Journal of Machine Learning Research 27 (2026) 1-61 Submitted 5/24; Revised 11/25; Published 1/26

A Symplectic Analysis of Alternating Mirror Descent

Jonas Katona JONAS.KATONA@YALE.EDU
Department of Applied & Computational Mathematics
Yale University
New Haven, CT 06511, USA

Xiuyuan Wang XIUYUAN.WANG@YALE.EDU
Department of Computer Science
Yale University
New Haven, CT 06511, USA

Andre Wibisono ANDRE.WIBISONO@YALE.EDU
Department of Computer Science
Yale University
New Haven, CT 06511, USA

Editor: Lam Nguyen

Abstract

Motivated by understanding the behavior of the Alternating Mirror Descent (AMD) algorithm for bilinear zero-sum games, we study the discretization of continuous-time Hamiltonian flow via the symplectic Euler method. We provide a framework for analysis using results from Hamiltonian dynamics and symplectic numerical integrators, with an emphasis on the existence and properties of a conserved quantity, the modified Hamiltonian (MH), for the symplectic Euler method. We compute the MH in closed-form when the original Hamiltonian is a quadratic function, and show that it generally differs from the other conserved quantity known previously in the literature. We derive new error bounds on the MH when truncated at orders in the stepsize in terms of the number of iterations, K, and use these bounds to show an improved $\mathcal{O}(K^{1/5})$ total regret bound and an $\mathcal{O}(K^{-4/5})$ duality gap of the average iterates for AMD. Finally, we propose a conjecture which, if true, would imply that the total regret for AMD scales as $\mathcal{O}\left(K^{\varepsilon}\right)$ and the duality gap of the average iterates as $\mathcal{O}\left(K^{-1+\varepsilon}\right)$ for any $\varepsilon > 0$, and we can take $\varepsilon = 0$ upon certain convergence conditions for the MH.

Keywords: symplectic integrators, numerical integration, Hamiltonian dynamics, algorithmic game theory, min-max optimization, alternating mirror descent

1. Introduction

In its original physics setting, Hamiltonian flow is used to describe the dynamics of physical systems where the total energy—i.e., the Hamiltonian function H—is conserved. Perhaps after some perturbations or upon a change of variables, Hamiltonian flows also appear in other scientific fields as dynamical models of systems, even when these systems do not describe physical processes. Some examples which motivate our work are as follows. In game theory, Hamiltonian structure appears in the dynamics of multi-agent systems (Hofbauer, 1996; Bailey and Piliouras, 2019); this perspective has been useful for understanding emergent phenomena such as chaotic behavior in the systems (Sato et al., 2002; Piliouras and Wang, 2021), and for designing more efficient multi-agent learning algorithms (Balduzzi et al., 2018). In online learning, Hamiltonian structure appears in the continuous-time view of best-response dynamics such as fictitious play (Ostrovski and van Strien, 2011; van Strien, 2011).

Hamiltonian flows also appear in many problems motivated by computer science. In optimization, Hamiltonian dynamics (with friction or damping) appear in the continuous-time view of accelerated gradient methods; this Hamiltonian perspective has been useful for deriving new optimization algorithms and extending the concept of accelerated methods to more general optimization problems, (see, e.g., Su et al., 2016; Wibisono et al., 2016; Betancourt et al., 2018; Wilson et al., 2019; Diakonikolas and Jordan, 2021; Wilson et al., 2021; Duruisseaux and Leok, 2022). A Hamiltonian flow perspective has also been used for deriving new descent methods for optimization (O'Donoghue and Maddison, 2019; Teel et al., 2019; De Luca et al., 2023; Wang, 2024). In sampling or probabilistic applications such as Bayesian inference, Hamiltonian dynamics appear as a crucial component of the Hamiltonian Monte Carlo (HMC) algorithm (Duane et al., 1987; Neal, 2011), which is a practical sampling algorithm with good empirical performance, and which is the default method in probabilistic programming packages such as Stan (Carpenter et al., 2017) and PyMC3 (Salvatier et al., 2016). The theoretical guarantees of HMC are strongly tied to conservation of energy in the discretization scheme used for implementing the Hamiltonian dynamics—see, e.g., Hoffman and Gelman (2014), Betancourt (2016), Bou-Rabee and Sanz-Serna (2018), Mangoubi and Vishnoi (2018), Lee and Vempala (2018), Chen and Vempala (2019), Betancourt (2019), Chen et al. (2020), Lee et al. (2020), and Kook et al. (2022).

