

# ESCAPING LIMIT CYCLES: GLOBAL CONVERGENCE FOR CONSTRAINED NONCONVEX-NONCONCAVE MINIMAX PROBLEMS

**Anonymous authors**

Paper under double-blind review

## ABSTRACT

This paper introduces a new extragradient-type algorithm for a class of nonconvex-nonconcave minimax problems. It is well-known that finding a local solution for general minimax problems is computationally intractable. This observation has recently motivated the study of structures sufficient for convergence of first order methods in the more general setting of variational inequalities when the so-called *weak Minty variational inequality* (MVI) holds. This problem class captures non-trivial structures as we demonstrate with examples, for which a large family of existing algorithms provably converge to limit cycles. Our results require a less restrictive parameter range in the weak MVI compared to what is previously known, thus extending the applicability of our scheme. The proposed algorithm is applicable to constrained and regularized problems, and involves an adaptive step-size allowing for potentially larger stepsizes. Our scheme also converges globally even in settings where the underlying operator exhibits limit cycles. Moreover, a variant with stochastic oracles is proposed—making it directly relevant for training of generative adversarial networks. For the stochastic algorithm only one of the stepsizes is required to be diminishing while the other may remain constant, making it interesting even in the monotone setting.

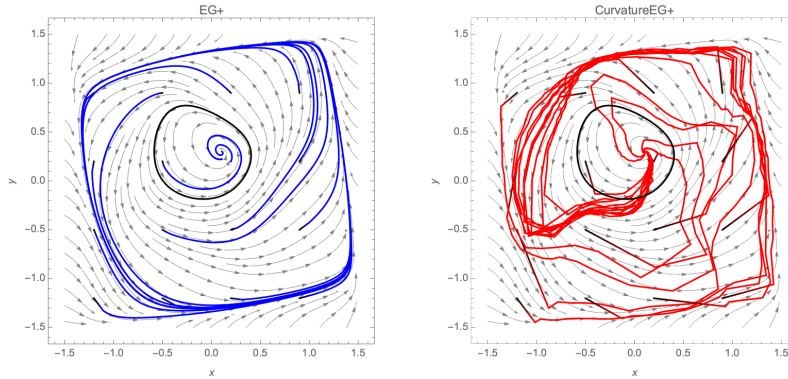
## 1 INTRODUCTION

Many machine learning applications, from generative adversarial networks (GANs) to robust reinforcement learning, result in nonconvex-nonconcave constrained minimax problems, which pose notorious difficulties to the scalable (stochastic) first order methods. Indeed, there is no shortage of results illustrating divergent or cycling behavior when going beyond minimization problems (Be-naim & Hirsch, 1999; Hommes & Ochea, 2012; Mertikopoulos et al., 2018b; Hsieh et al., 2021).

Traditionally, minimax problems have been studied for more than half a century under the umbrella of the variational inequalities (VIs). The extragradient-type algorithms from the VI literature was recently brought to the awareness of the machine learning community (Mertikopoulos et al., 2018a; Gidel et al., 2018; Böhm et al., 2020), and have provided a principled way of stabilizing training and avoiding Poincaré recursions. However, these results mostly concern the convex-concave setting.

In nonconvex-nonconcave minimax problems, or more generally nonmonotone variational inequalities (VIs), even finding a *local* solution is in general intractable. This has been made precise through exponential lower bound of the classical optimization type (Hirsch & Vavasis, 1987) and computational complexity results (Papadimitriou, 1994; Daskalakis et al., 2021b). This is in sharp contrast to minimization problems, where only finding a *global* solution is intractable. The recent result of (Hsieh et al., 2021) provides some intuition behind this difference by showing that the asymptotic limits of most schemes, including extragradient, can converge to attracting limit cycles.

To make progress in lieu of these negative results, Diakonikolas et al. (2021) proposes a simple generalization of extragradient, called (EG+), that can converge to a stationary point even for a class of nonmonotone problems provided that the *weak Minty variational inequality* (MVI) holds. This problem class is parametrized by a constant  $\rho$ , which controls the degree of nonconvexity. However, given the range of  $\rho$  in Diakonikolas et al. (2021), the new class is still too small to include even the simplest counterexample of Hsieh et al. (2021) for the general Robbins-Monro schemes.



**Figure 1:** Forsaken (Hsieh et al., 2021) provides an example where the weak MVI constant  $\rho$  does not satisfy algorithmic requirements of (EG+) and (EG+) does not converge to a stationary point but rather the attracting limit cycle (left). In contrast, adaptively choosing the extrapolation stepsize large enough with our new method, called (CurvatureEG+), is sufficient for avoiding the limit cycles (right). The repellent limit cycle is indicated in black and the stream plot shows the vectorfield  $Fz$ .

**Contributions** Building on the analysis in Diakonikolas et al. (2021), we propose a new adaptive scheme, called (CurvatureEG+), that converges even in the difficult counter example of Hsieh et al. (2021) as illustrated in Fig. 1. Our main contributions are summarized below.

1. We propose an adaptive extragradient-type algorithm that converges for a larger range of  $\rho$ , the parameter in the weak MVI assumption (cf. Assumption I(iii)) than previously known.
2. More importantly, we show that convergence is ensured if  $2\rho + \gamma_k > 0$ , where  $\gamma_k$  is the extrapolation stepsize. This is crucial since by selecting  $\gamma_k$  through a backtracking procedure larger stepsizes are allowed, which in turn implies convergence for more negative values of  $\rho$ , thus capturing a larger class of problems. In addition, we show that the linesearch eventually passes without triggering any backtrack if initialized based on the Jacobian of  $F$  (cf. Section 4).
3. We present a non-adaptive variant of our algorithm (CEG+), and show that for particular parameter choices (EG+) of Diakonikolas et al. (2021), and when  $\rho = 0$  the celebrated forward-backward-forward (FBF) algorithm of Tseng (2000) are recovered, thus unifying and generalizing both methods. We improve upon Diakonikolas et al. (2021) by not only relaxing the problem class but also the stepsize range. We show that our result are tight by providing a matching lower bound, thus providing a complete picture of (EG+) under weak MVI.
4. In the stochastic setting, similarly to Hsieh et al. (2020), we consider two separate stepsizes. Whereas they require both to be diminishing we show that the extrapolation step can in fact be picked constant—a modification which is critical for convergence when only the weak MVI holds.

**Related work** The community have resorted to various approaches to make progress for nonconvex-nonconcave minimax problems. One line of work focuses on deriving local convergence results (Mazumdar et al., 2019; Fiez & Ratliff, 2020; Heusel et al., 2017). For global results the two primary approaches has been to either assume a global oracle for the inner problem (Jin et al., 2019; Davis & Drusvyatskiy, 2018) or assume particular problem structure for the inner problem such as concavity (Rafique et al., 2019) or the Polyak-Łojasiewicz condition (Nouiehed et al., 2019).

We follow the same tradition of assuming structure, but from the general perspective of operator theory. The idea of studying minimax and related problems through the lens of variational inequality has a long history (Minty, 1962; Rockafellar, 1976; Polyak, 1987; Bertsekas, 1997), with recent renewed interest due to its relevance for minimax formulations (Mertikopoulos et al., 2018a; Gidel et al., 2018; Azizian et al., 2020).

One relaxation of the monotone case for which we have positive results is that of Minty variational inequalities (MVI) (Mertikopoulos et al., 2018a; Song et al., 2021; Zhou et al., 2017), which includes all quasiconvex-concave and starconvex-concave problems. Diakonikolas et al. (2021) introduced the relaxed condition of weak MVI. In the unconstrained setting they showed non-asymptotic convergence results under a restricted problem constant  $\rho$ . Similarly to us, Lee & Kim (2021a) extends the regime but under the stronger condition of cohypomonotonicity. They do so by studying a

more evolved variant of extragradient building on anchoring techniques. We instead directly improve upon the arguably simpler (EG+) with less tuning parameters and generalize it to new settings.

In the stochastic setting, usually the stepsize for the extrapolation step is diminishing. This is the case in [Böhm et al. \(2020\)](#) where they consider a forward-backward-forward type scheme. Despite claiming to avoid limit cycles, they remain in the monotone setting, where the limit cycles are non-attracting, as exemplified by a bilinear game. [Hsieh et al. \(2021\)](#) recently showed that a large family, which includes extragradient with diminishing stepsize, can converge to attracting limit cycles. Going beyond this restriction, [Hsieh et al. \(2020\)](#) interestingly considers two separate stepsizes similar to us. However, they only considering the more restrictive setting where MVI is satisfied, and still require both stepsizes to be decreasing.

## 2 PROBLEM FORMULATION AND PRELIMINARIES

In this paper we are interested in finding zeros of an operator (or set-valued mapping)  $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$  that is written as the sum of a Lipschitz continuous (but possibly nonmonotone) operator  $F$  and a maximally monotone operator  $A$ . That is, we wish to find  $z \in \mathbb{R}^n$  such that the general inclusion

$$0 \in Tz := Az + Fz \quad (2.1)$$

holds. The set of all such points is denoted by  $\mathbf{zer} T := \{z \in \mathbb{R}^n \mid 0 \in Tz\}$ . Throughout the paper problem (2.1) is studied under the following assumptions.

**Assumption I.** *In problem (2.1),*

- (i) *Operator  $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$  is a maximally monotone operator.*
- (ii) *Operator  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is  $L$ -Lipschitz continuous.*
- (iii) *Weak Minty inequality holds, i.e., there exists a nonempty set  $S^* \subseteq \mathbf{zer} T$  such that for all  $z^* \in S^*$  and some  $\rho \in (-\frac{1}{2L}, \infty)$*

$$\langle v, z - z^* \rangle \geq \rho \|v\|^2, \quad v \in Tz, \quad \text{for all } z \in \mathbb{R}^n. \quad (2.2)$$

Generally, we do not require the weak Minty assumption to hold at every  $z^* \in \mathbf{zer} T$ . In fact, as shown in [Theorem 3.1](#) nonemptiness of  $S^*$  is sufficient for ensuring that the limit points belong to  $\mathbf{zer} T$ . Interestingly, despite nonmonotonicity of  $F$ , global (as opposed to subsequential) convergence can be established when  $S^* = \mathbf{zer} T$ , an assumption that is still weaker than cohypomonotonicity.

