Asymmetric regularization mechanism for GAN training with Variational Inequalities

Spyridon C. Giagtzoglou Mark H.M. Winands Barbara Franci

Abstract— We formulate the training of generative adversarial networks (GANs) as a Nash equilibrium seeking problem. To stabilize the training process and find a Nash equilibrium, we propose an asymmetric regularization mechanism based on the classic Tikhonov step and on a novel zero-centered gradient penalty. Under smoothness and a local identifiability condition induced by a Gauss–Newton Gramian, we obtain explicit Lipschitz and (strong)-monotonicity constants for the regularized operator. These constants ensure last-iterate linear convergence of a single-call Extrapolation-from-the-Past (EFTP) method. Empirical simulations on an academic example show that, even when strong monotonicity cannot be achieved, the asymmetric regularization is enough to converge to an equilibrium and stabilize the trajectory.

I. INTRODUCTION

Generative Adversarial Networks (GANs) are a machine-learning model composed of two neural networks: the discriminator and the generator [1]. The purpose of the model is to train the generator to create realistic data (images, audio) by playing a "game" against the discriminator. In particular, the training of the two neural networks is a two-player zero-sum game that can inherently be modelled as a saddle-point problem. Saddle point problems are well studied in the literature [2], [3] and find applications in consensus problems, resource allocation or simply primal-dual problems [4], [5]. However, GANs highlighted some instabilities, like mode collapse, that can be seen as "best-response" pathologies. Moreover, the training dynamics are often rotational and poorly conditioned and the only available solutions are basically ways to change the game's payoff to make the learning process smoother [6], [7].

We rely instead on variational inequalities (VIs) and monotone operator theory to establish the properties of the operators involved and derive projected algorithms with convergence guarantees. We adopt a VI viewpoint because it exposes the saddle operator connected to the BAN game and its geometry, it allows to handle constraints via projections in a principled way, and directly connects to convergence guarantees for a variety of methods. Since the VI formulation highlights the rotational dynamics of adversarial training, stabilization strategies are needed to make the training reliable.For this reason, we introduce a regularization mechanism applied only on the discriminator side, that is, asymmetrically across the saddle blocks. This discriminator-only block-asymmetric regularization adds curvature where it is most needed while preserving the target equilibrium point. It also provides explicit analytic constants, specifically Lipschitz continuity and strong-monotonicity constants for the regularized operator, thereby enabling convergence guarantees for the proposed iterative method.

Among the available algorithms in the literature, we propose intrapolation-from-the-Past (EFTP) [8] because, although made of two projection steps, it requires only one gradient computation, hence providing a computationally convenient alternative to the more common Forward-Backward (FB) [9] and ExtraGradient (EG) [10] schemes. In particular, our contributions can be summarized as follows.

- We introduce and analyze two explicitly symmetric regularizations based on Tikhonov [11] and a zero-centered input-gradient penalty on the discriminator side. The penalty vanishes at equilibrium, preserving the original saddle point structure while improving conditioning.
- Under smoothness and a local Gauss–Newton identifiability condition, we derive explicit Lipschitz and strong-monotonicity constants for the regularized operator, showing that our regularization can infer stronger monotonicity properties on the original saddle operator.
- We propose EFTP as a training algorithm for GANs and show its convergence to a NE of the corresponding game, even in the regularized case.

We validate our analysis with some numerical simulations showing the effects of our proposed regularization in stabilizing the training process and the convergence properties of the proposed algorithm. Although our work takes GANs as a primary application, our results can be generalized to symmetric regularization, hence retrieving, for instance, the classic Tikhonov [11], [12]. As a consequence, the extension to n-players games is also possible, as long as some distinction between the two sides (generator and discriminator) is kept for the asymmetric case.

A. Notation

Let R be the set of all real number and Rn represents n-dimensional Euclidean space.

Probability: Let (X, X,pD) and (Z,Z,pZ) be probability spaces. X and Z are the domains of data and latent vectors, respectively (e.g., X = Rdx, Z = Rdz), X and Z are their σ-algebras, and pD and pZ are probability measures

---

S. C. Giagtzoglou was supported by the Dutch Research Council (NWO) M 1 grant, project number OCENW.M.22.433. S. C. Giagtzoglou and Mark H.M. Winands are with the Department of Advanced Computing Sciences, Maastricht University, Paul-Henri Spaaklaan 1, 6229 GS Maastricht, The Netherlands (spyridon.giagtzoglou@maastrichtuniversity.nl; m.winands@maastrichtuniversity.nl). B. Franci is with the Department of Mathematical Sciences, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy (barbara.franci@polito.it).

on $(\mathcal{X}, \mathcal{X})$ and $(\mathcal{Z}, \mathcal{Z})$, respectively. We write $x \sim p_D$ and $z \sim p_Z$ for samples drawn from these distributions.

Matrices: Given $A \in \mathbb{R}^{n\times n}$, we write $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ for its smallest and largest eigenvalues, respectively. Let $\mathbb{M}^n \subset \mathbb{R}^{n\times n}$ be the set of symmetric matrices of dimension n. For $A \in \mathbb{M}^n$: $A \succeq 0$ or $A \in \mathbb{M}^n_+$ means that A is positive semidefinite (PSD), i.e., $x^\top A x \geq 0$ for all $x \in \mathbb{R}^n$; $A \succ 0$ or $A \in \mathbb{M}^n_{++}$ means that A is positive definite (PD), i.e., $x^\top A x > 0$ for all $x \neq 0$. For vectors, $\|\cdot\|_2$ denotes the Euclidean norm. For matrices $A \in \mathbb{R}^{m\times n}$, $\|A\|$ denotes the spectral norm induced by $\|\cdot\|_2$, i.e. $\|A\| = \sup_{x\neq 0} \frac{\|Ax\|_2}{\|x\|_2} = \sigma_{\max}(A) = \sqrt{\lambda_{\max}(A^\top A)}$. If $A \in \mathbb{M}^n_+$, $\|A\| = \lambda_{\max}(A)$.

