochastic nash equilibrium problems. *SIAM Journal on Optimization*, 32 (2):573–603, 2022a.

Jinlong Lei and Uday V. Shanbhag. Stochastic nash equilibrium problems: Models, analysis, and algorithms. *IEEE Control Systems Magazine*, 42(4):103–124, 2022b.

Jinlong Lei, Uday V. Shanbhag, Jong-Shi Pang, and Suvrajeet Sen. On synchronous, asynchronous, and randomized best-response schemes for stochastic nash games. *Mathematics of Operations Research*, 45(1):157–190, 2020.

Patrice Marcotte and Daoli Zhu. Weak sharp solutions of variational inequalities. *SIAM Journal on Optimization*, 9(1):179–189, 1998.

Dov Monderer and Lloyd S. Shapley. Potential games. *Games and Economic Behavior*, 14(1): 124–143, 1996.

John Nash. Non-cooperative games. *Annals of Mathematics*, 54:286–295, 1951.

Angelia Nedić. Asynchronous broadcast-based convex optimization over a network. *IEEE Transactions on Automatic Control*, 56(6):1337–1351, 2011.

Jong-Shi Pang and Meisam Razaviyayn. A unified distributed algorithm for non-cooperative games. *Big Data over Networks*, pages 101–134, 2016.

Jong-Shi Pang and Gesualdo Scutari. Nonconvex games with side constraints. *SIAM Journal on Optimization*, 21(4):1491–1522, 2011.

Boris T. Polyak. *Introduction to optimization*. New York, Optimization Software, 1987.

Uma Ravat and Uday V. Shanbhag. On the characterization of solution sets of smooth and nonsmooth convex stochastic nash games. *SIAM Journal on Optimization*, 21(3):1168–1199, 2011.

Meisam Razaviyayn. *Successive convex approximation: Analysis and applications*. PhD thesis, University of Minnesota, 2014.

Herbert Robbins and David Siegmund. A convergence theorem for nonnegative almost supermartingales and some applications. *Optimizing Methods in Statistics*, pages 233–257, 1971.

Daniil Vankov, Angelia Nedić, and Lalitha Sankar. Last iterate convergence of popov method for non-monotone stochastic variational inequalities. *OPT2023: 15th Annual Workshop on Optimization for Machine Learning*, 2023.

Zhuoyu Xiao and Uday V. Shanbhag. Computing equilibria in stochastic nonconvex and nonmonotone games via gradient-response schemes. *arXiv preprint arXiv:2504.14056*, 2025.

Farzad Yousefian, Angelia Nedić, and Uday V. Shanbhag. Self-tuned stochastic approximation schemes for non-lipschitzian stochastic multi-user optimization and nash games. *IEEE Transactions on Automatic Control*, 61(7):1753–1766, 2016.