Hence, the accurate simulation of Hamiltonian flows is of great importance for applications across game theory, learning theory, and computer science at large. Unless one can explicitly integrate the Hamiltonian flow for the Hamiltonian in question (which is rare), one generally resorts to discretizing the flow by subdividing the time interval and computing iterative approximations to the ODE solution at those times. For these *numerical integrators*, we generally speak of a stepsize $\eta > 0$, such that as $\eta \to 0$, the numerical solution approaches the true solution (Atkinson, 1988).

Since the true Hamiltonian flow preserves H—the Hamiltonian function or "energy"—one might also expect a numerical integrator to conserve a quantity approximating H for sufficiently small η, which we call a *modified Hamiltonian* $\widetilde{H}_\eta$, such that as $\eta \to 0$, the modified Hamiltonian recovers the true Hamiltonian: $\widetilde{H}_\eta \to H$, in some notion of convergence. Analogous to how knowing H and its level sets are useful for geometric analyses of Hamiltonian flow, knowing $\widetilde{H}_\eta$ and its level sets would be similarly useful for analyzing the numerical integrator and its trajectories. However, a modified Hamiltonian $\widetilde{H}_\eta$ does not exist for many discrete-time approximations of Hamiltonian flow (Hairer et al., 2006, Chapter IX)—two counterexamples are given later in Section 3. Thus, guaranteeing that the numerical solution matches the behavior of the continuous-time dynamics can be difficult and subtle, especially when trying to develop such guarantees rigorously for algorithmic applications.

Motivated by the above problem, this paper studies the following two questions: How can one implement Hamiltonian flows as discrete-time algorithms with provable energy conservation errors, and how do these energy conservation errors inform the performance of related algorithms in a computer science context? Symplectic integration is the study of how to accurately discretize Hamiltonian flows as discrete-time algorithms and is a long-studied field with many classical results that are designed precisely to answer our first question (e.g., Channell and Scovel, 1990; Yoshida, 1990, 1993; Hairer et al., 2006). However, existing results are not sufficiently algorithmic for many applications of interest in algorithmic game theory and computer science. For instance, while all symplectic discretizations *in principle* conserve a modified Hamiltonian $\widetilde{H}_\eta$, unfortunately, $\widetilde{H}_\eta$ can only be described as a formal power series which generally does not converge (Field and Nijhoff, 2003). Furthermore, since the series for $\widetilde{H}_\eta$ can only be summed up to a finite number of terms in practice, it remains unanswered as to how using the truncated forms of the series for $\widetilde{H}_\eta$ induces an error in their conservation from an algorithmic perspective without assuming a priori that the numerical solution remains within a compact set (Hairer et al., 2006, Chapter IX).

This study also sets the ground for a framework of a rigorous mathematical machinery that can be used for deriving algorithmic guarantees and new algorithms in game theory, learning theory, and computer science applications which relate to Hamiltonian dynamics. One particular example comes from algorithmic game theory. The joint behavior of two agents in a bilinear zero-sum game using greedy strategies in continuous time can be described via a Hamiltonian flow, and different discretization methods of the Hamiltonian flow correspond to different strategies for the two players in the game (Hofbauer, 1996; Bailey and Piliouras, 2019). Out of these strategies, we focus on the *Alternating Mirror Descent* (AMD) algorithm for constrained zero-sum games (Wibisono et al., 2022), which we show is related by duality to the discretization of a Hamiltonian flow that preserves a modified Hamiltonian $\widetilde{H}_\eta$. The conservation of $\widetilde{H}_\eta$ and the near-conservation of its truncated forms have direct implications on regret guarantees of the AMD algorithm, which we describe further in Section 9. We use several key properties of $\widetilde{H}_\eta$ and its order-by-order truncations in η to generalize both the analyses done in Bailey et al. (2020) for the unconstrained setting of AMD with quadratic regularizers and the regret analysis of AMD in Wibisono et al. (2022).