$\|A\| = \max$

Operator Theory: For a nonempty closed convex set S, the Euclidean projector is $\Pi_S(x) := \operatorname{argmin}_{y \in S} \|y - x\|_2$. The normal cone is $N_S(x) := \{v : \langle v, y - x \rangle \leq 0, \forall y \in S\}$ when $x \in S$, and empty otherwise. A function $f : \mathbb{R}^n \to \mathbb{R}$ is μ-strongly convex (concave) if $f - \frac{\mu}{2}\|\cdot\|^2$ $\left(-f - \frac{\mu}{2}\|\cdot\|^2\right)$, $\mu > 0$, is convex. A mapping $F : \mathbb{R}^n \to \mathbb{R}^n$ is: L-Lipschitz continous on S if $\|F(u) - F(v)\| \leq L\|u - v\|$ for all $u, v \in S$ with $L > 0$; monotone if $\langle F(u) - F(v), u - v\rangle \geq 0$ for all $u, v \in S$, μ-strongly monotone if $\langle F(u) - F(v), u - v\rangle \geq \mu\|u - v\|^2$ with $\mu > 0$. Given $P \in \mathbb{M}^n_+$, the residual is $R_P(\omega) := P^{-1}(\omega - \Pi_S(\omega - P F(\omega)))$. For a mapping $F : \mathbb{R}^d \to \mathbb{R}^m$, we indicate with $J_F \in \mathbb{R}^{m\times d}$ the Jacobian. For a twice continuously differentiable function $f : \mathbb{R}^d \to \mathbb{R}$ we indicate with $\nabla^2_{uu} f \in \mathbb{M}$ the Hessian and with $\nabla^2_{xy} f$ the mixed partial derivatives.

II. PROBLEM DESCRIPTION

We model the training of GANs as a two-player zero-sum game between the generator $G_\theta : \mathcal{Z} \to \mathcal{X}$ (player G) and the discriminator $D_\varphi : \mathcal{X} \to \mathbb{R}$ (player D), with parameters $\theta \in \Theta \subseteq \mathbb{R}^{d_\theta}$ and $\varphi \in \Phi \subseteq \mathbb{R}^{d_\varphi}$. The strategy sets Θ and Φ capture architectural/regularization constraints We write $S := \Theta \times \Phi$ and $\omega = (\theta, \varphi)$. The two players min–max optimisation problem reads as

$$\min_{\theta \in \Theta} \max_{\varphi \in \Phi} L(\theta, \varphi) := \mathbb{E}_{x \sim p_D}[\Psi(D_\varphi(x))] + \mathbb{E}_{z \sim p_Z}[\Psi(-D_\varphi(G_\theta(z)))], \tag{1}$$

therefore this problem is a zero-sum game. A popular choice for the objective is $\Psi(t) = -\log(1 + e^{-t})$ but other choices can be considered, as long as the following assumptions (Assumption 3 in particular) are satisfied.

Accordingly, a pair $(\theta^\star, \varphi^\star) \in S$ is a Nash equilibrium iff

$$L(\theta^\star, \varphi) \leq L(\theta^\star, \varphi^\star) \leq L(\theta, \varphi^\star), \quad \forall (\theta, \varphi) \in S. \tag{2}$$

The min-max problem in (1) can be rewritten as a saddle problem and solved by means of a VI. To this aim, let the saddle operator associated [to problem (1)] be

$$F(\omega) = \begin{bmatrix} \nabla_\theta L(\theta, \varphi) \end{bmatrix}. \tag{3}$$

Then, the equilibria of the game in (1) can be characterized, thanks to the operator in (3), as the solutions of the VI(S, F)

$$\text{find } \omega^\star \in S \text{ s.t. } \langle F(\omega^\star), \omega - \omega^\star \rangle \geq 0, \ \forall \omega \in S. \tag{4}$$

We next state the standard conditions ensuring well-posedness of (1) and (4).

Assumption 1 (Local Sets): Θ and Φ are nonempty, closed, and convex.

Note that, under Assumption 1, the Euclidean projector Π_S, used in the later sections, is firmly non-expansive [13, sec 3.1].

Assumption 2 (Smoothness of the base operator): The saddle operator F in (3) is L_0-Lipschitz continous on S.

Assumption 3 (Curvature of the saddle objective): The objective function L in (1) is μ_θ-strongly convex in θ and μ_φ-strongly concave in φ on S, with moduli $\mu_\theta, \mu_\varphi \geq 0$

Remark 1: Note that in Assumption 3 the parameters μ_θ and μ_φ may be zero. This means that the objective function might be merely convex/concave (not strongly convex/strongly concave). Consequently, the symmetric part of the Jacobian J_F is only positive semidefinite and the saddle operator F is, at best, monotone on S.

A. Penalty Design

Without regularization, the GAN saddle field is typically rotational and ill-conditioned, therefore projected first-order methods tend to oscillate or require impractically small steps sizes [14]–[17]. To address these problems, we propose a regularization that injects curvature to suppress rotational drift and improve conditioning. The resulting field comes with explicit Lipschitz and (strong) monotonicity constants, which guide stepsize choices and enable convergence guarantees. Combining the base operator in (3) with a penalty produces

$$F_\gamma(\omega) = \begin{bmatrix} \nabla_\theta L(\theta, \varphi) - \nabla_\theta B(\theta, \varphi) \\ -\nabla_\varphi L(\theta, \varphi) + \nabla_\varphi B(\theta, \varphi) \end{bmatrix}. \tag{5}$$

We have a general, symmentric notation in (5), but later we propose an asymmetric one with $\nabla_\theta B(\theta, \varphi) = 0$. Therefore, we regularize only the discriminator side of the game and analyze the induced operator F_γ.

