EQUILIBRIUM STRUCTURE OF HIGH-RESOLUTION DIFFERENTIAL EQUATIONS FOR MIN–MAX OPTIMIZATION

Federico Praolini
Politecnico di Milano
{federico.praolini}@polimi.it

Tatjana Chavdarova
TU Wien
{tatjana.chavdarova}@tuwien.ac.at

ABSTRACT

High-resolution differential equations (HRDEs) provide refined continuous-time models for first-order methods in saddle-point and games. In this work, we show that the equilibrium structure of HRDEs need not coincide with the solution set of the underlying problem. We introduce a general framework to characterize equilibria of HRDEs and formalize the notion of *spurious equilibria*: stationary points of the continuous-time dynamics that are not solutions of the original game, as per their first-order characterization. We show that HRDEs associated with gradient descent–ascent are faithful in this sense, while HRDEs for extragradient and similar methods may admit additional equilibria induced by nonlinear algorithmic correction terms. We derive explicit conditions for the existence of such equilibria and analyze their stability. Our results highlight a structural limitation of continuous-time models beyond standard gradient flows and call for care when using HRDEs or related ODEs to reason about algorithmic solutions or behaviors.

1 INTRODUCTION

Understanding learning dynamics in games and saddle-point problems is substantially more challenging than in single-objective optimization. The interaction between players induces rotational components in the vector field, leading to oscillations and complex transient behavior that are difficult to analyze directly in discrete time (Korpelevich, 1976; Balduzzi et al., 2018; Mazumdar et al., 2020; Mertikopoulos et al., 2018; Adolphs et al., 2018; Chavdarova et al., 2019; Golowich et al., 2020). As a result, continuous-time viewpoints have become a fundamental tool for studying learning in games, providing tractable models that isolate the geometric and dynamical properties of algorithms (Hemmat et al., 2020; Tseng, 2000). Continuous-time models have been shown to clarify the role of rotations, mitigate oscillatory behavior, and provide convergence guarantees that are difficult to establish directly in discrete time (Hofbauer & Sandholm, 2009; Mertikopoulos et al., 2018; Daskalakis & Panageas, 2018; Chavdarova et al., 2023). They have also been instrumental in characterizing asymptotic behaviors such as cycling (Hsieh et al., 2021). In this line of work, ODEs serve as idealized limits that expose the qualitative structure of learning dynamics.

Two complementary paradigms underlie the use of continuous-time models in learning in games. In the first, a continuous-time limit is derived from a given discrete-time algorithm in order to approximate its behavior; in the second, one designs a continuous-time dynamical system with desirable properties and subsequently discretizes it to obtain a practical method. A key limitation of the former is that many distinct first-order min–max algorithms—such as gradient descent–ascent, extragradient, and optimistic variants—collapse to the same ODE. To address this, Chavdarova et al. (2023) introduce *high-resolution differential equations* (HRDEs), which retain higher-order information in the step size (see Appendix B.2) and yield distinct dynamics for different algorithms, enabling refined analyses of stability and last-iterate convergence.

In this work, relevant to both paradigms above, we study the equilibrium structure of HRDEs and related continuous-time models. We show that HRDEs associated with extragradient-type methods may admit equilibrium points that do not correspond to solutions of the underlying saddle-point problem, herein called *spurious equilibria* (SE). SEs arise from nonlinear algorithm-induced correction terms in the dynamics and are not present in classical gradient flows. While SEs are typically unstable in the examples we consider, their existence reveals a structural mismatch between the solution set of the continuous-time system and that of the original problem. Our results therefore, highlight a general caveat: equilibrium points of a continuous-time learning dynamic need not faithfully represent solutions of the underlying game.

The remainder of the paper formalizes these observations by introducing a general HRDE model, defining algorithm-induced solution sets, and characterizing conditions under which such spurious equilibria arise. A detailed discussion of related work is deferred to Appendix A.

2 SETTING & PRELIMINARIES

Setting. Given a non-linear function $f(\boldsymbol{x}, \boldsymbol{y}) \colon \mathbb{R}^d \to \mathbb{R}$, the saddle point problem is:

$$\min_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{y}\in\mathcal{Y}} f(\boldsymbol{x}, \boldsymbol{y}) \,, \qquad \text{with } \mathcal{X} \equiv \mathbb{R}^{d_1},\ \mathcal{Y} \equiv \mathbb{R}^{d_2},\ d_1 + d_2 = d \,. \tag{SP}$$

In the following, we will denote $\boldsymbol{z}$ the vector of states $[\boldsymbol{x}, \boldsymbol{y}]$. The vector field of the SP problem $V(\boldsymbol{z})$ and its Jacobian $J(\boldsymbol{z})$ are:

$$V(\boldsymbol{z}) := \begin{bmatrix} \nabla_{\boldsymbol{x}} f(\boldsymbol{z}) \\ -\nabla_{\boldsymbol{y}} f(\boldsymbol{z}) \end{bmatrix} \quad \text{and} \quad (V) \qquad J(\boldsymbol{z}) := \begin{bmatrix} \nabla^2_{\boldsymbol{x}} f(\boldsymbol{z}) & \nabla_{\boldsymbol{x}\boldsymbol{y}} f(\boldsymbol{z}) \\ -\nabla_{\boldsymbol{y}\boldsymbol{x}} f(\boldsymbol{z}) & -\nabla^2_{\boldsymbol{y}} f(\boldsymbol{z}) \end{bmatrix} \text{, resp.} \quad (J)$$

A point $\bar{\boldsymbol{z}}$ that nullifies the vector field is called herein a *problem equilibrium* (PE) of equation SP.

Discrete-time saddle-point algorithms. Many first-order algorithms for saddle-point problems (see Appendix B.3) can be written in the following generic discrete-time form:

$$\boldsymbol{z}_{k+1} = \boldsymbol{z}_k + \gamma\, \Phi(\boldsymbol{z}_k, \boldsymbol{z}_{k-1}) \,, \tag{G-Disc}$$

where $\gamma > 0$ is a step-size, and Φ a general algorithmic update such as gradient descent–ascent—with $\Phi \equiv -V(\boldsymbol{z})$, extragradient (Korpelevich, 1976), or lookahead (Chavdarova et al., 2021); refer to Appendix B.3.

GD– and EG–HRDEs. The high-resolution dynamical form of *gradient descent* reads as:

$$\begin{cases} \dot{\boldsymbol{z}}(t) = \boldsymbol{\omega}(t) \\ \dot{\boldsymbol{\omega}}(t) = -\frac{2}{\gamma}\boldsymbol{\omega}(t) - \frac{2}{\gamma} V(\boldsymbol{z}(t)) \,, \end{cases} \tag{GD-HRDE}$$

with $\boldsymbol{\xi} := \big(\boldsymbol{z}(t), \boldsymbol{\omega}(t)\big)$, γ denoting the selected step-size. Extragradient's continuous-time dynamic contain an additional term in the second equation as follows:

$$\begin{cases} \dot{\boldsymbol{z}}(t) = \boldsymbol{\omega}(t) \\ \dot{\boldsymbol{\omega}}(t) = -\frac{2}{\gamma}\boldsymbol{\omega}(t) - \frac{2}{\gamma} V(\boldsymbol{z}(t)) + 2J(\boldsymbol{z}(t))V(\boldsymbol{z}(t)) \,; \end{cases} \tag{EG-HRDE}$$

refer to (Chavdarova et al., 2023) for their derivation.

