PL CONDITIONS DO NOT GUARANTEE CONVERGENCE OF GRADIENT DESCENT-ASCENT DYNAMICS

JEAN-CHRISTOPHE MOURRAT

Abstract. We give an example of a function satisfying a two-sided Polyak-Łojasiewicz condition but for which a gradient descent-ascent flow line fails to converge to the saddle point, circling around it instead.

1. Introduction

A differentiable function $f : \mathbb{R}^d \to \mathbb{R}$ that is bounded from below is said to satisfy a Polyak-Łojasiewicz (PL) condition [8, 14] if there exists $C < +\infty$ such that for all $z \in \mathbb{R}^d$,

$$f(z) - \inf f \leqslant C|\nabla f(z)|^2. \tag{1.1}$$

Under this condition, all gradient descent trajectories converge to a minimizer of f at linear speed [5]. A natural question is whether analogous conditions can guarantee convergence for saddle-point problems. Min-max optimization problems of the form

$$\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} f(x, y)$$

arise in a variety of contexts, including robust optimization [1] and generative adversarial networks [4]. For definiteness, we set $\mathcal{X}$ and $\mathcal{Y}$ to be Euclidean spaces. A natural procedure for finding a saddle point is to iteratively take small steps in the direction of $(-\nabla_x f, \nabla_y f)$. In the regime of infinitesimally small steps, this amounts to studying the gradient descent-ascent (GDA) flow $(z(t))_{t \geqslant 0}$ solving

$$\partial_t z(t) = \begin{pmatrix} -\nabla_x f \\ \nabla_y f \end{pmatrix} (z(t)). \tag{1.2}$$

If the functions $(f(\cdot, y))_{y \in \mathcal{Y}}$ are uniformly strongly convex, and the functions $(f(x, \cdot))_{x \in \mathcal{X}}$ are uniformly strongly concave, then one can show that the GDA flow converges to a saddle point of f [2]. If f is only convex-concave (without the "strongly" qualifier), then there are counter-examples to convergence of the GDA flow, such as with the function $f : (x, y) \mapsto xy$ for which the flow circles around the origin. Yet under the sole convexity-concavity condition, other first-order algorithms such as the extragradient method do succeed in finding a saddle point for f (assuming that one exists) [6].

There has been significant effort to weaken the strong convexity-concavity assumption, for instance by imposing strong convexity in only one variable,