Remark 2: Solving the regularized problem is not equiv- alent to solving the original one in general. In particular, solutions of VI(S, F_γ) need not coincide with those of VI(S,F). However, if B vanishes at a saddle point of the unregularized operator, then $F_\gamma(\omega^\star) = F(\omega^\star)$. Moreover, as $\gamma \to 0$, bounded solution sequences of VI(S, F_γ) have cluster points that solve VI(S,F).

To see this, let $\{\gamma_k\}_{k>0} \subseteq \mathbb{R}$ be a sequence of regularization parameters and let $\{\omega_k\}_{k>0}$ be the corresponding sequence of solutions of VI(S, F_{γ_k}). Assume $\{\omega_k\}$ is bounded with $\omega_k \to \bar{\omega}$ along a subsequence. For any $y \in S$ we have $\langle F_{\gamma_k}(\omega_k), y - \omega_k \rangle \geq 0$ and by uniform convergence $F_{\gamma_k} \to F$ on a compact set containing $\{\omega_k\}$ [18, Thm. 4.3]. Therefore, passing to the limit results in $\langle F(\bar{\omega}), y - \bar{\omega} \rangle \geq 0$ for all $y \in S$, i.e., $\bar{\omega}$ is a solution of VI(S,F) [3, Cor. 5.1.8].

We consider two forms of regularization. The terms are added in the cost and later we differentiate to obtain the gradients of the regularization terms in (5).

a) Tikhonov regularization (Tik): Inspired by [11], let

$$B_{\mathrm{Tik}}(\theta,\varphi)=\frac{\gamma}{2}\|\varphi\|^2 \tag{6}$$

Compared to the standard Tikhonov regularization, in our case the term acts only on the discriminator block. Accordingly, $\nabla_\varphi B_{\mathrm{Tik}}=\gamma\varphi$, $\nabla^2_{\varphi\varphi}B_{\mathrm{Tik}}=\gamma I$, $\nabla_\theta B_{\mathrm{Tik}}\equiv 0$. This penalty introduces curvature in the φ-subspace while leaving the generator block unaffected.

Moreover, if we also regularize the generator with the same quadratic term, i.e.,

$$B_{\mathrm{fullTik}}(\theta,\varphi)=\frac{\gamma}{2}\left(\|\theta\|^2+\|\varphi\|^2\right), \tag{7}$$

we recover the classic (symmetric) Tikhonov regularization and the penalty adds identical curvature to both the generator and discriminator subspaces.

b) Symmetric (zero-centered) Gradient Penalty (SGP): As an alternative to the regularization introduced above, let

$$\begin{aligned}B_{\mathrm{SGP}}(\theta,\varphi)=\ &\frac{\gamma}{2}\mathbb{E}_{x\sim p_D}\left[\|\nabla_x D_\varphi(x)\|^2\right]\\&+\frac{\gamma}{2}\mathbb{E}_{z\sim p_Z}\left[\|\nabla_x D_\varphi(G_\theta(z))\|^2\right].\end{aligned} \tag{8}$$

This regularization penalizes the input gradient of φ toward zero over both real (x) and generated (z) supports. Despite its name [16], it is added exclusively to the discriminator payoff. By letting $g(x;\varphi):=\nabla_x D_\varphi(x)$ and $H_\varphi(x):=\nabla^2_{x,\varphi}D_\varphi(x)$, and by direct differentiation,

$$\begin{aligned}\nabla_\varphi B_{\mathrm{SGP}}=\gamma\Big(&\mathbb{E}_{x\sim p_D}\left[H_\varphi(x)^\top g(x;\varphi)\right]\\&+\mathbb{E}_{z\sim p_Z}\left[H_\varphi(G_\theta(z))^\top g(G_\theta(z);\varphi)\right]\Big).\end{aligned} \tag{9}$$

Using the chain rule through $x'=G_\theta(z)$ gives $\nabla_\theta B_{\mathrm{SGP}}=\gamma\mathbb{E}\left[J_\theta G_\theta(z)^\top H_{xx}(x')^\top g(x';\varphi)\right]$, $H_{xx}(x):=\nabla^2_x D_\varphi(x)$ and $J_\theta G_\theta(z)$ is the generator Jacobian. Note that, although the regularization is added only on the discriminator cost, it affects also the generator blocks.

Remark 3: The gradients in (8) are with respect to the input x, and not with respect to the parameter φ. This choice targets the function $D_\varphi(\cdot)$ in the data space: it penalizes sharpness of D_φ on the real and generated supports and is zero whenever D_φ is locally flat there. Consequently, the penalty vanishes at the equilibrium and preserves the target saddle point [16, Rem. D.8]. In particular, since $g(\cdot,\varphi^\star)\equiv 0$ at the equilibrium, (8) vanishes at the saddle point, thereby preserving the equilibrium in (2). This motivates the term ”zero-centered”, since both its value and its first derivative vanish at $(\theta^\star,\varphi^\star)$. Compared with a parameter penalty (e.g., Tikhonov in (6)), the input-gradient penalty acts in data space, it is invariant to simple reparameterizations, and, since it injects curvature away from the saddle point, it does not bias the equilibrium, i.e., γ need not be vanishing.

B. Regularization analysis

To understand how the penalties modify the geometry of the saddle field, we examine the curvature they induce on the discriminator block. Since the local behavior of the saddle operator is governed by its Jacobian, to certify explicit Lipschitz continuity and strong–monotonicity constants for the regularized mapping, we inspect the second-order structure of the proposed penalties.