3 DYNAMICS-INDUCED SPURIOUS EQUILIBRIA

General HRDE model. We generalize the introduced HRDEs, encompassing the models introduced above as well as related variants, as follows:

$$\begin{cases} \dot{\boldsymbol{z}}(t) = \boldsymbol{\omega}(t) \\ \dot{\boldsymbol{\omega}}(t) = -\frac{2}{\gamma}\boldsymbol{\omega}(t) + \Omega(\boldsymbol{z}(t)) \end{cases} \quad , \tag{G-HRDE}$$

where $\Omega(\boldsymbol{z}) \in \mathbb{R}^{[d\times 1]}$ is referred to as the *algorithmic drift*, capturing the algorithm-specific forcing induced by the underlying discrete method. Throughout, we assume that the algorithmic drift depends only on the state $\boldsymbol{z}$ and not on the velocity $\boldsymbol{\omega}$, and that $\Omega(\boldsymbol{z}) = 0$ whenever the saddle-point vector field V vanishes. These assumptions are satisfied by a broad class of HRDEs, including all the algorithms considered in this work. When necessary, we denote the drift explicitly as $\Omega^{SP}_{\text{ALG}}(\boldsymbol{z})$ to emphasize its dependence on the chosen algorithm ALG and saddle-point problem SP. System G-HRDE defines a $2\cdot d$–dimensional autonomous dynamical system with state $\boldsymbol{\xi} \in \mathbb{R}^{[2d\times 1]}$.

Dynamics Equilibria (DEs). A point $\boldsymbol{\xi}^\star$ is an *equilibrium of a dynamical system* if the time derivatives equal zero at that point, i.e., $\dot{\boldsymbol{\xi}}(t) = 0$. Equivalently, if initialized at a DE, the system remains stationary for all time.

For the HRDE system G-HRDE, its DE is a point $\boldsymbol{\xi}^\star = [\boldsymbol{z}^\star; \boldsymbol{\omega}^\star]^T$ satisfying $\dot{\boldsymbol{\xi}} = 0$. From G-HRDE structure, the following properties can be deduced.

Property 1 (G-HRDE)**.** *Every dynamics equilibrium (DE) has structure:* $\boldsymbol{\xi}^\star = \begin{bmatrix} \boldsymbol{z}^\star \\ \mathbf{0}^{[d\times 1]} \end{bmatrix}$.

Property 2 (G-HRDE)**.** *Every dynamics equilibrium (DE) satisfies:* $\Omega(\boldsymbol{z}^\star) = \mathbf{0}$ *.*

Property 3 (G-HRDE)**.** *A point $\boldsymbol{\xi}^\star$ is a DE if and only if it satisfies both property 1 and 2.*

We now address two natural questions regarding the relationship between equilibria of the saddle problem and equilibria of the induced dynamics:

1. ($\Rightarrow$) *Does every problem equilibrium (PE) $\bar{z}$ correspond to the $\boldsymbol{z}^\star$ component of a dynamics equilibrium (DE) $\boldsymbol{\xi}^\star$?*
2. ($\Leftarrow$) *Does the $\boldsymbol{z}^\star$ component of a dynamics equilibrium (DE) $\boldsymbol{\xi}^\star$ correspond to a problem equilibrium (PE) $\bar{z}$?*

The answers are *yes* to Question 1 and, in general, *no* to Question 2. The former follows directly from our standing assumption on the algorithmic drift: whenever the saddle-point vector field equation V vanishes, the algorithmic drift $\Omega(\boldsymbol{z})$ vanishes as well. Consequently, any PE induces a DE.

Property 4. *Every problem equilibrium $\bar{z}$ corresponds to the $\boldsymbol{z}^\star$ component of a dynamics equilibrium (DE) of equation G-HRDE.*

The converse implication does not hold in general. Indeed, $\Omega(\boldsymbol{z}^\star) = \mathbf{0}$ does not imply $V(\boldsymbol{z}^\star) = \mathbf{0}$, so a DE may exist at a point $\boldsymbol{z}^\star$ that is not a PE. In other words, the HRDE dynamics may admit more equilibria than the underlying saddle problem, while the reverse cannot occur. We refer to these additional equilibria as *spurious dynamics equilibria* and formalize this distinction below.

Definition 1. *Let $\mathcal{T}(f(\boldsymbol{z}))$ denote the set of problem equilibria (PEs) of the saddle problem SP:*

$$\mathcal{T}(f(\boldsymbol{z})) = \{\bar{z}_1, ..., \bar{z}_n\}\,.$$

*This set is called the **true solution set**.*

Definition 2. *Let $\mathcal{H}_{ALG}(f(\boldsymbol{z}))$ denote the set of the $\boldsymbol{z}_1^\star, \dots, \boldsymbol{z}_n^\star$ components all dynamic equilibria (DEs) $\boldsymbol{\xi}_1^\star, \dots, \boldsymbol{\xi}_n^\star$ of the HRDE associated with algorithm ALG applied to the saddle problem SP:*

$$\mathcal{H}_{ALG}(f(z)) = \{\boldsymbol{z}_1^\star, \dots, \boldsymbol{z}_n^\star\}\,.$$

*We refer to this set as the **algorithm solution set**.*

Property 4 yields the following inclusion; see Figure 1.

Property 5. *For any algorithm ALG, the algorithm solution set contains the true solution set:*

$$\mathcal{H}_{ALG}(f(\boldsymbol{z})) \supseteq \mathcal{T}(f(\boldsymbol{z}))\,.$$

We can now formally define spurious equilibria.

Definition 3 (spurious equilibria)**.** *For a given saddle problem SP and algorithm ALG, any element of $\mathcal{H}_{ALG}(f(\boldsymbol{z}))$ that does not belong to $\mathcal{T}(f(\boldsymbol{z}))$ is called a **spurious equilibrium** (SE). Equivalently, SEs are dynamics equilibria (DEs) that are not problem equilibria (PEs).*

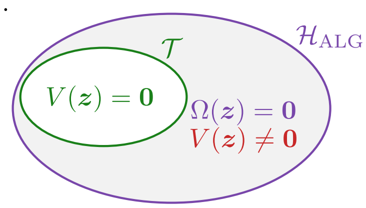


Figure 1: Pictorial: true solution set $\mathcal{T}$ (PEs) and algorithm solution set $\mathcal{H}_{\mathrm{ALG}}$ (DEs). Spurious equilibria correspond to elements of $\mathcal{H}_{\mathrm{ALG}} \setminus \mathcal{T}$—shaded area.