VIs provides a convenient abstraction for a range of problems. We mention some central examples below but otherwise defer to the overview in [Facchinei & Pang \(2007\)](#). Subsequently, we provide examples where the weak MVI holds.

**Example 1:** (minimax optimization). A comprehensive way to capture a wide range of applications in machine learning is to consider structured minimax problems of the form

$$\underset{x \in \mathbb{R}^{n_x}}{\text{minimize}} \underset{y \in \mathbb{R}^{n_y}}{\text{maximize}} \mathcal{L}(x, y) := \varphi(x, y) + g(x) - h(y), \quad (2.3)$$

where  $\varphi$  is not necessarily convex in  $x$  or concave in  $y$ . Functions  $g$  and  $h$  are proper extended real-valued lower semicontinuous and convex, with easy to compute proximal maps. Common examples for  $g$  and  $h$  involve regularizers such as  $\ell_1$ ,  $\ell_2$  norms, or indicator functions of sets allowing us to capture constrained minimax problems. The first order optimality condition associated with this problem may be written in the form of the structured inclusion (2.1) by letting  $Fz = (\nabla_x \varphi(x, y), -\nabla_y \varphi(x, y))$ ,  $Az = (\partial g(x), \partial h(y))$ .

As it will become clear in the next section (cf. [Algorithm 1](#)), the main computations involved in the proposed scheme are evaluations of  $F$  and resolvent  $J_A = (\text{id} + A)^{-1}$  where  $\text{id}$  denoted the identity operator. Recall that the resolvent of a maximally monotone operator is firmly nonexpansive with full domain (cf. [Bauschke & Combettes, 2017](#), Sect. 23)). If  $A = \partial f$  is the subdifferential operator of a convex function  $f$ , then its resolvent is the proximal mapping. For instance when  $A$  is as in [Example 1](#), then its resolvent is given by  $J_A(x, y) = (\mathbf{prox}_g(x), \mathbf{prox}_h(y))$ .

**Example 2:** ( $N$ -player games). More generally, we can consider a continuous game of  $N$  players in normal form. Denote the decision variables  $\mathbf{z} := (z_i; z_{-i}) := (z_1, \dots, z_N)$  and let the loss incurred by

the  $i^{\text{th}}$  player be  $\mathcal{L}_i(z_i; z_{-i}) = \varphi(z) + g_i(z_i)$  where  $\varphi$  is the payoff function and  $g_i$  typically enforce constraints on  $z_i$ . Then we seek a Nash equilibrium, which is any decision which is unilaterally stable, i.e.,

$$\mathcal{L}_i(z_i^*; z_{-i}^*) \leq \mathcal{L}_i(z_i; z_{-i}^*) \quad \forall z_i \text{ and } i \in [N] := \{1, \dots, N\}. \quad (2.4)$$

The corresponding first order optimality conditions may be written as  $Az = (\partial g_1(z_1), \dots, \partial g_N(z_N))$  and  $Fz = (\nabla_{z_1} \varphi(z), \dots, \nabla_{z_N} \varphi(z))$ .

A solution to (2.1) thus returns a candidate for which the first order condition of the above problems is satisfied. In the monotone case these two solution concepts coincide, while in the more general case of weak MVI, we provide examples where this still holds. In particular, we introduce in Section 5 a nonconvex-nonconcave minimax game which additionally exhibits limit cycles for  $Fz$ . As a consequence most schemes including gradient descent ascent, extragradient and optimistic gradient descent ascent do not converge to a stationary point globally (Hsieh et al., 2021). However, the global Nash equilibrium satisfies  $\rho > -1/2L$ , which we show is sufficient for global convergence of (CEG+).

The weak MVI condition is satisfied in certain reinforcement learning settings. Specifically, Diakonikolas et al. (2021); Daskalakis et al. (2021a) considers a two-player zero-sum game where the weak MVI holds, while neither MVI nor cohypomonotonicity holds. Interestingly, the formulation requires constraint—a condition they do not handle. We thus provide the first provable algorithm for this setting. Weak MVI also contains all quasiconvex-concave and starconvex-concave problems. For further examples, the literature on cohypomonotonicity (Bauschke et al., 2020) is relevant since it implies weak MVI, see for instance Lee & Kim (2021b, Example 1).

### 3 GENERALIZING EXTRAGRADIENT+

Our starting point is the Extragradient+ (EG+) algorithm of Diakonikolas et al. (2021) which is identical to extragradient (Korpelevich, 1976) except for the second stepsize being smaller. They only treat the inclusion (2.1) when  $A \equiv 0$ , and in our notation require  $\rho \in (-1/8L, 0]$ . Specifically,

$$\bar{z}^k = z^k - \gamma_k F z^k, \quad z^{k+1} = z^k - \bar{\alpha}_k \gamma_k F \bar{z}^k \quad (\text{EG+})$$

where they choose  $\gamma_k = 1/L$  and  $\bar{\alpha}_k = 1/2$  (Diakonikolas et al., 2021, Thm. 3.2).

We generalize this scheme in Algorithm 1 to take  $Az$  into account—one implication being that we capture constraint and regularized problems as well. In addition, the scheme we consider is adaptive in  $\bar{\alpha}_k$ . This will become useful, when we remove the knowledge of  $L$  completely in Section 4 by making the extrapolation step adaptive as well. We now show that the weaker requirement of  $\rho \in (-1/2L, \infty)$  suffices even for the more general inclusion (2.1).

The main convergence results of Algorithm 1 are established in the next theorem. The proof is largely inspired by recent developments in operator splitting techniques in the framework of monotone inclusions (Latafat & Patrinos, 2017; Giselsson, 2021). The key idea lies in interpreting each iteration of the algorithm as a projection onto a certain hyperplane, an interpretation that dates back to Solodov & Tseng (1996); Solodov & Svaiter (1999).

**Theorem 3.1.** *Suppose that Assumption 1 holds, and let  $\lambda_k \in (0, 2)$ ,  $\gamma_k \in ([-2\rho]_+, 1/L]$ ,  $\delta_k \in (-\gamma_k/2, \rho]$ ,  $\liminf_{k \rightarrow \infty} \lambda_k(2 - \lambda_k) > 0$ , and  $\liminf_{k \rightarrow \infty} (\delta_k + \gamma_k/2) > 0$ . Consider the sequences  $(z^k)_{k \in \mathbb{N}}$ ,  $(\bar{z}^k)_{k \in \mathbb{N}}$  generated by Algorithm 1. Then,*

$$\min_{k=0,1,\dots,m} \|H\bar{z}^k - Hz^k\|^2 \leq \frac{1}{\kappa(m+1)} \|z^0 - z^*\|^2, \quad (3.1)$$

where  $\kappa = \liminf_{k \rightarrow \infty} \lambda_k(2 - \lambda_k) \left( \frac{1}{2} + \frac{\delta_k}{\gamma_k} \right)^2$ . Moreover, the following holds

- (i)  $(\bar{z}^k)_{k \in \mathbb{N}}$  is bounded and its limit points belong to  $\mathbf{zer} T$ ;
- (ii) if in addition  $\limsup_{k \rightarrow \infty} \gamma_k < 1/L$  and  $S^* = \mathbf{zer} T$ , then  $(z^k)_{k \in \mathbb{N}}$ ,  $(\bar{z}^k)_{k \in \mathbb{N}}$  both converge to some  $z^* \in \mathbf{zer} T$ .

Note that whenever  $\limsup_{k \rightarrow \infty} \gamma_k < \frac{1}{L}$ , Lemma 1(ii) may be used to derive a similar inequality in terms of  $\|\bar{z}^k - z^k\|$  by lower bounding  $\|H\bar{z}^k - Hz^k\|$  in (3.1). We also remark that tighter rates may be obtained in the regime  $\rho \geq 0$ , however, this will not be pursued in this work.

**Algorithm 1** Deterministic algorithm for problem (2.1)INITIALIZE  $z^0 = z^{\text{init}} \in \mathbb{R}^n$ ,  $\lambda_k \in (0, 2)$ ,  $\gamma_k \in ([-2\rho]_+, 1/L]$ ,  $\delta_k \in (-\gamma_k/2, \rho]$ ,REPEAT for  $k = 0, 1, \dots$  until convergence1.1: Let  $\bar{z}^k = (\text{id} + \gamma_k A)^{-1}(z^k - \gamma_k Fz^k)$ 

1.2: Compute stepsize

$$\alpha_k = \frac{\delta_k}{\gamma_k} + \frac{\langle \bar{z}^k - z^k, H\bar{z}^k - Hz^k \rangle}{\|H\bar{z}^k - Hz^k\|^2},$$

where  $H = \text{id} - \gamma_k F$ .1.3: Update the vector  $z^{k+1} = z^k + \lambda_k \alpha_k (H\bar{z}^k - Hz^k)$ RETURN  $z^k$ 

## 3.1 NON-ADAPTIVE STEPSIZE VARIANT

Although we do not incur additional costs for evaluating the adaptive stepsize  $\alpha_k$  in [step 1.2](#), it proves instructive to present a variant with constant stepsize. As a result we compare the range of our stepsizes against [Diakonikolas et al. \(2021\)](#) showing an improvement by a factor of  $3/2$ . Moreover, in the monotone case ( $\rho = 0$ ), with a certain choice of stepsizes the algorithm reduces to the celebrated *forward-backward-forward* (FBF) algorithm of [Tseng \(2000\)](#). We remark that the relation of FBF to projection-type algorithms was noted in [Tseng \(2000\)](#), ([Giselsson, 2021](#), Sect. 6.2.1).

To this end, in this subsection consider the following non-adaptive variant of [Algorithm 1](#) that generalizes (EG+). Letting  $\bar{\alpha}_k \in (0, 1 + 2\delta_k/\gamma_k)$ :

$$\bar{z}^k = (\text{id} + \gamma_k A)^{-1}(z^k - \gamma_k Fz^k), \quad z^{k+1} = z^k + \bar{\alpha}_k (H\bar{z}^k - Hz^k). \quad (\text{CEG+})$$

The convergence of this algorithm is an immediate byproduct of [Theorem 3.1](#). To see this, note that the  $z^{k+1}$  update in [step 1.3](#) may be written as  $z^{k+1} = z^k + 2\eta_k \alpha_k (H\bar{z}^k - Hz^k)$ , for  $\eta_k \in (0, 1)$ . Therefore, convergence is still ensured for any  $\bar{\alpha}_k < 2\alpha_k$  as the difference may be absorbed by the relaxation parameter  $\eta_k$ . Note that by  $1/2$ -cocoercivity of  $H$  (cf. [Lemma .1\(i\)](#))

$$\bar{\alpha}_k < \frac{2\delta_k}{\gamma_k} + 1 \leq \frac{2\delta_k}{\gamma_k} + \frac{2\langle H\bar{z}^k - Hz^k, \bar{z}^k - z^k \rangle}{\|H\bar{z}^k - Hz^k\|^2} = 2\alpha_k, \quad (3.2)$$

establishing the validity of the prescribed stepsize range. The convergence of the non-adaptive variant is summarized in the next corollary that for simplicity is stated with constant parameters (dropping subscripts  $k$ ).