For the curvature analysis on the discriminator block, let the Gauss–Newton (GN) surrogate [19, Sec. 10.3] of the Hessian be $\nabla^2_{\varphi\varphi}B_{\mathrm{SGP}}\approx\gamma J_G(\omega)$, where

$$J_G(\omega):=\mathbb{E}\left[H_\varphi(x)^\top H_\varphi(x)+H_\varphi(x')^\top H_\varphi(x')\right] \tag{10}$$

with x being the real data inputs and x′ being the generated samples inputs.

Assumption 4 (Gauss–Newton identifiability): There exist a neighborhood $\mathcal{R}\subseteq\mathcal{S}$ and a constant $\lambda_0>0$ such that $J_G(\omega)\succeq\lambda_0 I$, for all $\omega\in\mathcal{R}$.

Remark 4: The Jacobian JF_γ of F_γ in (5) includes off-diagonal mixed blocks stemming from $\nabla^2_{\theta\varphi}(\cdot)$ terms. These drive rotational dynamics and can affect the desired strong-monotonicity needed for our iterative scheme. Assumption 4 enforces discriminator-side curvature guaranteeing that the φ-block of the symmetric part of the Jacobian has a uniform positive bound (see Lemma 1) that offsets the rotational coupling between generator and discriminator.

Assumption 5 (Continuity): The discriminator D_φ is twice continuously differentiable in both x and φ and on the relevant supports of $x\sim p_D$ and $x'=G_\theta(z)$, $z\sim p_Z$.

Let ξ be a generic input to indicate either a real sample $x\sim p_D$ or a generated sample $x'=G_\theta(z)$ with $z\sim p_Z$.

Assumption 6 (Bounded mixed Hessian): There exists a constant $C_H>0$ such that $\sup_{\xi\in\operatorname{supp}(p_D)\cup G_\theta(\operatorname{supp}(p_Z))}\mathbb{E}\left[\|H_\varphi(\xi)\|^2\right]\leq C_H^2$, where $H_\varphi(\xi):=\nabla^2_{x,\varphi}D_\varphi(\xi)$.

Lemma 1 (PSD and uniform bound for the Gramian): Let Assumptions 5–6 hold. Then $J_G(\omega)\succeq 0$ and

$$\|J_G(\omega)\|\leq 2C_H^2=:\kappa. \tag{11}$$

Proof: Using the definition of $J_G(\omega)$ in (10), to show that it is PSD, let $v\in\mathbb{R}^{d_\varphi}$, then $v^\top J_G(\omega)v=\mathbb{E}\left[\|H_\varphi(x)v\|^2+\|H_\varphi(x')v\|^2\right]\geq 0$, hence $J_G(\omega)\succeq 0$.

The bound in (11) instead follows by using linearity of expectation and the triangle inequality:

$$\begin{aligned}\|J_G(\omega)\|&=\left\|\mathbb{E}[H_\varphi(x)^\top H_\varphi(x)]+\mathbb{E}[H_\varphi(x')^\top H_\varphi(x')]\right\|\\&\leq\left\|\mathbb{E}[H_\varphi(x)^\top H_\varphi(x)]\right\|+\left\|\mathbb{E}[H_\varphi(x')^\top H_\varphi(x')]\right\|\\&\leq\mathbb{E}\left[\|H_\varphi(x)^\top H_\varphi(x)\|\right]+\mathbb{E}\left[\|H_\varphi(x')^\top H_\varphi(x')\|\right].\end{aligned}$$

Since the spectral norm is convex, by Jensen inequality, $\|J_G(\omega)\|\leq\mathbb{E}[\|H_\varphi(x)\|^2]+\mathbb{E}[\|H_\varphi(x')\|^2]$. Finally by Assumption 6 $\|J_G(\omega)\|\leq C_H^2+C_H^2=2C_H^2=:\kappa$, which proves the claim. ■

We can now show Lipschitz continuity and strong monotonicity of the reguralized operator in (5).

Proposition 1 (Lipschitz constant of F_γ): Under Assumptions 1, 2, 5 and 6, F_γ in (5) is L-Lipschitz continous on $\mathcal{S}$ with

$$L=L_0+\gamma\kappa_{\mathrm{tot}},\qquad \kappa_{\mathrm{tot}}\in[\kappa,\kappa+\tilde{\kappa}_\theta]. \tag{12}$$

where κ is as in (11), and $\tilde{\kappa}_\theta\geq 0$.

Algorithm 1 Forward–Backward

```
Require: initial $\omega^0 \in \mathcal{R}$, stepsize/preconditioner $P \succ 0$
for $k = 0, 1, 2, \ldots$ do
    $\omega^{k+1} \leftarrow \Pi_{\mathcal{R}}(\omega^k - P F_\gamma(\omega^k))$
end for
```

Algorithm 2 ExtraGradient

```
Require: initial $\omega^0 \in \mathcal{R}$, stepsize/preconditioner $P \succ 0$
for $k = 0, 1, 2, \ldots$ do
    $\tilde{\omega}^k \leftarrow \Pi_{\mathcal{R}}(\omega^k - P F_\gamma(\omega^k))$    ▷ look-ahead
    $\omega^{k+1} \leftarrow \Pi_{\mathcal{R}}(\omega^k - P F_\gamma(\tilde{\omega}^k))$
end for
```

Proof: By Assumption 2, the operator F is L_0-Lipschitz on $\mathcal{S}$. Therefore, it suffices to bound the Lipschitz constant of the regularization contribution $\nabla B(\omega) = (-\nabla_\theta B(\omega), \nabla_\varphi B(\omega))^\top$.