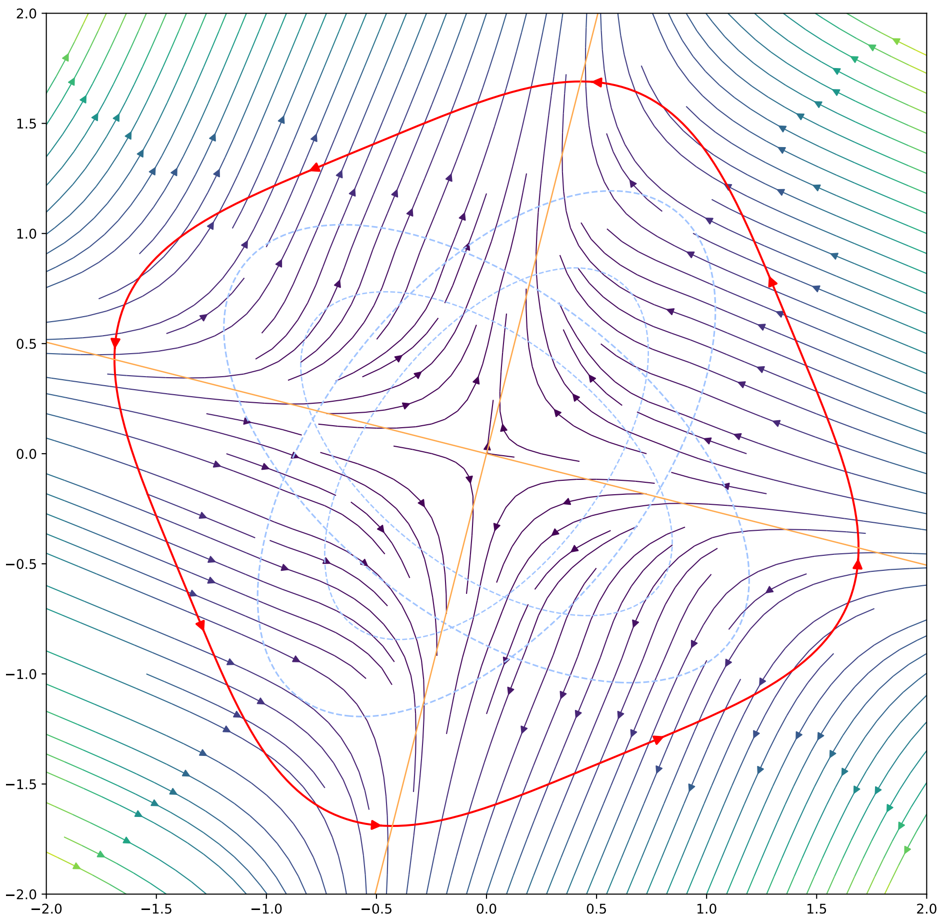

FIGURE 1. The flow lines with a color scale from dark blue to yellow are the level lines of the function f we build for Theorem 1.1 (the color scale indicates the magnitude of $\mathbf{v}$). The value of f is not shown and is prescribed along the two orange lines according to (3.2). These two lines are also the set of points at which the level line of f is horizontal or vertical (i.e., where $\partial_x f = 0$ or $\partial_y f = 0$). The red trajectory is a gradient descent-ascent flow line on f. The ellipses are the level lines of the quadratic forms inside the two occurrences of the function φ in (3.1), for the values $1/2$ and 1. In particular, the level lines of f are tangent to the vector field given in (3.4) on the region contained by every ellipse, and to the vector field given in (3.7) on the region that is outside of every ellipse.

or by replacing strong convexity-concavity with PL conditions [3, 7, 9, 12, 18]. We focus on the latter direction. We say that f satisfies a *two-sided PL condition* if the functions $(f(\cdot, y))_{y\in\mathcal{Y}}$ and $(-f(x,\cdot))_{x\in\mathcal{X}}$ satisfy a PL condition with a uniform constant. Under this condition, modified versions of the GDA flow were shown to converge to a saddle point in [3, 18]. The modifications crucially impose a sufficiently large separation of timescales between the evolutions of the variables x and y; see also [7, 9, 12] for related two-timescale approaches. Another positive result under the two-sided PL condition is that the GDA flow (1.2) itself converges to a saddle point if initialized sufficiently close to it (see Proposition 1.2 below).

In view of these results, one may expect that the two-sided PL condition in fact guarantees global convergence of the GDA flow to a saddle point (assuming one exists). The point of this paper is to show that this is not so.

Theorem 1.1. *There exists a $\mathcal{C}^\infty$ function $f : [-1,1]^2 \to \mathbb{R}$ with a unique critical point at the origin, and a constant $C < +\infty$ such that for every $x, y \in [-1,1]^2$,*

$$\begin{cases} f(x,y) - \inf_{x' \in [-1,1]} f(x',y) \leqslant C|\partial_x f(x,y)|^2, \\ \sup_{y' \in [-1,1]} f(x,y') - f(x,y) \leqslant C|\partial_y f(x,y)|^2, \end{cases} \tag{1.3}$$

and yet, for every $z(0)$ in some open subset of $[-1,1]^2$, the GDA flow given by (1.2) is periodic.

It is immediate to verify that if a function f satisfies (1.3) and admits a critical point, say at the origin (i.e. $\nabla f(0,0) = 0$), then this critical point is a saddle point, or more precisely, for every $x, y \in [-1,1]$, we have

$$f(0,y) \leqslant f(0,0) \leqslant f(x,0).$$

The function f that we build to show Theorem 1.1 is displayed on Figure 1 (up to a rescaling of the variables to bring them back to $[-1,1]^2$). For every (x,y) in a neighborhood of the origin, this function is given by $f(x,y) = \frac{\gamma}{2}x^2 + xy - \frac{\gamma}{2}y^2$, for some $\gamma \simeq 0.2531$. It is straightforward to check that this implies the convergence of the GDA flow towards the saddle point if one initializes the flow sufficiently close to the origin. This is not specific to this example, as we clarify in Proposition 1.2 below.

As we move away from the origin, we will progressively deform the function f so that the GDA flow then admits an integral of motion (i.e. a quantity that is preserved along the flow). This quantity is the L^4 norm after a rotation by $\pi/8$ (see also the red trajectory on Figure 1).

In the statement of Theorem 1.1, it is possible to replace the compact domain $[-1,1]^2$ by $\mathbb{R}^2$ if one so wishes. In order to do so, it indeed suffices to undo the said deformation, so that outside of a sufficiently large bounded region, the function f becomes again the quadratic form that we set it to be near the origin.

As announced, we also show for reference that the two-sided PL condition in (1.3) does guarantee local convergence to a saddle point. We state it in the two-dimensional case for convenience and consistency with the rest of the paper, but the argument can be generalized to higher dimensions. For every $r > 0$, we denote by B_r the open Euclidean ball of radius r centered at the origin.

Proposition 1.2 (local convergence)**.** *Let $f : [-1,1]^2 \to \mathbb{R}$ be a $\mathcal{C}^2$ function that has a critical point at the origin and is such that neither $f(0,\cdot)$ nor $f(\cdot,0)$ is a constant function, and let $C < +\infty$ be such that (1.3) holds for every $x, y \in [-1,1]^2$. There exists $r > 0$ such that for every $z(0) \in B_r$, the GDA flow given by (1.2) converges to the origin.*