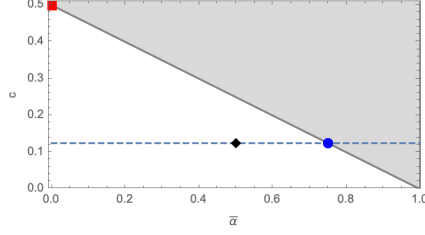
**Corollary 3.2** (Constant stepsize). *Suppose that [Assumption 1](#) holds, and let  $\gamma \in ([-2\rho]_+, 1/L]$ ,  $\delta \in (-\gamma/2, \rho]$ , and  $\bar{\alpha} \in (0, 1 + 2\delta/\gamma)$ . Consider the sequences  $(z^k)_{k \in \mathbb{N}}$ ,  $(\bar{z}^k)_{k \in \mathbb{N}}$  generated according to the update rule (CEG+). Then,*

$$\min_{k=0,1,\dots,m} \|H\bar{z}^k - Hz^k\|^2 \leq \frac{\|z^0 - z^*\|^2}{\kappa(m+1)}, \quad (3.3)$$

where  $\kappa = \bar{\alpha}(1 + \frac{2\delta}{\gamma} - \bar{\alpha})$ . Moreover, the claims of [Theorems 3.1\(i\)](#) and [3.1\(ii\)](#) hold true.

The setting of [Diakonikolas et al. \(2021\)](#) in (EG+) involves the stepsizes  $\gamma_k = 1/L$ ,  $\alpha_k = 1/2$ . Note that when restricting to  $A \equiv 0$ , the iterates (CEG+) simplify to this form owing to the fact that  $H\bar{z}^k - Hz^k = \bar{z}^k - F\bar{z}^k - Hz^k = -\gamma F\bar{z}^k$ . In comparison, in our setting if  $\delta = \rho = -1/8L$  (the smallest  $\rho$  permitted in [Diakonikolas et al. \(2021\)](#)) is selected, then based on our analysis in [Corollary 3.2](#) we may select  $\gamma_k = 1/L$ , and  $\bar{\alpha}_k \in (0, 3/4)$ , thus the upper bound for the second stepsize is  $3/2$  times that of [Diakonikolas et al. \(2021\)](#).

**Remark 3.3** (relation to FBF). In [Corollary 3.2](#) the range of stepsizes  $\gamma$ ,  $\bar{\alpha}$  may alternatively be set as  $\gamma \in ([-2\rho]_+, 1/L)$ ,  $\bar{\alpha} \in (0, 1 + 2\delta/\gamma]$ . This is due to the fact that if  $\gamma < 1/L$  (strictly), then  $H$  is strictly  $1/2$ -cocoercive. Therefore, in (3.2),  $1 + \frac{2\delta}{\gamma} < 2\alpha_k$  holds, and thus the stepsize  $\bar{\alpha} = 1 + \frac{2\delta}{\gamma}$  is permitted. Although this may appear to be of little practical significance, by setting  $\gamma \in (0, 1/L)$ ,  $\delta = \rho = 0$ , and  $\bar{\alpha} = 1$  in (CEG+), we obtain  $z^{k+1} = \bar{z}^k + \gamma Fz^k - \gamma F\bar{z}^k$ , which is the forward-backward-forward (FBF) algorithm of [Tseng \(2000\)](#), ([Bauschke & Combettes, 2017](#), Thm. 26.17)).  $\square$



**Figure 2:** The grey region indicates where convergence provably cannot be guaranteed by [Theorem 3.4](#). Dashed line indicates where  $\rho > -1/8L$ . This is the condition under which ([Diakonikolas et al., 2021, Thm. 3.2](#)) shows the first positive result ( $\blacklozenge$ ). [Corollary 3.2](#) improves their result by matching the lower bound for any  $\bar{\alpha}$ , in particular when  $\rho > -1/8L$  ( $\bullet$ ). The adaptive scheme in [Theorem 3.1](#) matches the smallest possible  $\rho$  for any (EG+) scheme with fixed stepsize ( $\blacksquare$ ).

### 3.2 LOWER BOUNDS

We show that the results in [Corollary 3.2](#) is tight by providing a matching lower bound when  $A = 0$ . We do so by fixing  $\bar{\alpha}_k$  and showing a stepsize dependent lower bound. In particular, we note that if we fix  $\bar{\alpha}_k = 1/2$  as in [Diakonikolas et al. \(2021, Theorem 3.2\)](#), then [Theorem 3.4](#) implies a lower bound of  $\rho > -1/4L$  for the (EG+) scheme. The lower bound is contextualized in [Fig. 2](#) by relating it to ours positive results and existing results in the literature.

**Theorem 3.4.** Consider a sequence  $(z^k)_{k \in \mathbb{N}}$  generated according to (EG+) fixing  $\gamma_k = \gamma = 1/L$  and  $\bar{\alpha}_k = \bar{\alpha} \in (0, 1)$ . Let  $c \geq \frac{1-\bar{\alpha}}{2}$ . Then, there exists an  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ ,  $n > 1$ , satisfying [Assumption I\(ii\)](#) and [Assumption I\(iii\)](#) with  $\rho = -\frac{c}{L} < 0$  for which the sequence will not converge.

### 3.3 STOCHASTIC SETTING

In this section we assume that one has access only to independent samples of  $F$  that will be denoted by  $\hat{F}(\cdot, \xi)$  depending on some random variable  $\xi$  whose distribution is revealed online by observations of i.i.d copies of  $\xi$ . We make the following standard assumptions.

**Assumption II.** For all  $z \in \mathbb{R}^n$

- (i)  $\mathbb{E}_\xi[\hat{F}(z, \xi)] = F(z)$
- (ii)  $\mathbb{E}_\xi[\|\hat{F}(z, \xi) - F(z)\|^2] \leq \sigma^2$

We proceed by presenting [Algorithm 2](#) which is the stochastic variant of (CEG+). Interestingly, our analysis does not impose further assumptions on  $\gamma_k$  (may be selected constant), while it is only  $\alpha_k$  that must satisfy the classical conditions  $\sum_{k=1}^{\infty} \alpha_k = \infty$ , and  $\sum_{k=1}^{\infty} \alpha_k^2 < \infty$  to guarantee convergence.

**Theorem 3.5.** Suppose that [Assumptions I and II](#) hold,  $\gamma_k \in ([-2\rho]_+, 1/L]$ , and let  $\alpha_k \in (0, \frac{1}{2} + \frac{\rho}{\gamma_k})$ . The sequence  $(z^k)_{k \in \mathbb{N}}$  generated by [Algorithm 2](#) satisfy

$$\mathbb{E}[\|z^{k+1} - z^*\|^2] \leq \mathbb{E}[\|z^k - z^*\|^2] - \eta_k \mathbb{E}[\|Hz^k - Hz^k\|^2] + 8\gamma_k^2 \alpha_k^2 \sigma^2, \quad (3.4)$$

where  $\eta_k = 2\alpha_k((\frac{1}{2} + \frac{\rho}{\gamma_k}) - \alpha_k)$ .

---

#### Algorithm 2 Stochastic algorithm for problem (2.1)

---

REQUIRE  $z^0 = z^{\text{init}} \in \mathbb{R}^n$ ,  $\gamma_k \in ([-2\rho]_+, 1/L]$ ,  $\alpha_k = \beta_k(\frac{1}{2} + \frac{\rho}{\gamma_k})$ ,  $\beta_k > 0$   
REPEAT for  $k = 0, 1, \dots$  until convergence  
2.1: Sample  $\xi_k \sim P$  and let  $F_k = \hat{F}(z^k, \xi_k)$   
2.2:  $\bar{z}^k = (\text{id} + \gamma_k A)^{-1}(z^k - \gamma_k F_k)$   
2.3: Sample  $\bar{\xi}_k \sim P$  and let  $\bar{F}_k = \hat{F}(\bar{z}^k, \bar{\xi}_k)$   
2.4:  $z^{k+1} = z^k + \alpha_k((\bar{z}^k - z^k) - \gamma_k(\bar{F}_k - F_k))$   
RETURN  $z^k$

---

Set constant stepsize  $\gamma_k = \gamma \in ([-2\rho]_+, 1/L]$  and  $\alpha_k = \beta_k(\frac{1}{2} + \frac{\rho}{\gamma})$  for some  $(\beta_k)_{k \in \mathbb{N}}$  with  $\beta_k \in (0, 1)$ . Then the following holds

$$\mathbb{E}[\|H\bar{z}^{k_*} - Hz^{k_*}\|^2] \leq \frac{\frac{1}{2}(\frac{1}{2} + \frac{\rho}{\gamma})^{-2}\|z^0 - z^*\|^2 + 4\gamma^2\sigma^2\sum_{k=0}^m\beta_k^2}{\sum_{k=0}^m\beta_k}, \quad (3.5)$$

where  $k_*$  is chosen from  $\{0, 1, \dots, m\}$  according to probability  $\mathcal{P}[k_* = k] = \frac{\beta_k}{\sum_{i=0}^m\beta_i}$ .

As discussed before when  $\gamma < \frac{1}{L}$ , [Lemma .1\(ii\)](#) may be used to derive a similar inequality in terms of  $\mathbb{E}[\|\bar{z}^{k_*} - z^{k_*}\|]$ .