(i) Discriminator-side contribution. For B_{Tik} in (6), we have $\nabla_\varphi B_{\mathrm{Tik}}(\omega) = \gamma\varphi$, whose Jacobian w.r.t. φ is γI; thus

$$\|\nabla_\varphi B_{\mathrm{Tik}}(\omega) - \nabla_\varphi B_{\mathrm{Tik}}(\omega')\| \le \gamma \|\omega - \omega'\|.$$

For B_{SGP} in (9), by the Gauss–Newton surrogate in (10) and Lemma 1, the φ-Jacobian of $\nabla_\varphi B_{\mathrm{SGP}}$ satisfies $\|\nabla^2_{\varphi\varphi} B_{\mathrm{SGP}}(\omega)\| \approx \gamma \|J_G(\omega)\| \le \gamma\kappa$, with κ as in (11). By the mean-value theorem for vector fields,

$$\|\nabla_\varphi B_{\mathrm{SGP}}(\omega) - \nabla_\varphi B_{\mathrm{SGP}}(\omega')\| \le \gamma\kappa \|\omega - \omega'\|.$$

Thus, in both B_{Tik} and B_{SGP} cases, the discriminator-side contribution is $\gamma\kappa$-Lipschitz (with $\kappa = 1$ for B_{Tik} and κ from (11) for B_{SGP}).

(ii) Generator-side contribution. For B_{Tik} in (6) we have $\nabla_\theta B_{\mathrm{SGP}}(\omega) = \gamma \mathbb{E}[J_{G_\theta}(z)^\top H_{xx}(x')^\top g(x';\varphi)]$. Under Assumptions 5–6 and bounded supports/parameters, $J_\theta G_\theta$, H_{xx} and g are uniformly bounded and Lipschitz continuous (see Remark 6). Therefore, $\nabla_\theta B_{\mathrm{SGP}}$ is Lipschitz continuous with some finite constant, i.e., $\tilde{\kappa}_\theta < \infty$.

Combining (i) and (ii), the regularization term is $\gamma(\kappa + \tilde{\kappa}_\theta)$-Lipschitz continuous. We have $\|F_\gamma(\omega) - F_\gamma(\omega')\| \le (L_0 + \gamma\kappa + \gamma\tilde{\kappa}_\theta)\|\omega - \omega'\|$, $\kappa_{\mathrm{tot}} \in [\kappa, \kappa + \tilde{\kappa}_\theta]$. ■

Remark 5: The quantity $\tilde{\kappa}_\theta$ in (12) is nonnegative by definition and it captures the θ-dependence of B. In the case of the Tikhonov regularization in (6), $\tilde{\kappa}_\theta = 0$. For the SGP regularization in (8), $\tilde{\kappa}_\theta > 0$ whenever the generator Jacobian $J_\theta G_\theta$ and the mixed Hessian H_φ are uniformly bounded on the supports of (x, z). This happens when x and z range over bounded sets (e.g., $x \in [0, 1]^{d_x}$, $z \in [-c, c]^{d_z}$), $G_\theta(Z)$ is bounded, the parameter sets Θ, Φ are bounded, and G_θ and D_φ are twice continuously differentiable on their domains.

Proposition 2 (Strong monotonicity of F_γ): Under Assumptions 3 and 4, the Jacobian J_{F_γ} of the regularized map F_γ in (5) satisfies

$$\frac{J_{F_\gamma}(\omega) + J_{F_\gamma}(\omega)^\top}{2} \succeq \mathrm{diag}(\mu_\theta I, (\mu_\varphi + \gamma\lambda_0) I) \tag{13}$$

on $\mathcal{R} \subseteq \mathcal{S}$. Therefore, F_γ is μ-strongly monotone on $\mathcal{R}$ with

$$\mu \ge \min\{\mu_\theta, \mu_\varphi + \gamma\lambda_0\}. \tag{14}$$

Proof: Let F_γ be as in (5). Then, its Jacobian is

$$J_{F_\gamma}(\omega) = \begin{bmatrix} A & -B_{\mathrm{mix}} \\ B_{\mathrm{mix}}^\top & C \end{bmatrix}(\omega), \tag{15}$$

where $A = \nabla^2_{\theta\theta} L - \nabla^2_{\theta\theta} B$, $B_{\mathrm{mix}} = \nabla^2_{\theta\varphi} L - \nabla^2_{\theta\varphi} B$ and $C = -\nabla^2_{\varphi\varphi} L + \nabla^2_{\varphi\varphi} B$. By Schwarz symmetry, the symmetric part is block diagonal: $\frac{J_{F_\gamma}(\omega) + J_{F_\gamma}(\omega)^\top}{2} = \mathrm{diag}(\nabla^2_{\theta\theta} L - \nabla^2_{\theta\theta} B, -\nabla^2_{\varphi\varphi} L + \nabla^2_{\varphi\varphi} B)$. By Assumption 3, $\nabla^2_{\theta\theta} L \succeq \mu_\theta I$ and $-\nabla^2_{\varphi\varphi} L \succeq \mu_\varphi I$. We consider the two regularization separately.

a) For Tikhonov regularization, $\nabla^2_{\varphi\varphi} B_{\mathrm{Tik}} = \gamma I$ and $\nabla^2_{\theta\theta} B_{\mathrm{Tik}} \equiv 0$, hence on any region in $\mathcal{R}$, $\frac{J_{F_\gamma}(\omega) + J_{F_\gamma}(\omega)^\top}{2} \succeq \mathrm{diag}(\mu_\theta I, (\mu_\varphi + \gamma) I)$.

b) For B_{SGP}, using the Gauss–Newton surrogate in (10) and Assumption 4, $\nabla^2_{\varphi\varphi} B_{\mathrm{SGP}} \succeq \gamma J_G(\omega) \succeq \gamma\lambda_0 I$ on $\mathcal{R}$, hence $\frac{J_{F_\gamma} + J_{F_\gamma}^\top}{2} \succeq \mathrm{diag}(\mu_\theta I, (\mu_\varphi + \gamma\lambda_0) I)$, on $\mathcal{R} \subset \mathcal{S}$ which is exactly (13).