We briefly mention a connection between the questions discussed here and recent works concerning the functional

$$F : \begin{cases} \mathcal{P}_H(\mathbb{T}^d) \times \mathcal{P}_H(\mathbb{T}^d) & \to \quad \mathbb{R} \\ (\mu,\nu) & \mapsto \quad \int g(x,y)\,\mathrm{d}\mu(x)\,\mathrm{d}\nu(y) + H(\mu) - H(\nu), \end{cases}$$

where $H(\mu) := \int_{\mathbb{T}^d} \log\left(\frac{\mathrm{d}\mu}{\mathrm{d}x}\right)\mathrm{d}\mu$ is the entropy of μ, $\mathcal{P}_H(\mathbb{T}^d)$ is the space of probability measures on the d-dimensional torus $\mathbb{T}^d$ with finite entropy, and $g : \mathbb{T}^d \times \mathbb{T}^d \to \mathbb{R}$ is a continuous function. In [16], the authors ask whether the Wasserstein GDA flow on F is convergent. Depending on the function g, the function F may not necessarily be convex-concave for the Wasserstein geometry. By [13], a two-sided PL assumption on F corresponds to assuming uniform log-Sobolev inequalities for the minimizers of $(F(\cdot,\nu))_{\nu\in\mathcal{P}(\mathbb{T}^d)}$ and $(-F(\mu,\cdot))_{\mu\in\mathcal{P}(\mathbb{T}^d)}$; this is automatic on $\mathbb{T}^d$ and would be a natural assumption if the torus $\mathbb{T}^d$ was replaced by the full space $\mathbb{R}^d$. In this infinite-dimensional context, local convergence results have been obtained in [15, 17], and two-timescale approaches have been considered in [10, 11].

The rest of the paper is organized as follows. In Section 2, we give simple sufficient conditions for a function defined on a compact interval of $\mathbb{R}$ to be PL, and for a function defined on a compact subset of $\mathbb{R}^2$ to satisfy the two-sided PL condition. We also prove Proposition 1.2. We then proceed to construct a function f that satisfies the requirements of Theorem 1.1 in Section 3. As discussed in the caption to Figure 1, we first build the level lines of f, which we identify as the flow lines of an explicit vector field. One key aspect of the construction is that the set of points where the level line is horizontal (i.e. the set of points where $\partial_x f = 0$) intersects any horizontal line only once (and similarly for the set of points of vertical tangency). Roughly speaking, the point of Section 2 is to show that this is essentially sufficient to guarantee that the function f satisfies the two-sided PL condition; the only additional information that we need next is that $\partial_x^2 f$ does not vanish wherever $\partial_x f = 0$, and similarly in the y direction. To complete the construction of the function f, we need to specify a value for each of the level lines. To facilitate this construction and also the verification of the two-sided PL condition, we do so by specifying the value of f on the set of points where $\partial_x f$ or $\partial_y f$ vanishes, which by our construction is a union of two lines (see the orange lines on Figure 1).

2. General properties of PL functions

Let I be a compact interval of $\mathbb{R}$. We say that a differentiable function $f : I \to \mathbb{R}$ satisfies the PL condition if there exists a constant $C < +\infty$ such that for every $x \in I$, we have

$$f(x) - \inf f \leqslant C|f'(x)|^2. \tag{2.1}$$

Proposition 2.1 (Criterion for PL condition)**.** *Let $f : I \to \mathbb{R}$ be a $\mathcal{C}^2$ function that is not constant. The function f satisfies the PL condition if and only if for every $x \in I$, we have the implication*

$$f'(x) = 0 \quad \Longrightarrow \quad f''(x) > 0. \tag{2.2}$$

Proof. We first show the direct implication, assuming that f satisfies the PL condition. Without loss of generality, we assume that $\inf f = 0$. We see from (2.1) that if we have $f'(x_0) = 0$ for some $x_0 \in I$, then we must have $f(x_0) = \inf f = 0$. Let J be a connected component of the open set $\{f > 0\}$. The set J is well defined since we assume that f is not constant. If this set is the entire interval I, then there is nothing to show. Else, one of the endpoints of J is different from $\{-1, 1\}$. For definiteness let us assume that the left endpoint of J, say x_0, is not -1. In this case we have $f(x_0) = 0$. We also note that f' cannot vanish in J (since this would imply the vanishing of f as previously observed), and since $f \geqslant 0$, we must have that $f' > 0$ on J. By (2.1), we deduce that for every $x \in J$,

$$\frac{f'(x)}{\sqrt{f(x)}} \geqslant \frac{1}{\sqrt{C}}.$$

Integrating this, we obtain that for every $x \in J$,

$$\sqrt{f(x)} \geqslant \frac{x - x_0}{2\sqrt{C}}.$$

This implies in particular that $f''(x_0) \geqslant \frac{1}{2C}$. We have seen that $f' > 0$ in J, so it is not possible that f vanishes at the right endpoint of J; in other words, we must have $J = (x_0, 1]$. Since $f''(x_0) > 0$, the same picture holds as well to the left of x_0: we must have $f' < 0$ everywhere on $[-1, x_0)$. The point x_0 is thus the unique point at which f' vanishes, and we have verified that $f''(x_0) > 0$, so the proof of the direct implication is complete.