Next, we derive complexity bounds based on [Theorem 3.5](#). Set  $\beta_k = \frac{\xi}{\sqrt{m+1}}$  for some constant  $\xi$ . Assume that  $m$  is large enough such that  $\beta_k < 1$ . By minimizing the left-hand-side in (3.5) we obtain  $\xi = \frac{1}{2\sqrt{2\sigma\gamma}}(\frac{1}{2} + \frac{\rho}{\gamma})^{-1}\|z^0 - z^*\|$ . Therefore, the algorithm will reach  $\mathbb{E}[\|H\bar{z}^k - Hz^k\|] \leq \epsilon$  accuracy for some  $k \in \{0, 1, \dots, m\}$  after at most  $m = \lceil \frac{8}{\epsilon^2}\sigma^2\gamma^2(\frac{1}{2} + \frac{\rho}{\gamma})^2\|z^0 - z^*\|^2 \rceil$  iterations.

We remark that stochastic variant of (EG+) was also studied in ([Diakonikolas et al., 2021](#), Thm. 4.5(i)) under a more restricted range of parameters  $\rho \in (\frac{1}{8\sqrt{2L}}, 0]$  and  $\gamma = \frac{1}{2\sqrt{2L}}$ .

#### 4 ADAPTIVELY TAKING LARGER STEPSIZES USING LOCAL CURVATURE

As made apparent in the analysis in [Section 3](#) (cf. [Appendix .1](#)) the bound on the smallest weak MVI constant  $\rho$  in [Assumption I\(iii\)](#) may be replaced with the requirement that  $\rho > -\kappa/2$  for all  $k \in \mathbb{N}$ . Therefore, larger stepsizes  $\gamma_k$  would guarantee global convergence for an even larger class of problems. Since a *global* Lipschitz constant is inherently pessimistic the natural question then become how to locally choose a maximal stepsize without diverging.

The proposed scheme involves a backtracking linesearch that uses the local curvature for its initial guess. The reason being that this will immediately pass, close enough to the solution  $z^*$ , by argument of continuity. More precisely, we will set the initial guess to something slightly smaller than  $\|JF(z^k)\|^{-1}$ , where  $JF(z)$  denotes the Jacobian of  $F$  at  $z$  and  $\|\cdot\|$  is the spectral norm. Note that, despite the use of second order information, the scheme remains efficient since  $\|JF(z)\|$  only requires one eigenvalue computation performed through Jacobian-vector product ([Pearlmutter, 1994](#)).

Given an initial point  $z^0 = z^{\text{init}}$  and  $\nu \in (0, 1)$ , the final scheme which we denote ([CurvatureEG+](#)) proceeds for  $k = 0, 1, \dots$  as follows:

1. Obtain  $\gamma_k$  and  $\bar{z}^k$  according to [Algorithm 3](#) with  $\gamma^{\text{init}} = \nu\|JF(z^k)\|^{-1}$  ([CurvatureEG+](#))
2. Compute  $z^{k+1}$  according to [steps 1.2 and 1.3](#) of [Algorithm 1](#)

We now make the above intuitive reasoning precise, by showing that the backtracking will terminate in finite time and that  $\gamma^{\text{init}}$  will be immediately accepted asymptotically.

**Lemma 4.1** (Lipschitz constant backtracking). *Suppose that  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is a Lipschitz continuous operator. Consider the linesearch procedure in [Algorithm 3](#). Then,*

- (i) *The linesearch terminates in finite time with  $\gamma \geq \min\{\gamma^{\text{init}}, \nu\tau/L\}$  where  $L$  denotes the Lipschitz constant of  $F$ ;*
- (ii) *Suppose that  $(z^k)_{k \in \mathbb{N}}$  converges to  $z^* \in \mathbf{zer} T$ . If  $F$  is continuously differentiable, and  $\gamma^{\text{init}} \in (0, \nu\|JF(z^k)\|^{-1})$  with  $\nu \in (0, 1)$ , then eventually the backtrack will never be invoked ( $\gamma^{\text{init}}$  would be accepted).*

We illustrate the behavior of this method in [Fig. 1](#) and in [Section 6](#).

#### 5 CONSTRUCTING TOY EXAMPLES

When [Assumption I\(iii\)](#) holds for negative  $\rho$ , limit cycles of the underlying operator  $F(z)$  can emerge. We illustrate this with simple polynomial examples for which all the properties of interest can be computed in closed form.

**Algorithm 3** Lipschitz constant backtracking

---

INITIALIZE  $z^k \in \mathbb{R}^n$ ,  $\tau \in (0, 1)$ ,  $\nu \in (0, 1)$   
**3.1:** Set initial guess  $\gamma = \gamma^{\text{init}}$ , and let  $G_\gamma(z^k) := (\text{id} + \gamma A)^{-1}(z^k - \gamma Fz^k)$   
    **while**  $\gamma \|F(G_\gamma(z^k)) - Fz^k\| > \nu \|G_\gamma(z^k) - z^k\|$  **do**  $\gamma \leftarrow \tau\gamma$   
RETURN  $\gamma_k = \gamma$  and  $\bar{z}^k = G_\gamma(z^k)$

---

**Definition 1** (PolarGame). A PolarGame denotes a two-player game whose associated operator  $F$  has limit cycles at  $\|z\|_2 = c_i$  for all  $i \in [k]$  where  $c_i \neq 0$ .

This turns out to be particularly easy to construct in polar coordinates as the name suggests (see [Appendix B.1](#)). Apart from introducing arbitrary number of limit cycles it also gives us control over  $\rho$ . This is illustrated in the following instantiations capturing three important cases.

**Example 3:** PolarGame Consider  $Fz = (\psi(x, y) - y, \psi(y, x) + x)$  where  $\|z\|_\infty < 1/10$  and  $\psi(x, y) = \frac{1}{4}ax(-1 + x^2 + y^2)(-1 + 4x^2 + 4y^2)$ . In particular, we have the following three cases:

- (i) When  $a = 3/2$  then  $-1/2L > \rho > -1/L$ ;
- (ii) When  $a = 1$  then  $-1/3L > \rho > -1/2L$ ;
- (iii) When  $a = 1/2$  then  $-1/9L > \rho > -1/8L$ .

For all cases  $F$  exhibits limit cycles at  $\|z\| = 1$  and  $\|z\| = 1/2$ . Proof is deferred to [Appendix B.2](#).

**Example 4:** Minimax In the particular case of constrained minimax problem we introduce the following polynomial game:

$$\underset{x:|x|<4/3}{\text{minimize}} \underset{y:|y|<4/3}{\text{maximize}} \phi(x, y) := xy + \psi(x) - \psi(y), \quad (\text{GlobalForsaken})$$

where  $\psi(z) = \frac{2z^6}{21} - \frac{z^4}{3} + \frac{z^2}{3}$ . We provide proof of the following properties in [Appendix B.3](#):

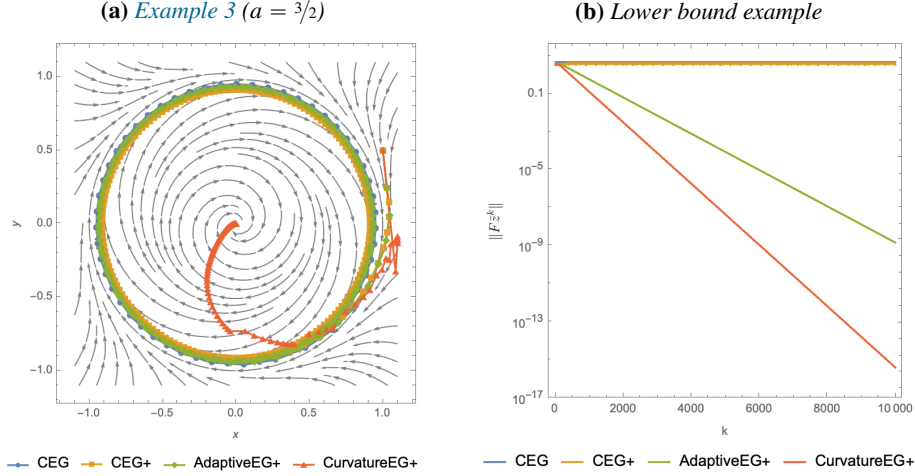
- (i) The origin is a global Nash equilibrium for which [Assumption I\(iii\)](#) holds with  $\rho > -1/2L$ .
- (ii) There exists a repellant limit cycle and an attracting limit cycle of  $F$ .

## 6 EXPERIMENTS

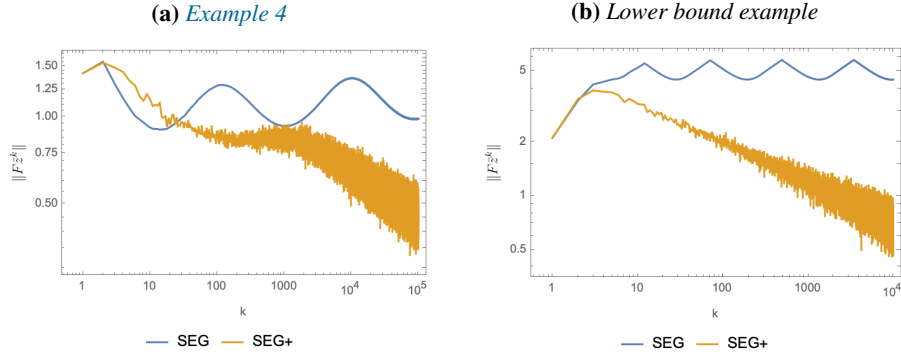
The algorithms considered in the experiments include the adaptive [Algorithm 1](#), ([CurvatureEG+](#)), and constant stepsize methods that can be seen as instances of ([CEG+](#)) for various choices of  $\gamma_k$  and  $\bar{\alpha}_k$ . When  $\gamma_k = 1/L$  and  $\bar{\alpha}_k = 1$  we recover a constrained variant of extragradient, which we denote CEG. When  $\bar{\alpha}_k = 1/2$  we denote the scheme CEG+. This is the direct generalization to the constraint setting of the EG+ scheme studied [Diakonikolas et al. \(2021\)](#). When  $\bar{\alpha}_k$  is chosen adaptively according to [Algorithm 1](#) we refer to it as AdaptiveEG+. Finally, when  $\gamma_k$  is additionally chosen adaptively we use the name ([CurvatureEG+](#)).

In the stochastic setting, we consider [Algorithm 2](#). When  $\gamma_k$  is diminishing and  $\alpha_k = 1$ , effectively both stepsizes diminish, and we recover a constraint variant of the popular stochastic extragradient scheme (see e.g. [Hsieh et al. \(2021, Algorithm 3\)](#)), which we refer to as SEG. When  $\gamma_k$  remains fixed as suggested by [Theorem 3.5](#), and only  $\alpha_k$  is decreasing, we refer to it as SEG+.