SE set of GD-HRDE. In this case, the algorithmic drift coincides with the vector field V scaled by the step size. Consequently, any point $\boldsymbol{z}^\star$ satisfying 2 must also satisfy $V(\boldsymbol{z}^\star) = 0$. Therefore, every dynamics equilibrium (DE) corresponds to a problem equilibrium (PE); the converse implication always holds by Lemma 4. This yields the following result.

Claim 1 (No-spuriousness of GD-HRDE)**.** *For any saddle problem $f(\boldsymbol{z})$, GD-HRDE satisfies:*

$$\mathcal{T}(f(\boldsymbol{z})) \equiv \mathcal{H}_{GDA}(f(\boldsymbol{z}))\,, \qquad \forall \gamma > 0\,. \tag{1}$$

DE and SE sets of EG-HRDE. We now turn to the EG-HRDE equation EG-HRDE. A dynamics equilibrium (DE) point of this system is obtained by imposing $\dot{\boldsymbol{\xi}} = \mathbf{0}$, yielding condition (see Appendix C):

$$(I - \gamma J(\boldsymbol{z}^\star))V(\boldsymbol{z}^\star) = 0. \tag{2}$$

Unlike the GD case, this condition may be satisfied at points $\boldsymbol{z}^\star$ for which $V(\boldsymbol{z}^\star) \neq \mathbf{0}$. Any $\boldsymbol{z}^\star$ that satisfies equation 2 and such that $V(\boldsymbol{z}^\star) \neq 0$ is a spurious equilibrium (SE). Further interpretation is provided in Appendix C.1. For two-dimensional systems, the following characterization holds (proved in Appendix C.2).

Claim 2 (Spuriousness of EG-HRDE: two-dimensional setting)**.** *Given a saddle problem $f(\boldsymbol{z})$, a dynamic equilibrium (DE) $\boldsymbol{z}^\star$ of the EG-HRDE is spurious if:*

$$1 - \gamma\,\mathrm{tr}\big(J(\boldsymbol{z}^\star)\big) + \gamma^2 \det\big(J(\boldsymbol{z}^\star)\big) = 0\,. \tag{3}$$

Condition (3) is necessary but not sufficient. It can nevertheless be used to partition the algorithm solution set $\mathcal{H}_{EG}(f(\boldsymbol{z}))$ into PEs and SEs.

Finally, observe that if the equilibrium condition 2 is linear and non-degenerate, it admits unique solution, which must coincide with the point where $V(\boldsymbol{z}) = \mathbf{0}$. This leads to the following general statement.

Claim 3 (Necessary condition for spurious equilibria.)**.** *Let ALG be an algorithm whose algorithmic drift $\Omega^{SP}_{ALG}(\boldsymbol{z})$ is linear in $\boldsymbol{z}$ and non degenerate. Then,*

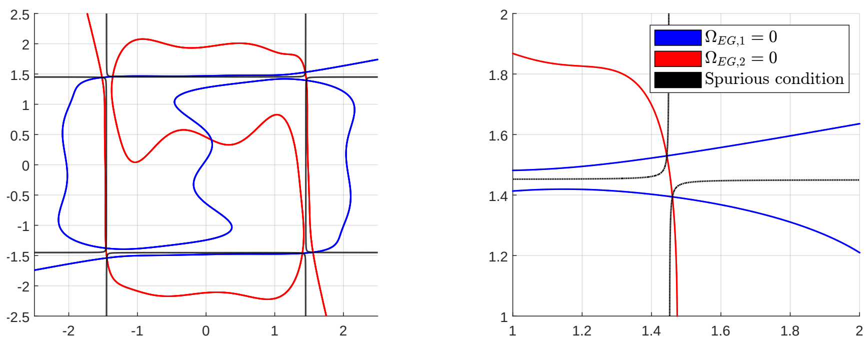
$$\mathcal{H}_{ALG}(f(\boldsymbol{z})) \equiv \mathcal{T}(f(\boldsymbol{z}))\,, \tag{4}$$

and the corresponding HRDE admits no spurious equilibria.

A criterion for assessing the local stability of dynamics equilibria is provided in Appendix D.1.

3.1 Numerical Examples

We illustrate the emergence of spurious equilibria by applying the EG-HRDE to the so-called *forsaken problem* (Hsieh et al., 2021). Figure 2 displays three curves in the (x, y) plane: the nullclines $\Omega_{EG,1}(\boldsymbol{z}) = 0$ and $\Omega_{EG,2}(\boldsymbol{z}) = 0$ corresponding to the first and second components of the algorithmic drift, together with the locus of points $\boldsymbol{z}$ satisfying the spuriousness condition equation 3. Since the nullclines are defined on the $\boldsymbol{\omega}_{1,2} = 0$ hyperplane, their intersection points correspond to DEs by Lemma 3. The PE near the origin is revealed by the intersection of the nullclines away from the spuriousness curve. In contrast, several additional intersections occur at points where the spuriousness condition holds, giving rise to SEs induced by the nonlinear correction terms of the extragradient dynamics. In this example, eight such SEs appear. Additional discussion of this example, along with further numerical illustrations, is provided in Appendix D.2, D.3, and D.4.



(a) EG-HRDE dynamics. (b) Zoomed view highlighting two SEs.

Figure 2: **EG-HRDE on Forsaken problem:** (x, y) are the players' states. The curves $\Omega_{EG,1}(\boldsymbol{z}) = 0$ and $\Omega_{EG,2}(\boldsymbol{z}) = 0$ are shown in red and blue, resp.; spuriousness condition equation 3 is shown in black. Intersections of the drift nullclines correspond to DEs; those lying on the black curve are SEs, while the remaining intersection corresponds to a PEs.

4 Discussion

We study the origin of spurious equilibria (SEs) in HRDEs and derive criteria to characterize equilibrium points and their local stability, supported by numerical examples. Our main message is that *equilibria of HRDEs need not coincide with solutions of the underlying game*. Importantly, our analysis is problem-agnostic: it does not rely on convexity, monotonicity, or related structural assumptions on the underlying operator. Instead, the phenomenon arises from the form of the algorithm-induced dynamics and can occur generically due to nonlinear correction terms. This mismatch is not specific to HRDEs, but applies more broadly to continuous-time models that depart from standard gradient descent, including ODEs that are explicitly designed and later discretized. SEs alter the invariant structure of the dynamics, affecting limit cycles and other nontrivial behaviors such as homoclinic and heteroclinic orbits. Although SEs are often unstable in our examples, their existence underscores a fundamental distinction between the well-posedness of continuous-time trajectories and the fidelity of their equilibrium sets, with implications for the design and interpretation of learning dynamics in games.

ACKNOWLEDGEMENTS

TC is currently supported by the Vienna Science and Technology Fund (WWTF) [Grant ID: 10.47379/VRG25019] and earlier by the FAIR (Future Artificial Intelligence Research) project, funded by the NextGenerationEU program within the PNRR-PE-AI scheme (M4C2, Investment 1.3, Line on Artificial Intelligence).