We now turn to the converse implication, assuming that for every $x \in I$, the implication (2.2) is valid. Under the stated assumption, if $f'(x) = 0$ for some $x \in I$, then in a neighborhood of x, we have that f' is negative to the left of x and is positive to the right of x. This implies that f' can vanish at most once. If f' does not vanish anywhere, then by continuity of f' and compactness we have that $\mathsf{c} := \inf |f'|^2 > 0$, so it suffices to take $C = (\sup f - \inf f)/\mathsf{c}$ to ensure that (2.1) holds. Now let us assume that f' vanishes at some $x_0 \in I$. We can find a neighborhood U of x_0 such that

$f'' \geqslant f''(x_0)/2 =: c > 0$ on U. Using also that f'' is bounded, we can integrate this to

$$\text{for every } x \in U, \quad f(x) - f(x_0) \leqslant \frac{\sup f''}{2}(x - x_0)^2 \quad \text{and} \quad |f'(x)| \geqslant c|x - x_0|.$$

The PL condition (2.1) is thus satisfied for every $x \in U$ provided that we choose $C \geqslant \sup f''/(2c^2)$. Since $|f'|^2$ is bounded away from zero in $I \setminus U$, we can again make sure that (2.1) is satisfied by enlarging $C < +\infty$ as necessary. ☐

For a function $f : I^2 \to \mathbb{R}$, we say that f satisfies the two-sided PL condition if there exists a constant $C < +\infty$ such that (1.3) is satisfied for every $x, y \in I^2$, with the interval $[-1, 1]$ there replaced by I. In other words, the functions $(f(\cdot, y))_{y \in I}$ and $(-f(x, \cdot))_{x \in I}$ all satisfy the PL condition with the same constant.

Proof of Proposition 1.2. Using Proposition 2.1 and the two-sided PL assumption, we see that $\partial_x^2 f(0,0) > 0$ and $\partial_y^2 f(0,0) < 0$. At the origin, the Jacobian of the vector field $(-\partial_x f, \partial_y f)$ is

$$J := \begin{pmatrix} -\partial_x^2 f & -\partial_x \partial_y f \\ \partial_x \partial_y f & \partial_y^2 f \end{pmatrix} (0,0).$$

The symmetric part of this matrix, namely $(J + J^\mathsf{T})/2$, is negative definite. Hence, the eigenvalues of J have negative real part, and the claim thus follows by the linear stability theorem. ☐

Proposition 2.2 (criterion for two-sided PL)**.** *Let $f : I^2 \to \mathbb{R}$ be a $\mathcal{C}^2$ function. In order for the function f to satisfy the two-sided PL condition, it suffices that for every $z \in I^2$, the following two implications hold:*

$$\partial_x f(z) = 0 \quad \Longrightarrow \quad \partial_x^2 f(z) > 0, \tag{2.3}$$

$$\partial_y f(z) = 0 \quad \Longrightarrow \quad \partial_y^2 f(z) < 0. \tag{2.4}$$

Proof. By symmetry, it suffices to establish the uniform PL condition with respect to the first variable. Consider the set

$$\mathcal{Z} := \{z \in I^2 \ : \ \partial_x f(z) = 0\}.$$

Since the function f is $\mathcal{C}^2$, for every $z \in \mathcal{Z}$, one can find a constant $c_z > 0$ and an open neighborhood U_z of z such that $\partial_x^2 f \geqslant c_z$ in U_z. Since the set $\mathcal{Z}$ is compact, we can cover it with a finite number of such neighborhoods, and thus we can build an open set $U \subseteq I^2$ containing $\mathcal{Z}$ and a constant $c > 0$ such that $\partial_x^2 f \geqslant c$ on U. Arguing as in the proof of Proposition 2.1, we see that the condition (2.1) holds for all the partial functions $(f(\cdot, y))_{y \in I}$ inside U with the constant $C = \sup \partial_x^2 f/(2c^2)$. On the complement $I^2 \setminus U$, the derivative $|\partial_x f|$ is bounded away from zero, and thus one can adjust the constant in the PL condition so that it holds everywhere. ☐

3. Construction of the function

In order to construct a function f satisfying the requirements of Theorem 1.1, we will start by constructing its level lines, which will be described as the flow lines of a particular vector field $\mathbf{v}$. Before defining $\mathbf{v}$ itself, we need to set up a handful of properties of the polynomial

$$P(T) := T^3 + T^2 + T - \frac{1}{3}.$$

Substituting $U - 1/3$ for T and applying Cardano's formula, we find that this polynomial has a unique real root γ given by

$$\gamma := \frac{1}{3}\left(\sqrt[3]{6\sqrt{2}+8} - \sqrt[3]{6\sqrt{2}-8} - 1\right) \simeq 0.2531,$$

and that

$$P(T) = (T-\gamma)\left(T^2 + aT + b\right),$$

where for convenience, we have set $a := \gamma + 1 \simeq 1.2531$ and $b := 1/(3\gamma) = \gamma^2 + \gamma + 1 \simeq 1.3171$. We let $\varphi : \mathbb{R}_+ \to [1, +\infty)$ denote a $\mathcal{C}^\infty$ function that is constant equal to 1 on $[0, 1/2]$, is non-decreasing, and is twice the identity on $[1, +\infty)$. Concretely, we can for instance take