We test the algorithms on the constructed examples and confirm their convergence guarantees. Specifically, we apply the algorithms to the minimax problem in [Example 4](#), the PolarGames in [Example 3](#), and a worst case construction, [Example 5](#), from the proof of the lower bound. For [Example 5](#) we choose the problem parameters such that  $\rho = -1/3L$  according to [\(A.7\)](#), and additionally add an  $\ell_\infty$ -ball constraint to keep the iterates bounded. To simulate the stochastic setting we add Gaussian noise to calls of  $F$ . Results for the deterministic setting and stochastic setting can be found in [Fig. 3](#) and [Fig. 4](#) respectively.



**Figure 3: Deterministic setting.** In (a) we have  $\rho < -1/2L$  for which [Theorem 3.4](#) provides lower bound for extrapolation stepsize  $\gamma_k = 1/L$ . However, adaptively choosing  $\gamma_k$  larger can converge as illustrated with [\(CurvatureEG+\)](#). The remaining PolarGames can be found in [Fig. 6](#). In addition, (b) confirms that CEG and CEG+ may indeed not converge even when  $\rho = -1/3L$ . In contrast, both AdaptiveEG+ and [\(CurvatureEG+\)](#) converges to the stationary point.



**Figure 4: Stochastic setting.** In (a) we test the stochastic algorithms on our nonconcave-nonconvex constrained minimax example. The cycling behavior of SEG is inline with [Hsieh et al. \(2021\)](#), who shows that the sequence generated by SEG can converge to limit cycles of the underlying operator  $F$ . On the other hand, SEG+ converges in terms of  $\|Fz^k\|$ . In (b) we also provide a more challenging example motivated by our lower bound. Nonetheless, SEG+ still converges in accordance with [Theorem 3.5](#).

## 7 CONCLUSION

This paper introduced an EG-type algorithm for a class of nonconvex-nonconcave minimax problems that satisfy the *weak Minty variational inequality* (MVI) assumption. The range of parameter in the weak MVI was extended compared to EG+ of [Diakonikolas et al. \(2021\)](#), and tightness of our results were demonstrated through construction of counter examples. In addition, EG+ ([Diakonikolas et al., 2021](#)), as well as the forward-backward-forward algorithm ([Tseng, 2000](#)) were all shown to be special cases of our scheme. Furthermore, [\(CurvatureEG+\)](#) was proposed that performs a backtracking linesearch on the extrapolation stepsize  $\gamma_k$  allowing for larger stepsizes and relaxes the condition  $\rho > \frac{1}{2L}$  to  $\rho > -\gamma_k/2$  which is often a much weaker condition. More importantly, it is shown that asymptotically the linesearch always passes with  $\gamma_k = \nu \|JF(z^k)\|^{-1}$  for any  $\nu \in (0, 1)$ , thus ratifying the name [\(CurvatureEG+\)](#). In the stochastic setting, unlike what is common in the literature, it was shown that it is only the second stepsize that must be diminishing. Future direction include exploring applications of the proposed algorithm in particular in the setting of GANs. It is also interesting to develop a variance reduced variant of the algorithm for finite sum minimax problems.

## REFERENCES

- Waïss Azizian, Ioannis Mitliagkas, Simon Lacoste-Julien, and Gauthier Gidel. A tight and unified analysis of gradient-based methods for a whole spectrum of differentiable games. In *International Conference on Artificial Intelligence and Statistics*, pp. 2863–2873. PMLR, 2020.
- Heinz H. Bauschke and Patrick L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*. CMS Books in Mathematics. Springer, 2017. ISBN 978-3-319-48310-8.
- Heinz H Bauschke, Walaa M Moursi, and Xianfu Wang. Generalized monotone operators and their averaged resolvents. *Mathematical Programming*, pp. 1–20, 2020.
- Michel Benaim and Morris W Hirsch. Mixed equilibria and dynamical systems arising from fictitious play in perturbed games. *Games and Economic Behavior*, 29(1-2):36–72, 1999.
- Dimitri P Bertsekas. Nonlinear programming. *Journal of the Operational Research Society*, 48(3): 334–334, 1997.
- Axel Böhm, Michael Sedlmayer, Ernő Robert Csetnek, and Radu Ioan Boț. Two steps at a time—taking gan training in stride with tseng’s method. *arXiv preprint arXiv:2006.09033*, 2020.
- Constantinos Daskalakis, Dylan J Foster, and Noah Golowich. Independent policy gradient methods for competitive reinforcement learning. *arXiv preprint arXiv:2101.04233*, 2021a.
- Constantinos Daskalakis, Stratis Skoulakis, and Manolis Zampetakis. The complexity of constrained min-max optimization. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pp. 1466–1478, 2021b.
- Damek Davis and Dmitriy Drusvyatskiy. Stochastic subgradient method converges at the rate  $\mathcal{O}(k^{-1/4})$  on weakly convex functions. *arXiv:1802.02988 [cs, math]*, February 2018.
- Jelena Diakonikolas, Constantinos Daskalakis, and Michael Jordan. Efficient methods for structured nonconvex-nonconcave min-max optimization. In *International Conference on Artificial Intelligence and Statistics*, pp. 2746–2754. PMLR, 2021.
- Francisco Facchinei and Jong-Shi Pang. *Finite-dimensional variational inequalities and complementarity problems*. Springer Science & Business Media, 2007.
- Tanner Fiez and Lillian Ratliff. Gradient descent-ascent provably converges to strict local minmax equilibria with a finite timescale separation. *arXiv preprint arXiv:2009.14820*, 2020.
- Gauthier Gidel, Hugo Berard, Gaëtan Vignoud, Pascal Vincent, and Simon Lacoste-Julien. A variational inequality perspective on generative adversarial networks. *arXiv preprint arXiv:1802.10551*, 2018.
- Pontus Giselsson. Nonlinear forward-backward splitting with projection correction. *SIAM Journal on Optimization*, 31(3):2199–2226, 2021. doi: 10.1137/20M1345062.
- Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017.
- M Hirsch and S Vavasis. Exponential lower bounds for finding Brouwer fixed points. In *Proceedings of the 28th Symposium on Foundations of Computer Science*, pp. 401–410, 1987.
- Cars H Hommes and Marius I Ochea. Multiple equilibria and limit cycles in evolutionary games with logit dynamics. *Games and Economic Behavior*, 74(1):434–441, 2012.
- Ya-Ping Hsieh, Panayotis Mertikopoulos, and Volkan Cevher. The limits of min-max optimization algorithms: Convergence to spurious non-critical sets. In *International Conference on Machine Learning*, pp. 4337–4348. PMLR, 2021.
- Yu-Guan Hsieh, Franck Iutzeler, Jérôme Malick, and Panayotis Mertikopoulos. Explore aggressively, update conservatively: Stochastic extragradient methods with variable stepsize scaling. *arXiv preprint arXiv:2003.10162*, 2020.

- Chi Jin, Praneeth Netrapalli, and Michael I. Jordan. What is local optimality in nonconvex-nonconcave minimax optimization? *arXiv:1902.00618 [cs, math, stat]*, June 2019.
- Galina M Korpelevich. The extragradient method for finding saddle points and other problems. *Matecon*, 12:747–756, 1976.
- Puya Latafat and Panagiotis Patrinos. Asymmetric forward–backward–adjoint splitting for solving monotone inclusions involving three operators. *Computational Optimization and Applications*, 68(1):57–93, Sep 2017.
- Sucheol Lee and Donghwan Kim. Fast extra gradient methods for smooth structured nonconvex-nonconcave minimax problems. *arXiv preprint arXiv:2106.02326*, 2021a.
- Sucheol Lee and Donghwan Kim. Semi-anchored multi-step gradient descent ascent method for structured nonconvex-nonconcave composite minimax problems. *arXiv preprint arXiv:2105.15042*, 2021b.
- Eric V. Mazumdar, Michael I. Jordan, and S. Shankar Sastry. On finding local Nash equilibria (and only local Nash equilibria) in zero-sum games. *arXiv:1901.00838 [cs, math, stat]*, January 2019.
- Panayotis Mertikopoulos, Bruno Lecouat, Houssam Zenati, Chuan-Sheng Foo, Vijay Chandrasekhar, and Georgios Piliouras. Optimistic mirror descent in saddle-point problems: Going the extra (gradient) mile. *arXiv preprint arXiv:1807.02629*, 2018a.
- Panayotis Mertikopoulos, Christos Papadimitriou, and Georgios Piliouras. Cycles in adversarial regularized learning. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 2703–2717. SIAM, 2018b.
- George J Minty. Monotone (nonlinear) operators in hilbert space. *Duke Mathematical Journal*, 29(3):341–346, 1962.
- Maher Nouiehed, Maziar Sanjabi, Tianjian Huang, Jason D Lee, and Meisam Razaviyayn. Solving a class of non-convex min-max games using iterative first order methods. *arXiv preprint arXiv:1902.08297*, 2019.
- Christos H Papadimitriou. On the complexity of the parity argument and other inefficient proofs of existence. *Journal of Computer and system Sciences*, 48(3):498–532, 1994.
- Barak A Pearlmutter. Fast exact multiplication by the hessian. *Neural computation*, 6(1):147–160, 1994.
- Boris T Polyak. *Introduction to optimization*. Optimization Software New York, 1987.
- Hassan Rafique, Mingrui Liu, Qihang Lin, and Tianbao Yang. Non-convex min-max optimization: Provable algorithms and applications in machine learning. *arXiv:1810.02060 [cs, math]*, January 2019.
- R. T. Rockafellar and R. J.-B. Wets. *Variational analysis*, volume 317. Springer Science & Business Media, 2009.
- R Tyrrell Rockafellar. Monotone operators and the proximal point algorithm. *SIAM journal on control and optimization*, 14(5):877–898, 1976.
- M. V. Solodov and P. Tseng. Modified projection-type methods for monotone variational inequalities. *SIAM Journal on Control and Optimization*, 34(5):1814–1830, 1996.
- Mikhail V Solodov and Benar F Svaiter. A hybrid projection-proximal point algorithm. *Journal of convex analysis*, 6(1):59–70, 1999.
- Chaobing Song, Zhengyuan Zhou, Yichao Zhou, Yong Jiang, and Yi Ma. Optimistic dual extrapolation for coherent non-monotone variational inequalities. *arXiv preprint arXiv:2103.04410*, 2021.
- Gerald Teschl. *Ordinary differential equations and dynamical systems*, volume 140. American Mathematical Soc., 2012.