REFERENCES

Leonard Adolphs, Hadi Daneshmand, Aurelien Lucchi, and Thomas Hofmann. Local saddle point optimization: A curvature exploitation approach. *arXiv:1805.05751*, 2018.

F. Alvarez, H. Attouch, J. Bolte, and P. Redont. A second-order gradient-like dissipative dynamical system with hessian-driven damping.: Application to optimization and mechanics. *Journal de Mathématiques Pures et Appliquées*, 81(8):747–779, 2002. ISSN 0021-7824.

Felipe Alvarez. On the minimizing property of a second order dissipative system in hilbert spaces. *SIAM Journal on Control and Optimization*, 38(4):1102–1119, 2000. doi: 10.1137/S0363012998335802.

Kenneth Arrow and Leonid Hurwicz. Gradient methods for constrained maxima. *Oper. Res.*, pp. 258–265, 1957. ISSN 0030-364X. doi: 10.1287/opre.5.2.258.

Hedy Attouch, Paul-Emile Maingé, and Patrick Redont. A second-order differential system with Hessian-driven damping; application to non-elastic shock laws. *Differential Equations and Applications*, 4(1):27–65, 2012.

Hedy Attouch, Juan Peypouquet, and Patrick Redont. Fast convex optimization via inertial dynamics with Hessian driven damping. *Journal of Differential Equations*, 261(10):5734–5783, November 2016. doi: 10.1016/j.jde.2016.08.020.

Hedy Attouch, Zaki Chbani, Jalal Fadili, and Hassan Riahi. First-order optimization algorithms via inertial systems with Hessian driven damping. *arXiv:1907.10536*, 2019.

David Balduzzi, Sebastien Racaniere, James Martens, Jakob Foerster, Karl Tuyls, and Thore Graepel. The mechanics of n-player differentiable games. In *ICML*, 2018.

Heinz H. Bauschke and Patrick L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, 2nd edition, 2017. ISBN 1441994661.

Tatjana Chavdarova, Gauthier Gidel, François Fleuret, and Simon Lacoste-Julien. Reducing noise in GAN training with variance reduced extragradient. In *NeurIPS*, 2019.

Tatjana Chavdarova, Matteo Pagliardini, Sebastian U Stich, François Fleuret, and Martin Jaggi. Taming GANs with Lookahead-Minmax. In *ICLR*, 2021.

Tatjana Chavdarova, Michael I. Jordan, and Manolis Zampetakis. Last-iterate convergence of saddle point optimizers via high-resolution differential equations. In *Minimax Theory and its Applications*, 2023.

Constantinos Daskalakis and Ioannis Panageas. The limit points of (optimistic) gradient descent in min-max optimization. In *NeurIPS*, pp. 9236–9246, 2018.

Constantinos Daskalakis, Andrew Ilyas, Vasilis Syrgkanis, and Haoyang Zeng. Training GANs with optimism. In *ICLR*, 2018.

Francisco Facchinei and Jong-Shi Pang. *Finite-dimensional Variational Inequalities and Complementarity Problems*. Springer, 2003.

Noah Golowich, Sarath Pattathil, Constantinos Daskalakis, and Asuman Ozdaglar. Last iterate is slower than averaged iterate in smooth convex-concave saddle point problems. In Jacob Abernethy and Shivani Agarwal (eds.), *Proceedings of Thirty Third Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pp. 1758–1784, 2020.

Uwe Helmke and John B. Moore. *Optimization and Dynamical Systems*. 1996.

Reyhane Askari Hemmat, Amartya Mitra, Guillaume Lajoie, and Ioannis Mitliagkas. Lead: Least-action dynamics for min-max optimization. *arXiv:2010.13846*, 2020.

Josef Hofbauer and William H. Sandholm. Stable games and their dynamics. *Journal of Economic Theory*, 144(4):1665–1693, 2009. doi: 10.1016/j.jet.2009.01.007.

Ya-Ping Hsieh, Panayotis Mertikopoulos, and Volkan Cevher. The limits of min-max optimization algorithms: convergence to spurious non-critical sets. In *ICML*, 2021.

Galina Michailovna Korpelevich. The extragradient method for finding saddle points and other problems. *Matecon*, 1976.

Eric Mazumdar, Lillian J. Ratliff, and S. Shankar Sastry. On gradient-based learning in continuous games. *SIAM Journal on Mathematics of Data Science*, 2(1):103–131, 2020. doi: 10.1137/18M1231298.

Panayotis Mertikopoulos and William H. Sandholm. Learning in games via reinforcement and regularization. *Mathematics of Operations Research*, 41(4):1297–1324, November 2016. ISSN 0364-765X.

Panayotis Mertikopoulos, Christos H. Papadimitriou, and Georgios Piliouras. Cycles in adversarial regularized learning. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2018.

Lars Mescheder, Andreas Geiger, and Sebastian Nowozin. Which training methods for GANs do actually converge? In *International Conference on Machine Learning*, pp. 3481–3490, 2018.

Michael. Muehlebach and Michael I. Jordan. A dynamical systems perspective on nesterov acceleration. In *ICML*, 2019.

Boris T Polyak. Some methods of speeding up the convergence of iteration methods. *USSR Computational Mathematics and Mathematical Physics*, 1964.

Leonid Denisovich Popov. A modification of the Arrow–Hurwicz method for search of saddle points. *Mathematical Notes of the Academy of Sciences of the USSR, 28(5):845–848*, 1980.

Aniket Sanyal and Tatjana Chavdarova. Understanding lookahead dynamics through laplace transforms. In *High-dimensional Learning Dynamics (NeurIPS Workshop)*, 2025.

Aniket Sanyal, Baraah A. M. Sidahmed, Rebekka Burkholz, and Tatjana Chavdarova. Frequency-based hyperparameter selection in games. In *AISTATS*, 2026.

J. Schropp and I. Singer. A dynamical systems approach to constrained minimization. *Numerical Functional Analysis and Optimization*, 21(3-4):537–551, 2000. doi: 10.1080/01630560008816971.

Bin Shi, Simon S. Du, Michael I. Jordan, and Weijie J. Su. Understanding the acceleration phenomenon via high-resolution differential equations. *arXiv:1810.08907*, 2018.

Weijie Su, Stephen Boyd, and Emmanuel J. Candès. A differential equation for modeling Nesterov's accelerated gradient method: Theory and insights. *Journal of Machine Learning Research*, 17 (153):1–43, 2016.

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

Michael Zhang, James Lucas, Jimmy Ba, and Geoffrey E Hinton. Lookahead optimizer: k steps forward, 1 step back. In *NeurIPS*, 2019.