Therefore F_γ is μ-strongly monotone on $\mathcal{R}$ with $\mu \ge \min\{\mu_\theta, \mu_\varphi + \gamma\lambda_0\}$ [3, Proposition 2.3.2.(c)]. □ ■

Remark 6: If $\mu_\theta = 0$ in Assumption 3, Proposition 2 implies only $\mu \ge 0$ i.e., the operator F_γ might be monotone. Adding generator-side curvature, e.g., a Tikhonov term $\frac{\tau}{2}\|\theta\|^2$ (in (7), the right-hand side) of Equation (13) becomes $\mathrm{diag}((\mu_\theta + \tau) I, (\mu_\varphi + \gamma\lambda_0) I)$ and gives $\mu \ge$

Algorithm 3 EFTP

```
Require: initial $\omega_0 \in \mathcal{R}$, stepsize/preconditioner $P \succ 0$
$y_0 \leftarrow \omega_0$;  $\hat{F}_0 \leftarrow F_\gamma(y_0)$    ▷ warm-start: one oracle call
for $k = 0, 1, 2, \ldots$ do
    $\omega_{k+1} \leftarrow \Pi_{\mathcal{R}}(\omega_k - P\hat{F}_k)$
    $y_{k+1} \leftarrow \Pi_{\mathcal{R}}(\omega_{k+1} - P\hat{F}_k)$
    $\hat{F}_{k+1} \leftarrow F_\gamma(y_{k+1})$    ▷ one oracle call per loop
end for
```

III. ALGORITHMS AND CONVERGENCE

We consider three projected first-order methods to find a Nash equilibrium of our saddle problem: Forward–Backward (FB), ExtraGradient (EG), and Extrapolation from the Past (EFTP). The details of these iterative schemes are detailed in Algorithms 1, 2 and 3, respectively. Note that we use the regularized operator in (5) for these iterative schemes.

FB is the classical projected gradient step, taking a single projection along $-F_\gamma(\omega_k)$ [9]. EG stabilizes the dynamics by evaluating at a look-ahead point $\tilde{\omega}_k$ and updating with $F_\gamma(\tilde{\omega}_k)$, which helps cancel rotational drift on monotone saddle fields [10]. EFTP achieves similar stabilization while reusing a stored evaluation: after a warm start with one oracle call, it needs only a single gradient call per iteration, reducing computational cost [8].

A. Convergence analysis

Let L be the Lipschitz constant of F_γ on S and μ its (strong)-monotonicity constant on the region $R \subseteq S$ as in Lemma 1 and Proposition 1. We recall that, according to Remark 6, μ might be 0.

Assumption 7 (Stepsizes): Let $R \subseteq S$ and $P \in M_d^{++}$, the step size sequences are such that for FB, $\|P\|_2 < \frac{2\mu\lambda_{\min}(P)}{L^2}$; for EG, $\|P\| \leq \frac{1}{L}$; for EFTP, $\|P\| \leq \frac{1}{4L}$.

Remark 7 (Scalar case): If $P = \eta I_d$, then the bounds reduce to $0 < \eta < \frac{2\mu}{L^2}$ (FB), $0 < \eta \leq \frac{1}{L}$ (EG), and $0 < \eta \leq \frac{1}{2L}$ (EFTP). Using the explicit bound $L = L_0 + \gamma k_{tot}$ from Proposition 1 produces a computable step-sizes.

We are now ready to state the convergence result for our iterative schemes.

Theorem 1 (FB vs. EG): Let Assumptions 1, 2 and 7 hold.

1) If F_γ is strongly monotone ($\mu > 0$) on R, then Algorithm 1 converges to a solution $\omega_\star$ of (4). The convergence is Q-linear.

2) If F_γ is monotone ($\mu = 0$) on S, then Algorithm 2 converges to a solution $\omega_\star$ of (4). If, in addition, F_γ is (strongly) monotone on R, the convergence is Q-linear.

Proof: 1) Consider the fixed-point map $T(\omega) = \Pi_S(\omega - P^{-1}F_\gamma(\omega))$. By [3, Thm. 12.1.2], and since F_γ is a contraction in the P^{-1}-induced norm. Therefore, $\omega^{k+1} = T(\omega^k)$ converges to the unique solution of VI(S, F_γ), and the Banach contraction theorem provides Q-linear convergence.

2) If F_γ is monotone on S and L-Lipschitz continuous, Algorithm 2 generates a Fejér-monotone sequence with respect to the solution set SOL(S,F_γ) and hence converges to a solution of VI(S,F_γ) [3, Thm 12.1.1], [13, §5.8]. If, in addition, F_γ is (locally) strongly monotone on R, then EG converges Q-linearly for sufficiently small η [3, Thm. 12.6.4]. ■

Remark 8 (Region of validity R): Throughout Sections II–III, $R \subseteq S$ denotes a convex neighborhood of a solution $\omega_\star$ on which our regularity assumptions hold and where the constants L (Lipschitz continuity) and μ (strong monotonicity) are valid. All convergence statements for FB, EG, and EFTP are made on R. If $R = S$, the results are global; otherwise they are local and require an initial point $\omega^0 \in R$.

We state the convergence of Algorithm 3 as a separate result since, as far as we know, it is a new scheme for games.