P. Tseng. A modified forward-backward splitting method for maximal monotone mappings. *SIAM Journal on Control and Optimization*, 38(2):431–446, 2000.

Zhengyuan Zhou, Panayotis Mertikopoulos, Nicholas Bambos, Stephen Boyd, and Peter W Glynn. Stochastic mirror descent in variationally coherent optimization problems. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

## .1 PROOFS OF SECTION 3

**Proof of Theorem 3.1.** Let  $H = \text{id} - \gamma_k F$ . By Step 1.1  $Hz^k \in \bar{z}^k + \gamma_k A\bar{z}^k$ . Therefore,

$$\frac{1}{\gamma_k}(Hz^k - H\bar{z}^k) \in A\bar{z}^k + F\bar{z}^k \quad (.1)$$

In what follows we will show that Algorithm 1 is equivalent to taking a forward-backward step followed by a correction step. Consider the updates

$$\bar{z}^k := (\text{id} + \gamma_k A)^{-1}(z^k - \gamma_k Fz^k), \quad (.2)$$

$$z^{k+1} = (1 - \lambda)z^k + \lambda \Pi_{\mathcal{D}_k}^S(z^k), \quad \text{where } \mathcal{D}_k := \left\{ w \mid \langle Hz^k - H\bar{z}^k, \bar{z}^k - w \rangle \geq \frac{\sigma_k}{\gamma_k} \|Hz^k - H\bar{z}^k\|^2 \right\}.$$

Note that

$$\langle H\bar{z}^k - Hz^k, \bar{z}^k - z^k \rangle + \frac{\sigma_k}{\gamma_k} \|H\bar{z}^k - Hz^k\|^2 \geq \left(\frac{1}{2} + \frac{\sigma_k}{\gamma_k}\right) \|H\bar{z}^k - Hz^k\|^2 \quad (.3)$$

where in the first inequality Lemma .1(i) was used. Hence, by (.3) the stepsize  $\alpha_k$  is positive and bounded away from zero. Moreover, if  $z^k \in \mathcal{D}_k$ , then from (.3) we may conclude that  $\|H\bar{z}^k - Hz^k\| \leq 0$  which implies that the generated sequence remains constant and  $\bar{z}^k \in \mathbf{zer} T$  (cf. (.1)).

The projection onto  $\mathcal{D}_k$  for any  $v \notin \mathcal{D}_k$  is given by

$$\Pi_{\mathcal{D}_k}^S(v) = v + \frac{-\frac{\sigma_k}{\gamma_k} \|Hz^k - H\bar{z}^k\|^2 + \langle \bar{z} - v, Hz^k - H\bar{z}^k \rangle}{\|Hz^k - H\bar{z}^k\|^2} (Hz^k - H\bar{z}^k)$$

Moreover, (.1) together with Assumption I(iii) at  $\bar{z}$  yields

$$\frac{1}{\gamma_k} \langle Hz^k - H\bar{z}^k, \bar{z}^k - z^* \rangle \geq \frac{\rho}{\gamma_k^2} \|Hz^k - H\bar{z}^k\|^2 \geq \frac{\sigma_k}{\gamma_k^2} \|Hz^k - H\bar{z}^k\|^2, \quad (.4)$$

thus ensuring  $z^* \in \mathcal{S}^* \subseteq \mathcal{D}_k$ . The projection onto  $\mathcal{D}_k$  is then given by  $\Pi_{\mathcal{D}_k}(z^k) = z^k + \alpha_k(H\bar{z}^k - Hz^k)$ , where  $\alpha_k$  is as in step 1.2.

Finally, since the projection  $\Pi_{\mathcal{D}_k}$  is firmly nonexpansive, it follows from (Bauschke & Combettes, 2017, Cor. 4.41) that the mapping  $(1 - \lambda_k)\text{id} + \lambda_k \Pi_{\mathcal{D}_k}$  is  $\lambda_k/2$ -averaged. Consequently, we may conclude that  $(z^k)_{k \in \mathbb{N}}$  is Fejér monotone relative to  $\mathcal{S}^*$  (Bauschke & Combettes, 2017, Prop. 4.35). That is for all  $z^* \in \mathcal{S}^*$

$$\begin{aligned} \|z^{k+1} - z^*\|^2 &\leq \|z^k - z^*\|^2 - \lambda_k(2 - \lambda_k)\alpha_k^2 \|H\bar{z}^k - Hz^k\|^2. \\ (.3) &\leq \|z^k - z^*\|^2 - \varepsilon_k \|H\bar{z}^k - Hz^k\|^2, \end{aligned}$$

where  $\varepsilon_k := \lambda_k(2 - \lambda_k)\left(\frac{1}{2} + \frac{\sigma_k}{\gamma_k}\right)^2$ . Since  $\liminf_{k \rightarrow \infty} \varepsilon_k > 0$  we may conclude that  $(\|H\bar{z}^k - Hz^k\|)_{k \in \mathbb{N}}$  converges to zero. Moreover, since  $(\|z^k - z^*\|)_{k \in \mathbb{N}}$  converges, the sequence  $(z^k)_{k \in \mathbb{N}}$  is bounded, and thus there exists an index set  $K \subseteq \mathbb{N}$  such that  $(z^k)_{k \in K}$  converges to some  $z \in \mathbb{R}^n$ . Since  $\gamma_k$  is bounded, and  $F$  and the resolvents  $(\text{id} + \gamma_k A)^{-1}$  are Lipschitz continuous (cf. (Bauschke & Combettes, 2017, Cor. 23.9)), so is their composition. Hence,  $(\bar{z}^k)_{k \in K}$  is also bounded and converges to some  $\bar{z} \in \mathbb{R}^n$ . Combined with the fact that  $(\|H\bar{z}^k - Hz^k\|)_{k \in \mathbb{N}}$  converges to zero, we may conclude from (.1) along with (Bauschke & Combettes, 2017, Prop. 20.38) and Lipschitz continuity of  $F$  that  $\bar{z} \in \mathbf{zer} T$ . Finally, if in addition  $\limsup_{k \rightarrow \infty} \gamma_k < 1/L_F$ , then  $(1 - \gamma_k L_F)\|\bar{z}^k - z^k\| \leq \|H\bar{z}^k - Hz^k\|$  (invoke Lemma .1(ii)). Therefore, also  $(\|\bar{z}^k - z^k\|)_{k \in \mathbb{N}}$  converges to zero, and thus  $z \in \mathbf{zer} T$ . Consequently, if Assumption I(iii) holds at all of the zeros of  $T$ , i.e., if  $\mathcal{S}^* = \mathbf{zer} T$ , then the second claim follows by invoking (Bauschke & Combettes, 2017, Thm. 5.5). □

**Proof of Corollary 3.2 (Constant stepsize).** The proof of convergence was already given prior to the statement of the corollary. It remains to derive (3.3). By Assumption I(iii) and owing to  $1/2$ -cocoercivity of  $H$  (cf. Lemma .1(i))

$$\begin{aligned} \langle z^k - z^*, H\bar{z}^k - Hz^k \rangle &= \langle \bar{z}^k - z^*, H\bar{z}^k - Hz^k \rangle + \langle z^k - \bar{z}^k, H\bar{z}^k - Hz^k \rangle \\ (.4) &\leq -\left(\frac{1}{2} + \frac{\rho}{\gamma_k}\right) \|H\bar{z}^k - Hz^k\|^2. \end{aligned} \quad (.5)$$

Therefore, provided that  $\bar{\alpha} > 0$  we have

$$\|z^{k+1} - z^*\|^2 = \|z^k - z^*\|^2 + \bar{\alpha}^2 \|H\bar{z}^k - Hz^k\|^2 + 2\bar{\alpha} \langle z^k - z^*, H\bar{z}^k - Hz^k \rangle$$

$$(5) \leq \|z^k - z^*\|^2 - \bar{\alpha}(2(\frac{1}{2} + \frac{\rho}{\gamma}) - \bar{\alpha})\|H\bar{z}^k - Hz^k\|^2.$$

Telescoping the above inequality yields the claimed inequality.  $\square$

**Proof of Theorem 3.5.** Let  $w^k = \hat{F}(z^k, \xi_k) - F(z^k)$ ,  $\bar{w}^k = \hat{F}(\bar{z}^k, \bar{\xi}_k) - F(\bar{z}^k)$  denote the additive noise associated with the stochastic oracle access at  $z^k$  and  $\bar{z}^k$ , respectively. Then, letting  $\hat{H}(z^k) = H(z^k) - \gamma_k w^k$ ,  $\hat{H}(\bar{z}^k) = H(\bar{z}^k) - \gamma_k \bar{w}^k$ , it holds that

$$\begin{aligned} \|\hat{H}(\bar{z}^k) - \hat{H}(z^k)\|^2 &= \|\gamma_k(w^k - \bar{w}^k) + H(\bar{z}^k) - H(z^k)\|^2 \\ &\leq 2\gamma_k^2\|w^k - \bar{w}^k\|^2 + 2\|H(\bar{z}^k) - H(z^k)\|^2 \\ &\leq 4\gamma_k^2(\|w^k\| + \|\bar{w}^k\|)^2 + 2\|H(\bar{z}^k) - H(z^k)\|^2, \end{aligned} \quad (6)$$

where the Young inequality was used in both inequalities. Using the law of total expectation, for  $z^* \in \mathbf{zer} T$

$$\begin{aligned} \mathbb{E}[\langle z^k - z^*, \hat{H}(\bar{z}^k) - \hat{H}(z^k) \rangle] &= \mathbb{E}[\mathbb{E}[\langle z^k - z^*, \hat{H}(\bar{z}^k) \rangle | \bar{\mathcal{F}}_k]] + \mathbb{E}[\mathbb{E}[\langle z^k - z^*, -\hat{H}(z^k) \rangle | \mathcal{F}_k]] \\ \text{Assumption II(i)} &= \mathbb{E}[\langle z^k - z^*, H(\bar{z}^k) - H(z^k) \rangle] \\ (5) &\leq -(\frac{1}{2} + \frac{\rho}{\gamma_k})\mathbb{E}[\|H\bar{z}^k - Hz^k\|^2]. \end{aligned} \quad (7)$$