Contributions. We address the questions above and set the framework for relating algorithms in game theory to symplectic integrators by showing the following:

1. Under a computable linear transformation and a Legendre transform, for any choice of Legendre-type regularizers in the bilinear zero-sum game setting and usual domain assumptions, the trajectories of the symplectic Euler method applied to a Hamiltonian system with a suitable Hamiltonian function transform into the trajectories of AMD. Thus, AMD inherits the properties which symplectic Euler has: A conserved quantity expressible via order-by-order corrections in η and a continuous-time flow which interpolates its discrete-time trajectory.

2. Order-by-order approximations in the stepsize, $\eta > 0$, to the modified Hamiltonian $\widetilde{H}_\eta$ for symplectic Euler—and hence, the associated conserved quantity for AMD as well—are approximately conserved by the trajectories of symplectic Euler at least up to order 3 in η; furthermore, we conjecture this statement holds for all orders in η, as we formalize in Conjecture 10.

3. Under some smoothness conditions and usual domain assumptions, AMD has an $\mathcal{O}(K^{1/5})$ total regret and an $\mathcal{O}(K^{-4/5})$ duality gap of the average iterates, where K is the number of iterations of AMD. If the conjecture introduced in Section 5 is true in its entirety, then these can be improved to $\mathcal{O}\left(K^{\varepsilon}\right)$ and $\mathcal{O}\left(K^{-1+\varepsilon}\right)$, respectively, for any $\varepsilon > 0$. Even more strongly, if $\widetilde{H}_\eta$ converges absolutely, then the total regret is bounded for all K, i.e., aforementioned complexities become $\mathcal{O}\left(1\right)$ and $\mathcal{O}\left(K^{-1}\right)$, respectively. In particular, our results generalize those derived in Wibisono et al. (2022), in which the authors show an $\mathcal{O}(K^{1/3})$ total regret and an $\mathcal{O}(K^{-2/3})$ duality gap of the average iterates for AMD.

Furthermore, we compute and analyze the closed-form of $\widetilde{H}_\eta$ when the original Hamiltonian H is quadratic. We find that $\widetilde{H}_\eta$ is *different* from the conserved quantity found previously in the literature (e.g., Bailey et al., 2020).

Roadmap. The structure of the rest of the paper is as follows. In Section 3, we review some basic definitions and properties of continuous-time Hamiltonian flow, state some foundational assumptions that will hold throughout the paper, and introduce the foundations of symplectic integrators—and in particular, the symplectic Euler method—alongside connections between symplectic integrators and algorithms used throughout algorithmic game theory and optimization. In Section 4, we introduce and review the bilinear zero-sum games and how to measure the performance of algorithms used in their design, especially in the context of AMD and how it is defined. In Section 5, we show how the AMD and symplectic Euler algorithms are related to one another via duality and a linear transformation. In Section 6, we define the modified Hamiltonian $\widetilde{H}_\eta$ for the symplectic Euler method via the Baker-Campbell-Hausdorff series, present several of its known characteristics, and introduce a conjecture describing how well approximations to $\widetilde{H}_\eta$ are conserved. In Section 8, we present an example of a case where we can compute $\widetilde{H}_\eta$ in closed-form—the multivariate quadratic case, under some mild restrictions—and discuss some properties which can be derived about the trajectories of symplectic Euler in that case using $\widetilde{H}_\eta$. In Section 9, we use various properties of $\widetilde{H}_\eta$ introduced in prior sections to derive the algorithmic performance guarantees mentioned above in our contributions. Finally, in Section 10 we conclude by discussing further implications of this paper and further questions to study.

2. Preliminaries

2.1 Notations

Let $\mathbb{N} = \{1, 2, \dots\}$ be natural numbers, $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ be the non-negative integers, and $\mathcal{Z} = \{\cdots, -2, -1, 0, 1, 2, \cdots\}$ be the integers. Let $\mathcal{P}, \mathcal{Q} \subseteq \mathbb{R}^d$, so $\mathcal{P} \times \mathcal{Q} \subseteq \mathbb{R}^{2d}$. We denote an element of $\mathcal{P} \times \mathcal{Q}$ by (p, q) where $p \in \mathcal{P}$ and $q \in \mathcal{Q}$.

For $n \in \mathbb{N}_0$, let $C^n(\mathcal{P} \times \mathcal{Q})$ denote the space of n-times continuously differentiable functions $F \colon \mathcal{P} \times \mathcal{Q} \to \mathbb{R}$. Let $C^\infty(\mathcal{P} \times \mathcal{Q})$ denote the space of infinitely-differentiable functions $F \colon \mathcal{P} \times \mathcal{Q} \to \mathbb{R}$.