A RELATED WORKS

Differential equations for minimization and monotone operator problems. Continuous-time viewpoints for optimization go back at least to the interpretation of iterative schemes as dynamical systems and to the use of inertial / second-order dynamics (e.g., heavy-ball–type limits) (Polyak, 1964; Arrow & Hurwicz, 1957; Helmke & Moore, 1996). A prominent line of work connects accelerated first-order methods to ODE limits and uses these limits to derive rates, qualitative behavior, and Lyapunov functions that can sometimes be transferred back to discrete time; a canonical example is the ODE perspective on Nesterov-type acceleration (Su et al., 2016). Related developments study second-order dissipative systems and their discretizations, including variants with geometric damping and monotone operator structure (Schropp & Singer, 2000; Attouch et al., 2012; Alvarez et al., 2002; Attouch et al., 2019; Muehlebach & Jordan, 2019). These works primarily target single-objective minimization (or monotone inclusions) where the vector field has (or can be reduced to) a gradient-like structure.

Differential equations for games, variational inequalities, and min–max dynamics. For saddle-point / game dynamics, naive continuous-time limits can mask algorithmic differences: several first-order methods for min–max (e.g., gradient descent, extragradient (Korpelevich, 1976), optimistic gradient descent (Popov, 1980)) share the same first-order ODE limit, despite sharply different discrete-time behavior (Hsieh et al., 2021; Chavdarova et al., 2023). This motivates *high-resolution* limits that retain higher-order information in the step size and can distinguish algorithms beyond the standard ODE approximation (Shi et al., 2018; Chavdarova et al., 2023). In parallel, a large body of work documents nonconvergence phenomena (cycles/limit cycles, sensitivity, divergence) for training games and min-max objectives and proposes stabilizing modifications (Daskalakis et al., 2018; Mescheder et al., 2018; Daskalakis & Panageas, 2018; Mazumdar et al., 2020; Mertikopoulos et al., 2018; Adolphs et al., 2018; Chavdarova et al., 2019). Recent theoretical results also show that for broad classes of problems, popular min–max methods can be attracted to spurious cycles, highlighting that the continuous-time viewpoint must be chosen carefully to remain faithful to the discrete algorithm and/or the underlying VI solution concept (Hsieh et al., 2021; Golowich et al., 2020; Chavdarova et al., 2023).

Solution sets and equilibrium fidelity. For classical inertial gradient dynamics arising in optimization, equilibria of the continuous-time system coincide with minimizers of the objective function under suitable convexity or geometric assumptions; see, e.g., Alvarez (2000); Attouch et al. (2016, and references therein). In this setting, the equilibrium structure of the ODE faithfully reflects the solution set of the underlying problem. In contrast, we show that high-resolution differential equations (HRDEs) associated with extragradient methods may admit additional equilibria that do not correspond to solutions of the underlying variational inequality. This phenomenon highlights a mismatch between algorithm-induced continuous-time equilibria and problem-defined solutions.

Related HRDE works focus on well-posedness, stability, and convergence properties of trajectories. For instance, Chavdarova et al. (2023, Proposition 1) establish uniqueness of trajectories under monotonicity and smoothness assumptions, but do not examine whether the equilibria of the HRDE coincide with VI solutions. Similarly, Sanyal & Chavdarova (2025); Sanyal et al. (2026) analyze stability and convergence of HRDE dynamics while implicitly assuming that algorithmic equilibria faithfully represent problem solutions.

Observations related to algorithm-induced solution mismatch also appear in the literature on operator splitting methods (e.g., Forward–Backward, Douglas–Rachford, and Peaceman–Rachford schemes). In these settings, it is well known that the fixed-point set of the splitting operator may differ from the solution set of the original monotone inclusion, necessitating additional recovery or projection mechanisms (Bauschke & Combettes, 2017).

Finally, we emphasize that our notion of a "true" solution is defined with respect to the standard variational inequality formulation, characterized by first-order optimality conditions. In a related but distinct line of work, several studies in learning dynamics and game theory point out that gradient-based dynamics may admit stationary points that fail to satisfy Nash equilibrium conditions, even when such points are dynamically stable (Hofbauer & Sandholm, 2009; Mertikopoulos & Sandholm, 2016). These observations further illustrate that equilibrium concepts induced by dynamics need not coincide with solution concepts defined by the underlying problem.

B Additional Preliminaries

B.1 Variational Inequality Viewpoint of equation SP

We can rewrite the zero-sum equation SP game using the operator V associated with f as the following Variational Inequality (Facchinei & Pang, 2003, VI) problem:

$$\textit{find } \boldsymbol{z}^\star \quad \text{s.t.} \quad \langle \boldsymbol{z} - \boldsymbol{z}^\star, V(\boldsymbol{z}) \rangle \geq 0, \quad \forall \boldsymbol{z} \in \mathbb{R}^d . \tag{VI}$$

When the domain is unconstrained, the above reduces simply to finding $\boldsymbol{z}^\star$ such that $V(\boldsymbol{z}^\star) = 0$.

B.2 Primer on HRDEs: Time rescaling and velocity variables

To derive a continuous-time model that captures higher-order effects of the discrete dynamics, HRDEs introduce an auxiliary velocity variable defined by

$$\omega_k := \frac{z_k - z_{k-1}}{\gamma}.$$

Interpreting the iterates as samples of a smooth trajectory $z(t)$ at times $t_k = k\gamma$, we formally identify

$$\omega_k \approx \dot{z}(t_k).$$

Using Taylor expansions of $z(t+\gamma)$ around t and matching terms up to second order in γ, the discrete-time recursion equation G-Disc yields a second-order differential equation governing the evolution of $(z(t), \omega(t))$.

B.3 Additional Methods and Their HRDEs

Discrete saddle-point methods. We briefly review several widely used first-order methods for saddle-point and variational inequality problems.

Gradient Descent–Ascent (GDA). A direct extension of gradient descent to saddle-point optimization is given by

$$\boldsymbol{z}_{n+1} = \boldsymbol{z}_n - \gamma V(\boldsymbol{z}_n), \tag{GDA}$$

where $\gamma \in [0,1]$ is a step size. In the context of zero-sum games, the operator V takes the form $(-\nabla_x f, \nabla_y f)$, so the update performs a descent step in the primal variable x and an ascent step in the dual variable y. This method is therefore referred to as *gradient descent–ascent* (GDA).

Extragradient (EG). The *extragradient* method augments GDA with an intermediate prediction step. Starting from $\boldsymbol{z}_n$, an extrapolated point

$$\boldsymbol{z}_{n+\frac{1}{2}} = \boldsymbol{z}_n - \gamma V(\boldsymbol{z}_n)$$

is first computed, and the update is then performed using the gradient evaluated at this extrapolated point:

$$\boldsymbol{z}_{n+1} = \boldsymbol{z}_n - \gamma V\big(\boldsymbol{z}_n - \gamma V(\boldsymbol{z}_n)\big). \tag{EG}$$

The extragradient method was introduced by Korpelevich (1976), who established its convergence for monotone (constrained) variational inequalities with Lipschitz operators. Subsequent work extended these results to convex–concave saddle-point problems over closed convex sets (Facchinei & Pang, 2003).