where  $\mathcal{F}_k$  and  $\bar{\mathcal{F}}_k$  represent the information available at  $z^k$  and  $\bar{z}^k$  updates, respectively. It follows from the above two inequalities that

$$\begin{aligned} \mathbb{E}[\|z^{k+1} - z^*\|^2] &= \mathbb{E}[\|z^k - z^*\|^2] + \alpha_k^2\mathbb{E}[\|\hat{H}(\bar{z}^k) - \hat{H}(z^k)\|^2] \\ &\quad + 2\alpha_k\mathbb{E}[\langle z^k - z^*, \hat{H}(\bar{z}^k) - \hat{H}(z^k) \rangle] \\ (7), (6), \text{Assumption II(ii)} &\leq \mathbb{E}[\|z^k - z^*\|^2] - 2\alpha_k(\frac{1}{2} + \frac{\rho}{\gamma_k})\mathbb{E}[\|H(\bar{z}^k) - H(z^k)\|^2] \\ &\quad + 2\alpha_k^2\mathbb{E}[\|H(\bar{z}^k) - H(z^k)\|^2] + 8\gamma_k^2\alpha_k^2\sigma^2, \end{aligned}$$

as claimed. Next, with constant stepsize  $\gamma$  and  $(\alpha_k)_{k \in \mathbb{N}}$  as prescribed in the lemma,  $\eta_k = 2\beta_k(1 - \beta_k)(\frac{1}{2} + \frac{\rho}{\gamma})^2 \geq 2\beta_k(\frac{1}{2} + \frac{\rho}{\gamma})^2$ . Using this and summing (3.4) over  $k$  we obtain

$$2(\frac{1}{2} + \frac{\rho}{\gamma})^2 \sum_{k=0}^M \beta_k \mathbb{E}[\|H\bar{z}^k - Hz^k\|^2] \leq \mathbb{E}[\|z^0 - z^*\|^2] + 4\sigma^2\gamma(1 + \gamma)(\frac{1}{2} + \frac{\rho}{\gamma})^2 \sum_{k=0}^M \beta_k^2.$$

Dividing both sides by  $2(\frac{1}{2} + \frac{\rho}{\gamma})^2 \sum_{k=0}^M \beta_k$  yields

$$\frac{1}{\sum_{k=0}^M \beta_k} \sum_{k=0}^M \beta_k \mathbb{E}[\|H\bar{z}^k - Hz^k\|^2] \leq \frac{\frac{1}{2}(\frac{1}{2} + \frac{\rho}{\gamma})^{-2}\|z^0 - z^*\|^2 + 2\gamma(1 + \gamma)\sigma^2 \sum_{k=0}^M \beta_k^2}{\sum_{k=0}^M \beta_k},$$

establishing the claimed inequality.  $\square$

**Proof of Lemma 4.1 (Lipschitz constant backtracking).** **4.1(i):** Since  $F$  is  $L$ -Lipschitz continuous the linesearch would terminate in finite steps. Either  $\gamma^{\text{init}}$  satisfies the condition, or else the back-track procedure is invoked, which in turn implies the previous candidate  $\gamma/\tau$  should have violated the condition leading to the claimed lower bound.

**4.1(ii):** Since the resolvent  $(\text{id} + \gamma A)^{-1}$  and  $F$  are Lipschitz continuous, so is their composition. Hence,  $G_\gamma(z^k) \rightarrow G(z^*)$ . Furthermore, by definition  $z^* - \gamma Fz^* \in G_\gamma(z^*) + \gamma A(G_\gamma(z^*))$ . Consequently, using monotonicity of  $A$  at  $G_\gamma(z^*)$  and  $z^*$ , and that  $-Fz^* \in Az^*$  yields  $0 \leq \langle z^* - \gamma Fz^* - G_\gamma(z^*) + Fz^*, G_\gamma(z^*) - z^* \rangle = -\|z^* - G_\gamma(z^*)\|^2$ . Thus  $G_\gamma(z^*) = z^*$ . Using the fact that both  $(G_\gamma(z^k))_{k \in \mathbb{N}}$  and  $(z^k)_{k \in \mathbb{N}}$  converges to  $z^* \in \mathbf{zer} T$ :

$$\lim_{k \rightarrow \infty} \frac{\|F(G_\gamma(z^k)) - Fz^k\|}{\|G_\gamma(z^k) - z^k\|} \leq \limsup_{z, z' \rightarrow z^*} \frac{\|Fz' - Fz\|}{\|z' - z\|} = \text{lip} F(z^*) = \|JF(z^*)\|,$$

where (Rockafellar & Wets, 2009, Thm. 9.7) was used. The claim follows from continuity of  $JF$  and the fact that  $(z^k)_{k \in \mathbb{N}}$  converges to  $z^*$ .  $\square$

## .2 AUXILIARY LEMMAS

**Lemma .1.** Let  $A : \mathbb{R}^n \rightarrow \mathbb{R}^n$  denote a single valued operator. Then,

- (i)  $A$  is 1-Lipschitz if and only if  $T = \text{id} - A$  is  $1/2$ -cocoercive.
- (ii) If  $A$  is  $L$ -Lipschitz, then  $T = \text{id} - \eta A$ ,  $\eta \in (0, 1/L)$ , is  $(1 - \eta L)$ -strongly monotone, and in particular  $\|Tu - Tv\| \geq (1 - \eta L)\|u - v\|$  for all  $u, v \in \mathbb{R}^n$ .

*Proof.* The first claim follows directly from (Bauschke & Combettes, 2017, Prop.4.11). That  $T$  is strongly monotone is a consequence of the Cauchy Schwarz inequality and Lipschitz continuity of  $A$ :

$$\langle Tv - Tu, v - u \rangle = \|v - u\|^2 - \eta \langle Av - Au, v - u \rangle \geq (1 - \eta L)\|v - u\|^2.$$

In turn, the last claim follows from the Cauchy-Schwarz inequality.  $\square$

## A LOWER BOUNDS

To prove the lower bound we introduce the following unconstrained bilinear minimax problem with an unstable critical point.

**Example 5:** Consider the following minimax problem:

$$\underset{x \in \mathbb{R}}{\text{minimize}} \underset{y \in \mathbb{R}}{\text{maximize}} f(x, y) := axy + \frac{b}{2}(x^2 - y^2), \quad (\text{A.1})$$

where  $b < 0$  and  $a > 0$ .

*Proof of Theorem 3.4.* The associated operator of Example 5 can easily be computed,

$$F(z) = (ay + bx, by - ax), \quad (\text{A.2})$$

where  $z = (x, y)$ . In this particular case, both  $L$  and  $\rho$  turns out to be constants. By simple calculation we have,

$$\|JF(z)\| = \frac{b}{a^2 + b^2}, \quad \rho = \sqrt{a^2 + b^2} \quad (\text{A.3})$$

where  $\|\cdot\|$  is the spectral norm. Since the curvature is constant it equates the global Lipschitz constant,  $L = \|JF(z)\|$ .

By linearity of  $F$ , one step of (EG+) is conveniently also a linear operator. Specifically,

$$z^{k+1} = Tz^k \quad \text{with} \quad T := \begin{pmatrix} \frac{(1-\bar{\alpha})a^2+b(-\bar{\alpha}\sqrt{a^2+b^2}+\bar{\alpha}b+b)}{a^2+b^2} & -\frac{a\bar{\alpha}(\sqrt{a^2+b^2}-2b)}{a^2+b^2} \\ \frac{a\bar{\alpha}(\sqrt{a^2+b^2}-2b)}{a^2+b^2} & \frac{(1-\bar{\alpha})a^2+b(-\bar{\alpha}\sqrt{a^2+b^2}+\bar{\alpha}b+b)}{a^2+b^2} \end{pmatrix}. \quad (\text{A.4})$$

We know that a linear dynamical system is globally asymptotically stable if and only if the spectral radius of the linear mapping is strictly less than 1.

Let  $\lambda_1, \lambda_2$  be the eigenvalues of  $T$ . Then the spectral radius is the largest absolute value of the eigenvalues. For  $T$  this becomes,

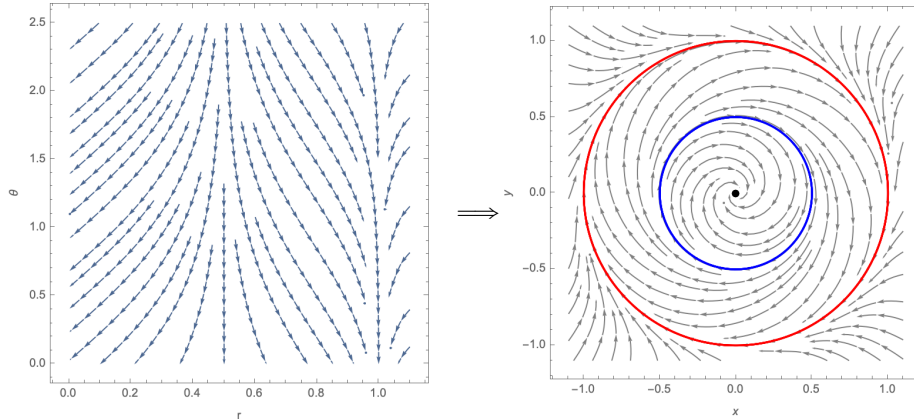
$$\max_i |\lambda_i| = \sqrt{\frac{(2(\bar{\alpha}-1)\bar{\alpha}+1)a^2 - 2\bar{\alpha}(\bar{\alpha}+1)b(\sqrt{a^2+b^2}-b) + b^2}{a^2+b^2}}. \quad (\text{A.5})$$

So we can ask what  $c$  in  $\rho = -\frac{c}{L}$  needs to be for the sequence  $(z^k)_{k \in \mathbb{N}}$  to converge. Solving for  $c$  in this equality with  $\max_i |\lambda_i| < 1$ , we obtain,

$$c < \frac{1 - \bar{\alpha}}{2}, \quad (\text{A.6})$$

provided that we pick

$$\frac{\sqrt{1-c^2}}{c} = -\frac{a}{b}. \quad (\text{A.7})$$



**Figure 5:** We can construct the desired properties in polar coordinates  $(r, \theta)$  and then subsequently transform it into a vectorfield in cartesian coordinates  $(x, y)$ . This is illustrated by PolarGame with attracting limit cycles at radius  $\|z\| = 1$  and repellant limit cycle at  $\|z\| = 1/2$  for the associated operator  $F(z)$  as indicated in red and blue respectively.