Given any differentiable $F : \mathcal{P} \times \mathcal{Q} \to \mathbb{R}$, we denote its gradient vector by

$$\nabla F(p, q) = \begin{pmatrix} \nabla_p F(p, q) \\ \nabla_q F(p, q) \end{pmatrix} \in \mathbb{R}^{2d}$$

where $\nabla_p F(p,q) = \frac{\partial}{\partial p} F(p,q) \in \mathbb{R}^d$ is the partial derivative in the p coordinate, and $\nabla_q F(p,q) = \frac{\partial}{\partial q} F(p,q) \in \mathbb{R}^d$ is the partial derivative in the q coordinate, both represented as column vectors.

For column vectors $u, v \in \mathbb{R}^d$, we denote their ℓ_2-inner product by $u^\top v$.

Given a function $f : \mathbb{R}^d \to \mathbb{R}$, we call f *L-smooth of order k* if f is k-times continuously differentiable and there exists a constant $L > 0$ such that for all $x \in \mathbb{R}^d$, $||\nabla^k f(x)||_{\text{op}} \leq L$, where $||\cdot||_{\text{op}}$ is the operator norm. Concretely, this means $|\nabla^k f(x)[v^{\otimes k}]| \leq L$, for all v with $||v||_2 = 1$.

2.2 Convex optimization

Let W be a subset of $\mathbb{R}^d$ with non-empty interior. A function $\Psi : W \to \mathbb{R}$ is a *Legendre-type function* if Ψ is strictly convex, $\Psi \in C^1(W)$, $\nabla\Psi : W \to \mathbb{R}^d$ is injective, and $||\nabla\Psi(w)|| \to \infty$ as w approaches ∂W (the boundary of W) (Rockafellar, 1996, Section 26). For any Legendre-type function $\Psi : W \to \mathbb{R}$, we may also define its convex conjugate $\Psi^* : \nabla\Psi(W) \to \mathbb{R}$ as follows:

$$\Psi^*(p) := \sup_{w \in W} \left\{ w^\top p - \Psi(w) \right\} \quad \forall p \in \nabla\Psi(W).$$

The gradient of the convex conjugate of Ψ is the inverse map of the gradient of Ψ—i.e., $\nabla\Psi^* = (\nabla\Psi)^{-1}$—see Calafiore and El Ghaoui (2014) or Theorem 26.5 from Rockafellar (1996).

The *Bregman divergence* of a Legendre-type function $\Psi : W \to \mathbb{R}$ is the function $D_\Psi : W \times W \to \mathbb{R}$ defined as

$$D_\Psi(w, \widetilde{w}) = \Psi(w) - \Psi(\widetilde{w}) - \nabla\Psi(\widetilde{w})^\top (w - \widetilde{w})$$

for $w, \widetilde{w} \in W$. Since Ψ is strictly convex, $D_\Psi(w, \widetilde{w}) \geq 0$ and $D_\Psi(w, \widetilde{w}) = 0$ if and only if $w = \widetilde{w}$.

The Bregman divergence can recover standard measures of "distance," e.g., the Euclidean distance (which is symmetric) when Ψ is the quadratic function, or the Kullback-Leibler divergence (which is asymmetric) when Ψ is the negative entropy function on the simplex.

2.3 Modes of convergence

The following definitions and properties are taken from Ross (2013). Let $\{a_k\}_{k \in \mathbb{N}_0}$ be a sequence of real or complex numbers. We say that the series $S := \sum_{k=0}^{\infty} a_k$ converges if and only if for all $\varepsilon > 0$, there exists an $N = N(\varepsilon) \in \mathbb{N}_0$ and $L \in \mathbb{R}$ such that for all $n \geq N$, $|\sum_{k=0}^{n} a_k - L| < \varepsilon$. We say S converges conditionally if S converges but $\lim_{N \to \infty} \sum_{k=0}^{N} |a_k| = \infty$. Finally, S converges absolutely if both S and $\sum_{k=0}^{\infty} |a_k|$ converge.