Optimistic Gradient Descent–Ascent (OGDA). The *optimistic gradient descent–ascent* (OGDA) method, originally proposed by Popov (1980), incorporates information from the previous iterate:

$$\boldsymbol{z}_{n+1} = \boldsymbol{z}_n - 2\gamma V(\boldsymbol{z}_n) + \gamma V(\boldsymbol{z}_{n-1}). \tag{OGDA}$$

By leveraging a simple extrapolation of past gradients, OGDA has been shown to improve stability and convergence behavior in several saddle-point settings.

Lookahead–Minmax (LA). Building on earlier work in optimization (Zhang et al., 2019), Chavdarova et al. (2021) proposed the *Lookahead–Minmax* (LA) algorithm for min–max optimization. At each iteration n, the method proceeds as follows: (i) a copy of the current iterate is formed,

$\tilde{\boldsymbol{z}}_n \leftarrow \boldsymbol{z}_n$; (ii) the copy $\tilde{\boldsymbol{z}}_n$ is updated $k \geq 1$ times using a chosen base optimizer, yielding $\tilde{\boldsymbol{z}}_{n+k}$; (iii) the next iterate is obtained by interpolating between the current point and the prediction:

$$\boldsymbol{z}_{n+1} = \boldsymbol{z}_n + \alpha(\tilde{\boldsymbol{z}}_{n+k} - \boldsymbol{z}_n), \quad \alpha \in [0,1]. \tag{LA}$$

Typically, GDA equation GDA is used as the base optimizer, denoted as *LA*k*–GDA*.

HRDEs. The HRDE of the Lookahead (Chavdarova et al., 2021) method with parametric (α, k) is the following:

$$\begin{cases} \dot{\boldsymbol{z}}(t) = \boldsymbol{\omega}(t) \\ \dot{\boldsymbol{\omega}}(t) = -\frac{2}{\gamma}\boldsymbol{\omega} - \frac{2k\alpha}{\gamma}V(\boldsymbol{z}(t)) + (\sum_{i=1}^{k-1} i)2\alpha J(\boldsymbol{z}(t))V(\boldsymbol{z}(t))\,. \end{cases} \tag{LA-GD-HRDE}$$

For generalizing purposes, we will rewrite the sum $\sum_{i=1}^{k-1} i$ with $k \in \mathbb{N}$ as its real valued equivalent $\frac{(k-1)k}{2}$ with $k \in \mathbb{R}$:

$$\begin{cases} \dot{\boldsymbol{z}}(t) = \boldsymbol{\omega}(t) \\ \dot{\boldsymbol{\omega}}(t) = -\frac{2}{\gamma}\boldsymbol{\omega} - \frac{2k\alpha}{\gamma}V(\boldsymbol{z}(t)) + (k-1)k\alpha J(\boldsymbol{z}(t))V(\boldsymbol{z}(t))\,. \end{cases} \tag{LAk-GD-HRDE}$$

C PROOF OF EXTRAGRADIENT

By considering EG-HRDE and by setting $\dot{\boldsymbol{\xi}} = 0$ we have that:

$$\begin{cases} 0 = \boldsymbol{\omega}^\star, \\ 0 = -\frac{2}{\gamma}\,\boldsymbol{\omega}^\star(t) - \frac{2}{\gamma}V(\boldsymbol{z}^\star) + 2J(\boldsymbol{z}^\star)V(\boldsymbol{z}^\star) \end{cases} \tag{5}$$

If we consider $d = 2$, we can simply derive the following:

$$\begin{aligned} &-V(\boldsymbol{z}^\star) + \gamma J(\boldsymbol{z}^\star)V(\boldsymbol{z}^\star) = 0, \\ &\big(I^{[2\times 2]} - \gamma J(\boldsymbol{z}^\star)\big)V(\boldsymbol{z}^\star) = 0, \\ &\big(I^{[2\times 2]} - \gamma J(\boldsymbol{z}^\star)\big)V(\boldsymbol{z}^\star) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \\ &\begin{bmatrix} 1 - \gamma J_{11}(\boldsymbol{z}^\star) & -\gamma J_{12}(\boldsymbol{z}^\star) \\ -\gamma J_{21}(\boldsymbol{z}^\star) & 1 - \gamma J_{22}(\boldsymbol{z}^\star) \end{bmatrix} \begin{bmatrix} V_1(\boldsymbol{z}^\star) \\ V_2(\boldsymbol{z}^\star) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \end{aligned} \tag{6}$$

From which system (3) is derived:

$$\begin{cases} J_{11}(\boldsymbol{z}^\star)V_1(\boldsymbol{z}^\star) + J_{12}(\boldsymbol{z}^\star)V_2(\boldsymbol{z}^\star) = V_1(\boldsymbol{z}^\star)/\gamma\,, \\ J_{21}(\boldsymbol{z}^\star)V_1(\boldsymbol{z}^\star) + J_{22}(\boldsymbol{z}^\star)V_2(\boldsymbol{z}^\star) = V_2(\boldsymbol{z}^\star)/\gamma\,. \end{cases} \tag{7}$$

Not that the assumption of $d = 2$ was made just to simplify the notation. These procedures can be repeated for any d.

C.1 INTERPRETATION OF EQUATION 2

Any solution of system derived in equation 2 is the set of $\boldsymbol{z} \in \mathcal{H}_{EG}(f(\boldsymbol{z}))$. Any solution of such a system is in fact a point that nullify the algorithmic drift, and, consequently correspond to a dynamics equilibrium (DE) $\boldsymbol{\xi} = \begin{bmatrix} \boldsymbol{z} \\ \boldsymbol{0} \end{bmatrix}$. It is immediate to notice that any problem equilibrium also solves equation 7, as it is clear that it could exist a certain value of $\boldsymbol{z}$ such that V is non-zero while still guaranteeing a null algorithmic drift. This latter solution is evidently created by the nonlinear interaction between V and J. The existence of spurious equilibria (SEs) is a matter of interest because of two main reasons: first of all, the existence of SEs may alter the fundamental dynamics of the explored dynamical systems, and may even affect the behavior of other invariants like limit cycles. The second reason for which the study of SEs is of interest is simply because if, for some parametric condition, a SE happen to be locally attractive, the solver algorithm could converge to a non existing solution.