The left hand side of (A.7) is positive since  $c > 0$  by assumption and  $c < \frac{1-\bar{\alpha}}{2} < 1/2$  for  $\bar{\alpha} \in (0, 1)$ . By assumption on  $a$  and  $b$  the right hand side span the hole positive reals. Thus the equality in (A.7) is feasible and provides a specification for the worst case example that we can still converge in with (EG+). On the other hand, since it is a linear system, we simultaneously learn that picking  $c$  any larger would imply non convergence through  $\max_i |\lambda_i| \geq 1$  (given  $z^0 \neq 0$ ). We can trivially embed problem (A.4) into a higher dimension to generalize the result. This completes the proof.  $\square$

We provide Mathematica code to verify each step of the above proof.

## B TOY EXAMPLES

### B.1 CONSTRUCTING A POLARGAME (DEFINITION 1)

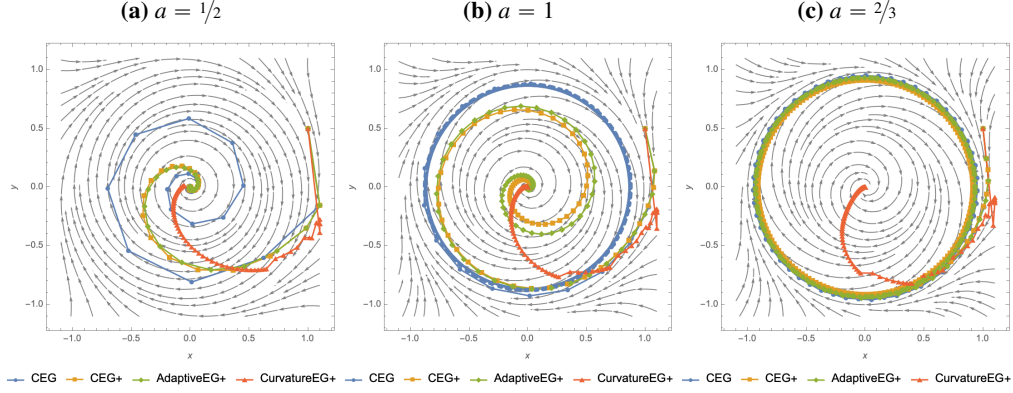
Recall Definition 1 which considers an  $F$  with limit cycles at  $r \in \{c_1, \dots, c_k\}$  where  $c_i \neq 0$  for all  $i \in [k]$ . Such a vectorfield can be constructed by departing from the following dynamics in polar coordinates,

$$\begin{aligned} \frac{\partial r}{\partial t} &= -a \cdot r(t) \prod_{i=1}^k (r(t) + c_i) \cdot (r(t) - c_i) \\ \frac{\partial \theta}{\partial t} &= -b \cdot r(t), \end{aligned} \tag{B.1}$$

with  $a, b \neq 0$ . Transforming this dynamics into cartesian coordinates yields the desired vectorfield,  $F$ , while subsequently integrating with respect to  $x$  and  $y$  yields the two potentials associated with the two players. Note that the roots  $\{-c_i\}_{i=1}^k$  for the polynomial defining  $\dot{r}$  are not strictly necessary for showing existence of limit cycles, but leads to a simpler form for  $Fz$ . We illustrate the construction in Fig. 5.

**Proposition 1.** Let  $F(z) = (\dot{x}, \dot{y})$  be the evolution in cartesian coordinates of the associated vectorfield in polar coordinates defined by Eq. (B.1). Then the only stationary point of  $F$  is at the origin and there exists a limit cycle at  $r = c_i$  for all  $i \in [k]$ .

*Proof.* Recall  $r = \sqrt{x^2 + y^2}$ . It is easy to see from Eq. (B.1) that the only stationary point is at  $r = 0$ . By construction,  $\dot{r}$  is a polynomial with roots  $c_i$  for all  $i \in [k]$ , so any trajectory starting on the circle defined by  $r = c_i$  remains in that set. However,  $\dot{\theta}$  is strictly nonzero. As a consequence  $F(z)$  is nonzero, so  $r = c_i$  must define a limit cycle, which proves the claim.  $\square$



**Figure 6:** *Example 3* for different values of  $a$ . Note that even extragradient may escape the limit cycles even though  $\rho < 0$ . This is not in conflict with the negative results of [Hsieh et al. \(2021\)](#) since the stepsize is not diminishing. However, in the general case extragradient will not converge as shown by the lower bound in [Theorem 3.4](#).

## B.2 PROOF FOR PROPERTIES OF [EXAMPLE 3](#)

The operator  $F$  defined in [Example 3](#) is obtained by constructing the associated dynamics in polar coordinates,

$$\begin{aligned}\frac{\partial r}{\partial t} &= -a \cdot r(t) \cdot (r(t) + 1) \cdot (r(t) - 1) \cdot (r(t) + 1/2) \cdot (r(t) - 1/2) \\ \frac{\partial \theta}{\partial t} &= -r(t).\end{aligned}\tag{B.2}$$

This can easily be verified by a change of variables. From [Proposition 1](#) it then follows, that there must exist a limit cycle at  $\|z\| = 1$  and  $\|z\| = 1/2$ . To verify the conditions on  $\rho$  we compute the closed form solution to  $\rho$  and  $L$  in Mathematica:

- (i) For  $a = 3/2$  we have  $\rho = -\frac{3456}{17113}$  and  $L = \frac{\sqrt{48248185933-521268\sqrt{7816416361}}}{10000}$
- (ii) For  $a = 1$  we have  $\rho = -\frac{576}{4177}$  and  $L = \frac{\sqrt{5374798437-173756\sqrt{871962929}}}{5000}$
- (iii) For  $a = 1/2$  we have  $\rho = -\frac{1152}{16465}$  and  $L = \frac{\sqrt{5449798437-173756\sqrt{890712929}}}{10000}$

It can easily be verified that the stated conditions for  $\rho$  in [Example 3](#) are met for the values above. This completes the proof.

We provide Mathematica code verifying the construction of  $F$  and the closed form solutions to  $L$  and  $\rho$ .

## B.3 PROOF FOR PROPERTIES OF [EXAMPLE 4](#)

**Proposition 2.** *Let  $F$  be the associated operator of  $\phi$  in ([GlobalForsaken](#)) defined as  $Fz = (\nabla_x \phi(x, y), -\nabla_y \phi(x, y))$ . Define the radius as  $r = \|z\|$ . Then,  $Fz$  has a stable critical point at the origin, at least one attracting limit cycle in the region defined by  $\sqrt{3/2} < r < 2$  and at least one repellant limit cycle within  $r \leq \sqrt{3/2}$ .*

*Proof.* We follow a similar argument as in [Hsieh et al. \(2021, D.2\)](#). We can compute the associated operator  $F$ ,

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} \frac{4x^5}{7} - \frac{4x^3}{3} + \frac{2x}{3} + y \\ -x + \frac{4y^5}{7} - \frac{4y^3}{3} + \frac{2y}{3} \end{pmatrix}.\tag{B.3}$$

With a change of variables into polar coordinates  $(r, \theta)$  we get that  $r = \sqrt{x^2 + y^2}$  evolves as,

$$\dot{r} = -\frac{1}{42}r(9r^4 \cos(4\theta) - 14r^2 \cos(4\theta) + 15r^4 - 42r^2 + 28). \quad (\text{B.4})$$

When  $r = \sqrt{3/2}$  this reduces to  $\dot{r} = \frac{3\cos(4\theta)+5}{56\sqrt{6}}$  and we observe that  $\dot{r} > 0$  for any  $\theta$ . Likewise for  $r = 2$ , we have that  $\dot{r} = -\frac{4}{21}(22\cos(4\theta)+25)$  which implies  $\dot{r} < 0$ . Since there is no stationary point in the region  $\mathcal{S} = \{(r, \theta) : \sqrt{3/2} < r < 2\}$  it then follows from the Poincaré-Bendixson theorem (Teschl, 2012, Theorem 7.16) that there must exist at least one attracting limit cycle in  $\mathcal{S}$ . Further, it is easy to see from  $F(z)$  that  $(0, 0)$  is a critical point and that it is stable by inspection of the Jacobian  $JF(z)$ . Since  $\mathcal{S}$  is trapping, we then know from Poincaré–Hopf index theorem, that there must exist a repellent limit cycles in the region defined by  $r < \sqrt{3/2}$ . This completes the proof.  $\square$

**Proposition 3.** *The origin in (GlobalForsaken) is a global Nash equilibrium and satisfies Assumption I(iii) with  $\rho > -1/2L$ .*

*Proof.* To verify that  $(0, 0)$  is indeed a global Nash equilibrium we need to check that the solution cannot be unilaterally improved. In other words, they should coincide with  $(x^*, y^*)$  where

$$\begin{aligned} x^* &= \mathbf{arg\,min}_x \phi(x, 0) \\ y^* &= \mathbf{arg\,max}_y \phi(0, y). \end{aligned} \quad (\text{B.5})$$

We can easily verify this with `Minimize` in Mathematica, since the functions are polynomial for which a closed form solutions to the global optimization problem will be returned.

To find  $\rho$  for  $z^* = (0, 0)$  we solve the global minimization problem,

$$\mathbf{minimize}_z \frac{\langle Fz, z - z^* \rangle}{\|Fz\|}, \quad (\text{B.6})$$

for which a closed form solution can be found with Mathematica, which when numerically evaluated is approximately  $-0.119732$ .

We need to compute  $L$  to ensure  $\rho > -1/L$ . In our case of convex constraints,  $C$ , we have that  $L = \sup_{z \in C} \|JF(z)\|$  where  $\|\cdot\|$  denotes the spectral norm (Rockafellar & Wets, 2009, Thm. 9.2 and 9.7). Under our constraint  $\|z\|_\infty < 4/3$ , this can similarly be computed in closed form, yielding  $L = \frac{\sqrt{2470645}}{567}$ . So  $-\frac{1}{2L} \approx -0.180363$  which satisfy the condition  $\rho > -\frac{1}{2L}$ . This completes the proof.  $\square$