Theorem 2 (EFTP): Let Assumptions 1-7 hold.

1) If F_γ is monotone ($\mu = 0$) on R, then Algorithm 3 converges to a solution $\omega_\star$ of (4).

2) If F_γ is strongly monotone ($\mu > 0$) on R, then Algorithm 3 is a contraction on R and $\|\omega^{k+1} - \omega_\star\| \leq q\|\omega^k - \omega_\star\|$ with $q \in (0,1)$. Consequently, Algorithm 3 converges Q-linearly to a solution $\omega_\star$ of (4). Moreover, the number of iterations to reach $\|\omega^{k+1} - \omega^*\| \leq \varepsilon$ scales as $O(\log(1/\varepsilon))$.

Proof: Let $\omega_\star \in$ SOL(S,F_γ) be the (locally) unique solution in R. Then, for 1), convergence follows mutatis mutandis from [8, Theorem 1]. In particular, our conservative step size $\|P\| \leq \frac{1}{4L}$ ensures the bound $\|P\| \leq \frac{1}{3L}$ used in [8]. If, instead, we instead have strong monotonicity, analogously to [14, Theorem 1] which shows $\|\omega^{k+1} - \omega^*\| \leq \left(1 - \frac{\mu}{4L}\right)\|\omega^k - \omega^*\|$ leading to $q \in (0,1)$ by Assumption 7. Accordingly, by [3, Thm. 2.1.21(c)], the number of iterations to reach $\|\omega^{k+1} - \omega_\star\| \leq \varepsilon$ scales as $O(\log(1/\varepsilon))$. ■

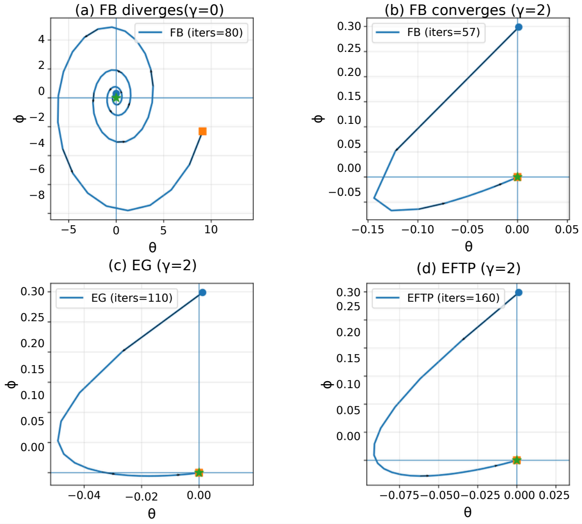

Fig. 1: Trajectories with discriminator-only curvature: (a) FB diverges for $\gamma = 0$; (b) FB converges for $\gamma = 2$; (c) EG ($\gamma = 2$); (d) EFTP ($\gamma = 2$). Blue dots indicate the starting point and squares the end points; the star is the solution. The legends report iteration till convergence.

IV. NUMERICAL SIMULATIONS

To illustrate our asymmetric design mechanism, we study a bilinear toy example. We assume the discriminator is linear in the input, and specialize the GAN loss to $L(\theta,\varphi) = \mathbb{E}_{x\sim p_D}[\varphi^\top x] - \mathbb{E}_{z\sim p_z}[\varphi^\top G_\theta(z)] = \varphi^\top(\mathbb{E}[x] - \mathbb{E}[G_\theta(z)])$. In particular, we set $\mathbb{E}[x] = 0$ and $\mathbb{E}[G_\theta(z)] = -a\theta$ with $a > 0$. Since the discriminator is linear in its input, $g(x;\varphi) = \nabla_x D_\varphi(x) = \varphi$ is constant in x. In this case, the zero-centered input-gradient penalty simplifies to $B_{SGP}(\theta,\varphi) = \frac{\gamma}{2}\mathbb{E}\left[\|g(x;\varphi)\|^2 + \|g(G_\theta(z);\varphi)\|^2\right] = \frac{\gamma}{2}\|\varphi\|^2$, which exactly matches the Tikhonov regularization in (6). With $\omega = (\theta, \varphi)^\top \in \mathbb{R}^2$, the associated saddle operator is

$$F_\gamma(\omega) = \begin{bmatrix} a\varphi \\ -a\theta + \gamma\varphi \end{bmatrix}, \quad JF_\gamma = \begin{bmatrix} 0 & a \end{bmatrix}. \tag{16}$$

Note that F_γ is monotone for $\gamma \geq 0$, but not strongly monotone [3, Sec. 2.3]. We implemented (16) with $a > 0$ and varied $\gamma \geq 0$. For stepsize we used Assumption 7 with L as in Proposition 1. Under this premises, FB exhibits oscillatory behavior and can stall unless γ is large enough and the stepsize is tuned conservatively, i.e., according to Assumption 7. This can be observed in Figure 1(a), with $\gamma = 0$ and Figure 1(b) with $\gamma = 2$. For EG and EFTP instead, the second gradient step cancels the rotational drift

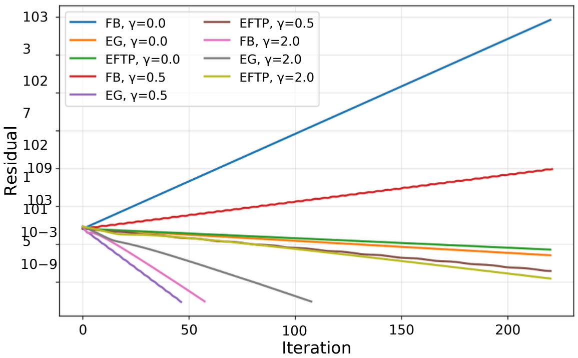

Fig. 2: Residual vs. iterations on the bilinear toy for $\gamma = \{0, 0.5, 2\}$.