C.2 PROOF OF CLAIM 2

Consider system (7), impose $d = 2$ and set:
$V(\boldsymbol{z}^\star) \neq 0,\ J_{11}(\boldsymbol{z}^\star) \neq \frac{1}{\gamma},\ J_{22}(\boldsymbol{z}^\star) \neq \frac{1}{\gamma},\ J_{12}(\boldsymbol{z}^\star) \neq 0,\ J_{21}(\boldsymbol{z}^\star) \neq 0$. It is obtained:

$$\begin{cases} \frac{V_2(\boldsymbol{z}^\star)}{V_1(\boldsymbol{z}^\star)} = \frac{1-\gamma J_{11}(\boldsymbol{z}^\star)}{\gamma J_{12}(\boldsymbol{z}^\star)}, \\ \frac{V_2(\boldsymbol{z}^\star)}{V_1(\boldsymbol{z}^\star)} = \frac{\gamma J_{21}(\boldsymbol{z}^\star)}{1-\gamma J_{22}(\boldsymbol{z}^\star)}. \end{cases}$$

$$\frac{1-\gamma J_{11}(\boldsymbol{z}^\star)}{\gamma J_{12}(\boldsymbol{z}^\star)} = \frac{\gamma J_{21}(\boldsymbol{z}^\star)}{1-\gamma J_{22}(\boldsymbol{z}^\star)}. \tag{8}$$

$$1 - \gamma(J_{11}(\boldsymbol{z}^\star) + J_{22}(\boldsymbol{z}^\star)) + \gamma^2(J_{11}(\boldsymbol{z}^\star)J_{22}(\boldsymbol{z}^\star) - J_{12}(\boldsymbol{z}^\star)J_{21}(\boldsymbol{z}^\star)) = 0.$$

$$1 - \gamma \mathrm{tr}(J(\boldsymbol{z}^\star)) + \gamma^2 \det(J(\boldsymbol{z}^\star)) = 0. \tag{9}$$

This last equation (9), which coincide with condition (3), holds only for two dimensional systems and enunciate a necessary condition for a solution to be spurious. Any $\boldsymbol{z}$ that solves (9) is such that the algorithmic drift terms are equal i.e. $\Omega_1(\boldsymbol{z}) = \Omega_2(\boldsymbol{z})$ where $\Omega(\boldsymbol{z}) = [\Omega_1(\boldsymbol{z})\ \Omega_2(\boldsymbol{z})]^T$ and at the same time make $V(\boldsymbol{z}) \neq 0$. Any solution point is such that $\Omega(\boldsymbol{z}) = 0$ (property 2), therefore, if a solution point also solves condition (9) then it must be spurious. If instead a solution point does not solves condition (9) then it is true.

D OMITTED RESULTS

For convenience, Table 1 summarizes the solution notation used in the main part.

Table 1: Solution notation summary.

Point	**Meaning**	**Associated set**
$\xi^\star \in \mathbb{R}^{2d} = (\boldsymbol{z}^\star, \omega^\star)$	dynamical system solution	$\mathcal{H}_{ALG}(f(z)) = \{\boldsymbol{z}_1^\star, \dots, \boldsymbol{z}_n^\star\}$
$\bar{z} \in \mathbb{R}^d$	VI solution, where $V(\bar{z}) = 0$	$\mathcal{T}(f(z)) = \{\bar{z}_1, ..., \bar{z}_n\}$

D.1 STABILITY OF SOLUTIONS

When we refer to *stability of a solution* we refers to the behavior of the algorithm dynamics at least in a neighborhood of it. Since a solution is just an equilibrium point of a continuous time dynamical system, here we use the Lyapunov stability definition and all the related properties when addressing the solution stability. First, we study the stability of the elements of $\mathcal{H}_{ALG}$ for any G-HRDE; consider the linearized system state matrix of G-HRDE:

$$A_{ALG}(\boldsymbol{\xi}) = \begin{bmatrix} 0^{[d\times d]} & I^{[d\times d]} \\ \frac{\partial \Omega_{ALG}(\boldsymbol{z})}{\partial z} & -\frac{2}{\gamma} I^{[d\times d]} \end{bmatrix} = \begin{bmatrix} 0^{[d\times d]} & I^{[d\times d]} \\ L_{ALG}(\boldsymbol{z}) & -\beta I^{[d\times d]} \end{bmatrix}. \tag{10}$$

Given an equilibrium point $\xi^\star$, if every $\Re\{\mathrm{eig}(A_{ALG}(\boldsymbol{\xi}^\star))\} < 0$ then $\boldsymbol{z}^\star$ is locally asymptotically stable. If such a condition is not respected, then the convergence to the solution is impossible as the solution itself is unstable.
The matrix $L_{ALG}(\boldsymbol{z})$ is any arbitrary matrix in $\mathbb{R}^{[d\times d]}$ which changes for a chosen algorithm ALG and correspond to the Jacobian of the algorithmic drift $\Omega_{ALG}(z)$. Here we show that the stability of each dynamics equilibrium (DE) can be inferred by studying matrix $L_{ALG}(\boldsymbol{z})$ only.

Claim 4. ***Asymptotic stability of G-HRDE*** *Every eigenvalue λ_i of matrix $A_{ALG}(\boldsymbol{\xi}) \in \mathbb{R}^{[2d\times 2d]}$ have a negative real part if and only if every eigenvalue μ_j of matrix $L_{ALG}(\boldsymbol{z}) \in \mathbb{R}^{[d\times d]}$ is such that*

$$\Re\left\{\sqrt{\beta^2 + 4\mu_j}\right\} < \beta,\ \forall j = 1, ..., d.$$

Proof. Consider a block matrix:

$$\Gamma = \begin{bmatrix} A & B \\ C & D \end{bmatrix}, \tag{11}$$

if each block can commute, i.e. $AB = BA,\ AC = CA,\ AD = DA,\ \dots$, then:

$$\det(\Gamma) = \det(AD - BC). \tag{12}$$

To compute the eigenvalues of matrix $A_{ALG}(\xi)$ we have:

$$\text{eig}(A_{ALG}(\boldsymbol{\xi})) = \det(\lambda I - A_{ALG}(\boldsymbol{\xi})) = \det\left(\begin{bmatrix} \lambda I^{[d\times d]} & -I^{[d\times d]} \\ -L_{ALG}(\boldsymbol{z}) & (\lambda + \frac{2}{\gamma})I^{[d\times d]} \end{bmatrix}\right). \tag{13}$$

Matrix (13) is composed of four $[d \times d]$ blocks that can commute, therefore:

$$\det\left(\begin{bmatrix} \lambda I^{[d\times d]} & -I^{[d\times d]} \\ -L_{ALG}(\boldsymbol{z}) & (\lambda + \frac{2}{\gamma})I^{[d\times d]} \end{bmatrix}\right) = \det\left(\lambda(\lambda + \frac{2}{\gamma})I^{[d\times d]} - L_{ALG}(\boldsymbol{z})\right) = \chi(\lambda). \tag{14}$$

Let now μ be an eigenvalue of $L_{ALG}(\boldsymbol{z})$; μ is such that

$$\det(\mu - L_{ALG}(\boldsymbol{\xi})) = 0. \tag{15}$$

By comparing (15) and (14) we find that each eigenvalue of $A_{ALG}(\xi)$ solves:

$$\lambda(\lambda + \beta) - \mu_j = 0,\ \forall j = 1, .., d. \tag{16}$$

Where $\beta = \frac{2}{\gamma}$. The $2 \times d$ eigenvalues of $A_{ALG}(\boldsymbol{\xi})$ are therefore the solutions of:

$$\lambda_i = \frac{-\beta \pm \sqrt{\beta^2 + 4\mu_j}}{2},\ \forall j = 1, ..., d. \tag{17}$$