$$\varphi(t) = \begin{cases} 1 & \text{if } t \leqslant \frac{1}{2}, \\ 1 + (2t-1)\,\rho(2t-1) & \text{if } t \in (\frac{1}{2}, 1), \\ 2t & \text{if } t \geqslant 1, \end{cases}$$

where $\rho : (0,1) \to (0,1)$ is given by $\rho(u) := e^{-1/u} \big/ (e^{-1/u} + e^{-1/(1-u)})$. With this in place, we define, for every $x, y \in \mathbb{R}$,

$$\mathbf{v}(x,y) = \begin{pmatrix} \mathbf{v}_1(x,y) \\ \mathbf{v}_2(x,y) \end{pmatrix} := \begin{pmatrix} (\gamma y - x)\varphi(x^2 + axy + by^2) \\ (y + \gamma x)\varphi(y^2 - axy + bx^2) \end{pmatrix}. \tag{3.1}$$

We say that a trajectory $(M(t))_{t \in I}$, where I is an interval of $\mathbb{R}$ and $M(t) \in \mathbb{R}^2$, is a *flow line of* $\mathbf{v}$ if for every $t \in I$, we have

$$\partial_t M(t) = \mathbf{v}(M(t)).$$

We say that it is a *complete flow line of* $\mathbf{v}$ if it is a flow line of $\mathbf{v}$ and the interval I is maximal for this property. We also introduce the notation

$$\mathcal{X} := \{(x,y) \in \mathbb{R}^2 \ : \ x = \gamma y \text{ or } y = -\gamma x\}.$$

Proposition 3.1. *Let $(M(t))_{t \in I}$ be a complete flow line of $\mathbf{v}$. Exactly one of these four possibilities is valid.*

(0) The flow line stays put at the origin for all times.
(1) The flow line intersects $\mathcal{X}$ exactly once.
(2) The flow line never intersects $\mathcal{X}$, and $\lim_{t \to +\infty} M(t) = 0$.
(3) The flow line never intersects $\mathcal{X}$, and $\lim_{t \to -\infty} M(t) = 0$.

Proof. The origin is indeed a fixed point. For the other cases, we make use of the change of variables $\ell_1 := x - \gamma y$ and $\ell_2 := y + \gamma x$, so that $\mathcal{X} = \{\ell_1 = 0\} \cup \{\ell_2 = 0\}$, and we also denote $\varphi_1 := \varphi(x^2 + axy + by^2) \geqslant 1$ and $\varphi_2 := \varphi(y^2 - axy + bx^2) \geqslant 1$. We use the notation $M(t) = (x(t), y(t))$ and use an upper dot to indicate the time derivative along the flow line. Using $\dot{x} = \mathbf{v}_1 = -\ell_1\varphi_1$ and $\dot{y} = \mathbf{v}_2 = \ell_2\varphi_2$, we see that

$$\dot{\ell}_1 = -\ell_1\varphi_1 - \gamma\ell_2\varphi_2, \qquad \dot{\ell}_2 = -\gamma\ell_1\varphi_1 + \ell_2\varphi_2.$$

We introduce the four open quadrants $Q_{\pm\pm}$ defined by

$$Q_{\pm\pm} := \{\pm\ell_1 > 0 \quad \text{and} \quad \pm\,\ell_2 > 0\}.$$

Step 1. We first argue that the flow line can cross $\{\ell_1 = 0\} \cup \{\ell_2 = 0\}$ at most once. On the half-line $\{\ell_1 = 0,\, \ell_2 > 0\}$, we have $\dot{\ell}_1 = -\gamma\ell_2\varphi_2 < 0$, so every crossing there goes from Q_{++} into Q_{-+}. On $\{\ell_2 = 0,\, \ell_1 > 0\}$, we have $\dot{\ell}_2 = -\gamma\ell_1\varphi_1 < 0$, so every crossing goes from Q_{++} into Q_{+-}. Similarly, the two boundary segments $\{\ell_1 = 0,\, \ell_2 < 0\}$ and $\{\ell_2 = 0,\, \ell_1 < 0\}$ are crossed from Q_{--} into Q_{+-} and Q_{-+} respectively. In particular, every trajectory that enters Q_{-+} never leaves it, and similarly for Q_{+-}. Conversely, every backward-in-time trajectory that enters Q_{++} or Q_{--} never leaves it. Since every crossing of $\{\ell_1 = 0\} \cup \{\ell_2 = 0\}$ away from the origin corresponds to a transition from $\{Q_{++}, Q_{--}\}$ into $\{Q_{-+}, Q_{+-}\}$, at most one such crossing can occur along any trajectory.