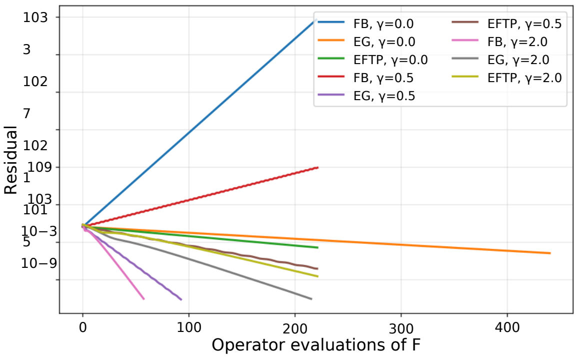

Fig. 3: Residual as a function of gradient evaluations.

and results in convergence (Figure 1(c)-(d)). Moreover, EFTP with one oracle call per half step matches the stabilizing effect of EG on monotone fields while using stored gradients hence while being computationally lighter.

The residual plots in Figures 2–3 show how γ stabilizes the adversarial dynamics: FB diverges for $\gamma = 0$, slows for moderate curvature, and contracts linearly for $\gamma = 2$. For all γ, the field remains monotone but not strongly monotone, so only look-ahead methods (EG, EFTP) achieve convergence. In particular, EG and EFTP exhibit stable, near- linear decrease that improves with larger γ, whereas FB is unstable for small curvature and stabilizes as γ increases. Concerning the computational burden (Figure 3), EG uses two gradient steps per iteration and EFTP only one per half step: both eliminate the rotational drift but EFTP achieves similar performances with fewer gradient calls.

V. CONCLUSION

We studied GAN training through a VI formulation and introduced a block-asymmetric, zero-centered input-gradient penalty applied to the discriminator only. This design preserves the target saddle point while certifiably improving conditioning. Under smoothness and a local Gauss–Newton (GN) identifiability condition, we derived explicit Lipschitz continuity and strong-monotonicity constants for the regularized operator and established linear convergence of EFTP scheme. Extending the analysis to stochastic oracles with variance reduction or adaptive block preconditioners tasks would further test the asymmetric design.

REFERENCES

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial networks,” Communications of the ACM, vol. 63, no. 11, pp. 139–144, 2020.

[2] W. Li, X. Zeng, and L. Pavel, “Primal-dual accelerated mirror-descent method for constrained bilinear saddle-point problems,” IEEE Transactions on Automatic Control, 2025.

[3] F.FacchineiandJ.-S.Pang,Finite-dimensionalvariationalinequalities and complementarity problems. Springer, 2003.

[4] D.GoldsztajnandF.Paganini,“Proximalregularizationforthesaddle point gradient dynamics,” IEEE Transactions on Automatic Control, vol. 66, no. 9, pp. 4385–4392, 2020.

[5] Y.Huang,Z.Meng,J.Sun,andW.Ren,“Aunifieddistributedmethod for constrained networked optimization via saddle-point dynamics,” IEEE Transactions on Automatic Control, vol. 69, no. 3, pp. 1819–1826, 2024.

[6] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, “Improved training of wasserstein gans,” Advances in neural information processing systems, vol. 30, 2017.

[7] A.Jolicoeur-Martineau,“Therelativisticdiscriminator:Akeyelement missing from standard gan,” arXiv preprint arXiv:1807.00734, 2018.

[8] L.D.Popov,“Amodificationofthearrow-hurwiczmethodforsearch of saddle points,” Mathematical notes of the Academy of Sciences of the USSR, vol. 28, pp. 845–848, 1980.

[9] B. Franci and S. Grammatico, “Forward–backward algorithms for stochastic nash equilibrium seeking in restricted strongly and strictly monotone games,” IEEE Transactions on Automatic Control, vol. 67, no. 9, pp. 4881–4893, 2022.

[10] G. M. Korpelevich, “The extragradient method for finding saddle points and other problems,” Matecon, vol. 12, pp. 747–756, 1976.

[11] A.N.Tikhonov,A.V.Goncharsky,V.V.Stepanov,andA.G.Yagola, Regularizationmethods. Dordrecht:SpringerNetherlands,1995,pp. 7–63.

[12] F. Fabiani and B. Franci, “A stochastic generalized nash equilibrium model for platforms competition in the ride-hail market,” in 2022 IEEE 61st Conference on Decision and Control (CDC). IEEE, 2022, pp. 4455–4460.

[13] E. K. Ryu and S. Boyd, “Primer on monotone operator methods,” Applied and Computational Mathematics, vol. 15, no. 1, pp. 3–43, 2016.

[14] G. Gidel, H. Berard, G. Vignoud, P. Vincent, and S. Lacoste-Julien, “A variational inequality perspective on generative adversarial networks,” arXiv preprint arXiv:1802.10551, 2018.

[15] L. Mescheder, S. Nowozin, and A. Geiger, “The numerics of gans,” Advances in neural information processing systems, vol. 30, 2017.

[16] L. Mescheder, A. Geiger, and S. Nowozin, “Which training methods for gans do actually converge?” in International conference on machinelearning. PMLR,2018,pp.3481–3490.

[17] V. Nagarajan and J. Z. Kolter, “Gradient descent gan optimization is locally stable,” Advances in neural information processing systems, vol. 30, 2017.

[18] A.Shapiro,“Sensitivityanalysisofparameterizedvariationalinequalities,” Mathematics of Operations Research, vol. 30, no. 1, pp. 109–126, 2005.

[19] J. Nocedal and S. J. Wright, Numerical Optimization, 2nd ed. Springer, 2006.