Each pair of eigenvalues of $A_{ALG}(\xi)$ correspond to a solution of (17) for each j. If $\mu_j \in \mathbb{C}$, to ensure stability, we must impose that the largest λ_i for each μ_j is negative:

$$\Re\left\{\sqrt{\beta^2 + 4\mu_j}\right\} < \beta,\ \forall j = 1, ..., d. \tag{18}$$

□

D.2 Discussion: EG-HRDE on forsaken problem

The forsaken problem is well known in literature for having one (locally attractive) solution located close to the origin and two limit cycles around it. The inner limit cycle is unstable while the outer one is stable, resulting in most of the initializations to converge to the outer limit cycle instead of the solution point. In figure 2a the location of the problem equilibrium is made evident as it coincide with the intersection of the two blue and red curves near the origin, however, eight(!) different dynamics equilibrium (DEs) appears in other regions of space. All these solution points are spurious as they belong to the locus of point of condition (3) (highlighted in Figure 2b), meaning that in those points it holds: $\dot{\boldsymbol{\xi}} = 0\ \wedge\ V(\boldsymbol{z}) \neq 0$. A further notice is that the location of the SEs is function of the step size γ; even if in this example all the SEs are found to be *unstable* i.e. not locally attractive, this property may change for a different value of γ which has here been set to $\gamma = 0.1$.

D.3 Example: bilinear game

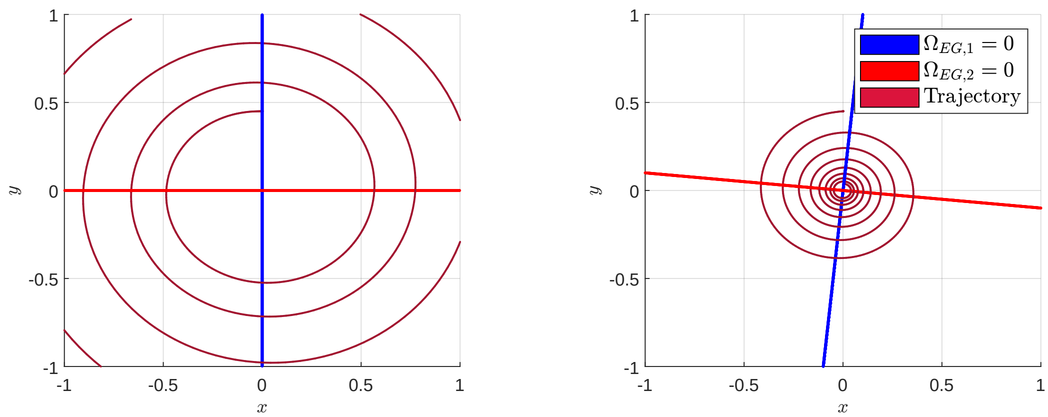
In figure 3 an example is given for the **bilinear game** case, where the GD-HRDE and the EG-HRDE are compared. The red and blue lines are the nullclines of the algorithmic drifts computed for both algorithm and, at the origin, the only problem equilibrium is found. For the GD-HRDE case, the L matrix is always equal to:

$$L_{GDA} = \begin{bmatrix} 0 & -\beta \\ \beta & 0 \end{bmatrix}.\ \mu_{1,2} = 0 \pm i\beta. \tag{19}$$

Never admitting negative real part eigenvalues. For the EG-HRDE case instead, the L matrix is always equal to:

$$L_{EG} = \begin{bmatrix} -2 & -\beta \\ \beta & -2 \end{bmatrix}.\ \mu_{1,2} = -2 \pm i\beta. \tag{20}$$

Implying that every equilibrium point for the bilinear game solved with EG-HRDE is attractive; being the one located in the origin the only equilibrium point it can be concluded that the EG-HRDE always converges to the origin.



(a) Bilinear problem with GD-HRDE.

(b) Bilinear problem with EG-HRDE.

Figure 3: Comparison of the high resolution differential equations of GD and EG dynamics for the bilinear problem.

D.4 Extension for quadratic games

Here we define a quadratic game any problem of the form:

$$\begin{aligned} f(x,y) &= xy + ax^2 - by^2. \\ V(\boldsymbol{z}) &= \begin{bmatrix} y + 2ax \\ -x + 2by \end{bmatrix}. \\ J(\boldsymbol{z}) &= \begin{bmatrix} 2a & 1 \\ -1 & 2b \end{bmatrix}. \end{aligned} \tag{QG}$$

Where $a, b \in \Re^+$. Notice how the bilinear game is a particular case of the quadratic one. We will now show the following:

Claim 5. ***Non-spuriousness of quadratic games***
Any quadratic game (QG) solved with GD-HRDE, EG-HRDE or LAk-GD-HRDE does admit one unique dynamics equilibrium (DE) which stability can be inferred by criterion (18). The equilibrium is the problem equilibrium (PE).

Proof. To show this, we just need to compute the algorithmic drift $\Omega^{QG}_{ALG}(\boldsymbol{z})$ for the considered game and ALG. If the algorithmic drift is linear in $\boldsymbol{z}$, then due to claim 3 any solution of $\Omega^{QG}_{ALG}(z) = 0$ is a problem equilibrium (PE). If the solution is unique, the claim is proven. First we compute $\Omega^{QG}_{ALG}(z)$ for EG-HRDE and LAk-GD-HRDE

$$\Omega^{QG}_{EG}(\boldsymbol{z}) = \begin{bmatrix} (y + 2ax)(4a - \beta) + 2(-x + 2by) \\ (-x + 2by)(4b - \beta) + 2(-y - 2ax) \end{bmatrix}. \tag{21}$$

$$\Omega^{QG}_{LAk}(\boldsymbol{z}) = k\alpha \begin{bmatrix} (y + 2ax)((k-1)2a - \beta) + (k-1)(-x + 2by) \\ (-x + 2by)((k-1)2b - \beta) + (k-1)(-y - 2ax) \end{bmatrix}. \tag{22}$$

In both cases, all the elements of $\Omega^{QG}_{ALG}(\boldsymbol{z})$ are linear in x and y, meaning that the value of z for which

$$\Omega^{QG}_{ALG}(\boldsymbol{z}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \tag{23}$$

is unique unless the two elements of Ω are always identical, which is not true for any value of a and b. The algorithmic drift of GD-HRDE is also linear in $\boldsymbol{z}$ as it equals $-\beta V(\boldsymbol{z})$ and admits one unique solution. It can be easily verified that $\Omega^{QG}_{ALG}(\mathbf{0}) = \mathbf{0}$ for the three cases, meaning that the origin is a problem solution and is the only one since the uniqueness of (23). ☐