Throughout this paper, we note how absolute convergence allows one to rearrange the order of a series without affecting its convergence or limit. For an absolutely convergent series of real or complex numbers, summing that series' terms in any order will still result in the same value. More precisely, if S converges absolutely, then for any permutation $\sigma : \mathbb{N}_{>0} \to \mathbb{N}_{>0}$, we have $S = \sum_{k=0}^{\infty} a_{\sigma(k)}$. The Riemann rearrangement theorem further states that the converse of the former statement is true, although this theorem does not necessarily hold for sequences outside of the reals or complex numbers.

3. Symplectic Euler discretization of Hamiltonian flow

For now, let $\mathcal{P}, \mathcal{Q} \subseteq \mathbb{R}^d$ be closed sets with non-empty interiors, and denote $\mathcal{Z} = \mathcal{P} \times \mathcal{Q}$ the *phase space*. We write an element $z \in \mathcal{Z}$ of the phase space as $z = (p, q)$ where $p \in \mathcal{P}$ and $q \in \mathcal{Q}$.

Suppose we are given a *Hamiltonian function* $H \colon \mathcal{Z} \to \mathbb{R}$, which is an arbitrary differentiable function. Let $\Omega = \begin{pmatrix} 0 & -I_d \\ I_d & 0 \end{pmatrix}$ be the *symplectic matrix*, where $I_d \in \mathbb{R}^{d\times d}$ is the identity matrix. Recall that the **Hamiltonian flow** generated by Hamiltonian H (with respect to the symplectic matrix Ω) is the flow of the differential equation:

$$\dot{z}(t) = \Omega \nabla H(z(t)) \tag{1}$$

starting from any $z(0) \in \mathcal{Z}$. Here, $\dot{z}(t) = \frac{d}{dt} z(t)$ is the velocity, which is the time derivative of the position z.

We define the **energy** of the flow at time t to be the value of the Hamiltonian function $H(z(t))$. One notable feature of Hamiltonian flow is the conservation of energy: For all $t \geq 0$,

$$H(z(t)) = H(z(0)).$$

Indeed, we can compute:

$$\tfrac{d}{dt} H(z(t)) = \left(\nabla H(z(t))\right)^\top \dot{z}(t) = \left(\nabla H(z(t))\right)^\top \Omega \, \nabla H(z(t)) = 0,$$

where the last equality holds since $\Omega = -\Omega^\top$ is skew-symmetric.

Where necessary, we will make the following separability assumption:

Assumption 1 *The Hamiltonian function H is separable, i.e. there exist differentiable functions $F \colon \mathcal{P} \to \mathbb{R}$ and $G \colon \mathcal{Q} \to \mathbb{R}$ such that*

$$H(p, q) = F(p) + G(q)$$

for all $(p, q) \in \mathcal{Z}$. We define the extensions $\widetilde{F} \colon \mathcal{Z} \to \mathbb{R}$ and $\widetilde{G} \colon \mathcal{Z} \to \mathbb{R}$ by $\widetilde{F}(p, q) = F(p)$ and $\widetilde{G}(p, q) = G(q)$ so that we can also write $H(z) = \widetilde{F}(z) + \widetilde{G}(z)$ for all $z \in \mathcal{Z}$.

For the sake of simplicity, in the remainder of the paper, we drop the tildes from $\widetilde{F}$ and $\widetilde{G}$, as it will be clear from context whether the functions F and G or their extensions are being used. When the Hamiltonian is separable as $H = F + G$, we can write the Hamiltonian flow (1) in terms of the coordinates of $z(t) = (p(t), q(t))$ as:

$$\dot{p}(t) = -\nabla_q H(p(t), q(t)) = -\nabla G(q(t)) \tag{2a}$$

$$\dot{q}(t) = \nabla_p H(p(t), q(t)) = \nabla F(p(t)). \tag{2b}$$

To simulate the trajectories of continuous-time Hamiltonian flow under Hamiltonian H, we can discretize the Hamiltonian flow using a numerical integrator. Two of the simplest numerical integrators known in numerical analysis literature are the forward (explicit) and backward (implicit) Euler methods (Atkinson, 1988). For applications in min-max game theory, these integrators when applied to the Hamiltonian flow (2) correspond to when the two players follow simultaneous mirror descent and simultaneous proximal mirror descent, respectively (Wibisono et al., 2022). For forward Euler applied to the Hamiltonian flow (2), we have

$$p_{k+1} = p_k - \eta \nabla G(q_k), \quad q_{k+1} = q_k + \eta \nabla F(p_k), \tag{3}$$

while for backward Euler applied to the Hamiltonian flow (2), we have

$$p_{k+1} = p_k - \eta \nabla G(q_{k+1}), \quad q_{k+1} = q_k + \eta \nabla F(p_{k+1}). \tag{4}$$