Step 2. We now argue that if a flow line stays forever in Q_{++}, then it must converge to the origin. Recalling that $\dot{\ell}_1 = -\ell_1\varphi_1 - \gamma\ell_2\varphi_2$ and that $\varphi_1, \varphi_2 \geqslant 1$, we see that in order for ℓ_1 to remain positive for all times, we must have that ℓ_1 and ℓ_2 tend to zero as t tends to infinity. The argument for the other quadrants is similar. □

Since the vector field $\mathbf{v}$ is smooth, it is not possible for two flow lines to intersect. In order to define the function f, it thus suffices to prescribe its value on every complete flow line. Let $(M(t))_{t\in I}$ be a complete flow line. If this flow line never intersects $\mathcal{X}$, or if it stays put at the origin, then we set $f(M(t)) = 0$ for every $t \in I$. If the flow line does intersect $\mathcal{X}$, then by Proposition 3.1, there exists a unique time $t_0 \in I$ such that $M(t_0)$ belongs to $\mathcal{X}$. Writing $(x_0, y_0) = M(t_0)$, we set, for every $t \in I$,

$$f(M(t)) = \frac{\gamma}{2}x_0^2 + x_0y_0 - \frac{\gamma}{2}y_0^2.$$

We note that the prescribed value is never zero in this case, since for every $x, y \in \mathbb{R}$, we have

$$f(\gamma y, y) = \frac{1}{2}(\gamma^3 + \gamma)y^2 \quad \text{and} \quad f(x, -\gamma x) = -\frac{1}{2}(\gamma^3 + \gamma)x^2. \tag{3.2}$$

We recall that we denote by B_r the open Euclidean ball of radius r centered at the origin.

Proposition 3.2. *There exists $r > 0$ such that for every $(x, y) \in B_r$, we have*

$$f(x,y) = \frac{\gamma}{2}x^2 + xy - \frac{\gamma}{2}y^2. \tag{3.3}$$

Proof. Step 1. In this step, we introduce some notation and a convenient change of coordinates. For every $x, y \in \mathbb{R}$, we write $g(x,y) := \frac{\gamma}{2}x^2 + xy - \frac{\gamma}{2}y^2$ and

$$\mathbf{w}(x,y) := \begin{pmatrix} \gamma y - x \\ y + \gamma x \end{pmatrix}. \tag{3.4}$$

Since the quadratic forms $x^2 + axy + by^2$ and $y^2 - axy + bx^2$ are positive definite (recall that $T^2 + aT + b$ has no real root), there exists $r_0 > 0$ such that both are at most $\frac{1}{2}$ on B_{r_0}. Since $\varphi = 1$ on $[0, \frac{1}{2}]$, we have $\mathbf{v} = \mathbf{w}$ on B_{r_0}.

We set $\mu := \sqrt{1 + \gamma^2}$ and introduce the linear coordinates

$$u := \gamma x + (1 + \mu)y, \qquad v := \gamma x + (1 - \mu)y.$$

A direct computation shows that along the flow lines of $\mathbf{w}$, we have $\dot{u} = \mu u$ and $\dot{v} = -\mu v$, so that uv is preserved along the flow. Moreover, for every $x, y \in \mathbb{R}$, we have

$$g(x,y) = \frac{\gamma}{2}x^2 + xy - \frac{\gamma}{2}y^2 = \frac{1}{2\gamma}\, uv. \tag{3.5}$$

Setting $\kappa := (\mu + 1)/(\mu - 1)$, one can check that, in (u, v) coordinates,

$$\mathcal{X} = \{u = \kappa v\} \cup \{u = -v\}.$$

In particular, the set $\mathcal{X}$ meets every open quadrant of the (u, v)-plane: the line $\{u = \kappa v\}$ enters the first and third quadrants, while $\{u = -v\}$ enters the second and fourth.

Step 2. We now show that $f = g$ on B_r for a suitable $r \in (0, r_0]$.

Consider first a point $(x, y) \in B_r$ with $g(x, y) \neq 0$, i.e. $c := uv \neq 0$. The flow line of $\mathbf{w}$ through this point is an arc of the hyperbola $\{uv = c\}$. Each branch of this hyperbola lies in a single open quadrant and crosses $\mathcal{X}$ exactly once, at a point whose distance to the origin is $O(\sqrt{|c|})$. Since u and v are linear in (x, y), we have $|c| = O(r^2)$ for $(x, y) \in B_r$. Furthermore, on a branch of $\{uv = c\}$ the quantity $u^2 + v^2 = u^2 + c^2/u^2$ is convex in u^2, so the subarc connecting any point to the crossing point lies within the ball whose radius is the larger of the two endpoint distances. Therefore, choosing $r > 0$ small enough, we can ensure that the flow line of $\mathbf{w}$ through (x, y) reaches $\mathcal{X}$ without leaving B_{r_0}.

Since $\mathbf{v} = \mathbf{w}$ on B_{r_0}, this arc is equally part of a flow line of $\mathbf{v}$, which thus intersects $\mathcal{X}$. By definition, f is constant along this flow line and equals g at the crossing point; since g is also constant along this flow line by (3.5), we conclude that $f(x, y) = g(x, y)$.

It remains to treat $\{g = 0\} \cap B_r$. The four half-axes $\{u = 0, \pm v > 0\}$ and $\{v = 0, \pm u > 0\}$ are the only flow lines of $\mathbf{w}$ that converge to the origin (as

$t \to +\infty$ for the first pair, as $t \to -\infty$ for the second). The flow lines of $\mathbf{v}$ that converge to the origin as t tends to $\pm\infty$ from Proposition 3.1 must therefore, once restricted to B_r, coincide with the lines along which g vanishes. □

Proposition 3.3. *The function f is $\mathcal{C}^\infty$ on $\mathbb{R}^2$, has a unique critical point at the origin, and for every $R < +\infty$, it satisfies the two-sided PL condition on $[-R, R]^2$.*

Proof. By Proposition 3.2, there exists $r > 0$ such that the identity (3.3) holds for every $(x, y) \in B_r$. In particular, the restriction of f to B_r is $\mathcal{C}^\infty$ and has a unique critical point at the origin. Consider now a point $z \in \mathcal{X} \setminus \{0\}$; locally around z, the set $\mathcal{X}$ is a line that is neither horizontal nor vertical. Since exactly one of the coordinates of $\mathbf{v}$ vanishes at z, we see that $\mathbf{v}$ is transversal to $\mathcal{X}$ at z. By the implicit function theorem, we can assert that f is $\mathcal{C}^\infty$ in a neighborhood of z. Now, for $z \in \mathbb{R}^2 \setminus \{0\}$ arbitrary, we learn from Proposition 3.1 that there is a characteristic line that connects it either to a point in $\mathcal{X}$, or to a point in B_r. Hence, there is a $\mathcal{C}^\infty$ diffeomorphism that preserves f and that maps z to a point at which f is $\mathcal{C}^\infty$. This implies that f is $\mathcal{C}^\infty$ at z as well. By (3.2), we see that the gradient of f never vanishes on $(\mathcal{X} \cup B_r) \setminus \{0\}$. By the same diffeomorphism argument, this implies that the gradient of f does not vanish anywhere on $\mathbb{R}^2 \setminus \{0\}$.

It remains to show that for every $R < +\infty$, the function f satisfies the two-sided PL condition on $[-R, R]^2$. By Proposition 2.2, it suffices to verify that (2.3) and (2.4) are valid. These conditions are valid at the origin, by Proposition 3.2. Let $z = (x, y) \in \mathbb{R}^2 \setminus \{0\}$ be a point such that $\partial_x f(z) = 0$. Since f has no critical point besides the origin, we must have that $\partial_y f(z) \neq 0$. Since ∇f must be orthogonal to $\mathbf{v}$, it follows that $\mathbf{v}_2(z) = 0$, that is, $y + \gamma x = 0$ (and in particular $z \in \mathcal{X}$). For definiteness, let us assume that $x > 0$. Using that $\partial_x f(z) = 0$ and (3.2), we see that $\partial_y f(z) > 0$. Differentiating in x the relation $\nabla f \cdot \mathbf{v} = 0$ and using that $\partial_x f(z) = 0$ and $\mathbf{v}_2(z) = 0$, we find that

$$\left(\partial_x^2 f\, \mathbf{v}_1 + \partial_y f\, \partial_x \mathbf{v}_2\right)(z) = 0. \tag{3.6}$$

Since $\mathbf{v}_2(x', y')$ is of the form $(y' + \gamma x')$ times some function taking only positive values, we have that $\partial_x \mathbf{v}_2(z) > 0$. Since $x > 0$ and $y < 0$, we have $\mathbf{v}_1(x, y) < 0$. Recalling also that $\partial_y f(z) > 0$, we conclude from (3.6) that $\partial_x^2 f(z)$ must be strictly positive, as desired. The other cases can be handled similarly. □

We are now ready to prove the theorem.

Proof of Theorem 1.1. Up to a change of scale, we may as well show the statement of Theorem 1.1 with the domain $[-1, 1]^2$ replaced by $[-R, R]^2$, for some $R < +\infty$ of our choosing. For $R < +\infty$ sufficiently large, we have for every $(x, y) \in \mathbb{R}^2 \setminus B_R$ that

$$\mathbf{v}(x, y) = 2\begin{pmatrix} -y^3 P(x/y) \\ -x^3 P(-y/x) \end{pmatrix} = 2\begin{pmatrix} -x^3 - x^2 y - x y^2 + \frac{y^3}{3} \\ y^3 - x y^2 + x^2 y + \frac{x^3}{3} \end{pmatrix}. \tag{3.7}$$

For every $x, y \in \mathbb{R}$, we define

$$g(x,y) := 3x^4 + 4x^3y + 6x^2y^2 - 4xy^3 + 3y^4.$$

On $\mathbb{R}^2 \setminus B_R$, we have

$$\begin{pmatrix} \partial_x g \\ \partial_y g \end{pmatrix} = 6 \begin{pmatrix} -\mathbf{v}_1 \\ \mathbf{v}_2 \end{pmatrix}. \tag{3.8}$$

Let $(z(t))_{t\in I}$ be a GDA trajectory that stays in $\mathbb{R}^2 \setminus B_R$. By the definition in (1.2), we have

$$\partial_t(g(z(t))) = \begin{pmatrix} \partial_x g \\ \partial_y g \end{pmatrix} \cdot \begin{pmatrix} -\partial_x f \\ \partial_y f \end{pmatrix} (z(t)).$$

Using (3.8) and recalling that ∇f and $\mathbf{v}$ are orthogonal, we see that g remains constant along the trajectory of $(z(t))_{t\in I}$. We now observe that every level line of g is the boundary of an L^4 ball rotated by $\pi/8$. Indeed, this follows from the observation that, for $\alpha := \tan(\pi/8) = \sqrt{2} - 1$, we have

$$\begin{aligned} &|x + \alpha y|^4 + |y - \alpha x|^4 \\ &\quad = (1+\alpha^4)x^4 + 4\alpha(1-\alpha^2)x^3y + 12\alpha^2x^2y^2 - 4\alpha(1-\alpha^2)xy^3 + (1+\alpha^4)y^4 \\ &\quad = 2\alpha^2 g(x,y). \end{aligned}$$

It is therefore clear that the level lines of g are closed curves that circle around the origin, and that some of those stay outside of B_R. We choose R' such that one such level line is contained in $[-R', R']^2$. A GDA flow starting from a point on this level line will forever circle around it, since g must remain constant along the GDA flow and the function f has no critical point along this level line, by Proposition 3.3. By this same proposition, the restriction of the function f to $[-R', R']^2$ thus satisfies all the requirements to yield the validity of Theorem 1.1. □

Acknowledgements. I would like to warmly thank Loucas Pillaud-Vivien for introducing me to this problem and for helpful feedback. I acknowledge the support of the ERC MSCA grant SLOHD (101203974), and of the French National Research Agency (ANR) under the France 2030 grant ANR-24-RRII-0002 operated by the Inria Quadrant Program.

References

[1] Aharon Ben-Tal, Laurent El Ghaoui, and Arkadi Nemirovski. *Robust Optimization*. Princeton University Press, Princeton, NJ, 2009.

[2] Vladimir Fedorovich Dem'yanov and Aleksandr Borisovich Pevnyi. Numerical methods for finding saddle points. *USSR Comp. Math. and Math. Phys.*, 12(5):11–52, 1972.

[3] Thinh Doan. Convergence rates of two-time-scale gradient descent-ascent dynamics for solving nonconvex min-max problems. In *Learning for Dynamics and Control Conference*, pages 192–206. PMLR, 2022.

[4] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, volume 27, pages 2672–2680, 2014.

[5] Hamed Karimi, Julie Nutini, and Mark Schmidt. Linear convergence of gradient and proximal-gradient methods under the Polyak-Łojasiewicz condition. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 795–811. Springer, 2016.

[6] Galina M. Korpelevich. An extragradient method for finding saddle points and for other problems. *Ekonom. i Mat. Metody*, 12(4):747–756, 1976.

[7] Tianyi Lin, Chi Jin, and Michael Jordan. On gradient descent ascent for nonconvex-concave minimax problems. In *International Conference on Machine Learning*, pages 6083–6093. PMLR, 2020.

[8] Stanislaw Łojasiewicz. A topological property of real analytic subsets. *Coll. du CNRS, Les équations aux dérivées partielles*, 117(87-89):2, 1963.

[9] Songtao Lu, Ioannis Tsaknakis, Mingyi Hong, and Yongxin Chen. Hybrid block successive approximation for one-sided non-convex min-max problems: algorithms and applications. *IEEE Transactions on Signal Processing*, 68:3676–3691, 2020.

[10] Yulong Lu. Two-scale gradient descent ascent dynamics finds mixed Nash equilibria of continuous games: A mean-field perspective. In *International Conference on Machine Learning*, pages 22790–22811. PMLR, 2023.

[11] Chao Ma and Lexing Ying. Provably convergent quasistatic dynamics for mean-field two-player zero-sum games. In *International Conference on Learning Representations*, 2022.

[12] 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. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

[13] Felix Otto and Cédric Villani. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.*, 173(2):361–400, 2000.

[14] Boris Teodorovich Polyak. Gradient methods for minimizing functionals. *Zh. Vychisl. Mat. Mat. Fiz.*, 3(4):643–653, 1963.

[15] Geuntaek Seo, Minseop Shin, Pierre Monmarché, and Beomjun Choi. Local exponential stability of mean-field Langevin descent-ascent in Wasserstein space. Preprint, arXiv:2602.01564.

[16] Guillaume Wang and Lénaïc Chizat. Open problem: Convergence of single-timescale mean-field Langevin descent-ascent for two-player zero-sum games. In *The Thirty Seventh Annual Conference on Learning Theory*, pages 5345–5350. PMLR, 2024.

[17] Guillaume Wang and Lénaïc Chizat. Local convergence of mean-field Langevin dynamics: from gradient flows to linearly monotone games. Preprint, arXiv:2602.11999.

[18] Junchi Yang, Negar Kiyavash, and Niao He. Global convergence and variance reduction for a class of nonconvex-nonconcave minimax problems. In *Advances in Neural Information Processing Systems*, volume 33, pages 1153–1165, 2020.

(Jean-Christophe Mourrat) Department of Mathematics, ENS Lyon and CNRS, Lyon, France