As shown in Appendix B of Wibisono et al. (2022), when H is convex, H increases monotonically for forward Euler and decreases monotonically for backward Euler. In fact, the increase or decrease in H, respectively, is exponentially fast when H is strongly convex. Hence, to achieve some sort of energy conservation in discrete time, a more sophisticated discretization scheme is required.

A **symplectic integrator** is a numerical integrator $F_\eta : \mathcal{Z} \to \mathcal{Z}$—in the sense that $F_\eta(p_k, q_k) = (p_{k+1}, q_{k+1})$—for Hamiltonian flow which preserves the symplectic form, i.e., $(JF_\eta)\,\Omega\,(JF_\eta)^\top = \Omega$ for any $\eta > 0$ on all of $\mathcal{Z}$, where JF_η is the Jacobian matrix of F_η. More generally, a change of coordinates of the Hamiltonian phase space which preserves the symplectic form is called a *canonical transformation* (see, e.g., Chapter 2 in Part I of Lazutkin (1993) or Section 5.3 in Jose and Saletan (1998)), and hence, all symplectic integrators are canonical transformations.

Under certain conditions, a symplectic integrator conserves a modified Hamiltonian $\widetilde{H}_\eta$ expressible as a formal power series in η and whose level sets approximate those of H (Yoshida, 1990, 1993; Hairer et al., 2006). Moreover, the Hamiltonian flow generated by $\widetilde{H}_\eta$ exactly interpolates the discrete-time trajectories of the symplectic integrator. Some examples of symplectic integrators include the symplectic Euler, leapfrog, and Störmer–Verlet methods—Chapter VI of Hairer et al. (2006) gives a comprehensive overview of different symplectic integrators and how to derive new ones of arbitrarily high accuracy. Generally speaking, the modified Hamiltonian $\widetilde{H}_\eta$ of a given symplectic integrator might not converge (see, e.g., Field and Nijhoff (2003) or Chapter IX in Hairer et al. (2006)), in which case one must approximate $\widetilde{H}_\eta$ by taking the series up to some order in η.

By the discussion above, the forward and backward Euler methods are non-symplectic methods. The simplest symplectic integrator is the symplectic Euler method (Channell and Scovel, 1990; Yoshida, 1993), which applies the explicit Euler method for updating one component of the Hamiltonian flow (2), and the implicit Euler method for updating the other. Concretely, the **symplectic Euler method** with stepsize $\eta > 0$ is defined as follows: At each iteration $k \in \mathbb{N}_0$, from the current iterate $(p_k, q_k) \in \mathcal{Z}$, we compute the next iterate by the updates

$$p_{k+1} = p_k - \eta \nabla_q H(p_{k+1}, q_k) = p_k - \eta \nabla G(q_k) \tag{5a}$$

$$q_{k+1} = q_k + \eta \nabla_p H(p_{k+1}, q_k) = q_k + \eta \nabla F(p_{k+1}), \tag{5b}$$

since we assumed that the Hamiltonian $H = F + G$ is separable (Assumption 1). We will define the modified Hamiltonian $\widetilde{H}_\eta$ for symplectic Euler and introduce some of its properties in Section 6.

For the game-theoretic application, the symplectic Euler updates (5) correspond to the Alternating Mirror Descent algorithm in the dual space after a linear transformation; see Section 5. Before characterizing this correspondence, we review the necessary background on game theory and mirror descent in Section 4 below.

4. Alternating Mirror Descent

In this paper, we are also concerned with the problem of finding the Nash equilibrium of zero-sum games with a bilinear payoff matrix $A \in \mathbb{R}^{d\times d}$ for $d \in \mathbb{N}$ in *constrained* strategy spaces $\mathcal{A}$ and $\mathcal{B}$, which are both closed and convex subsets of $\mathbb{R}^d$:

$$\min_{a\in\mathcal{A}} \max_{b\in\mathcal{B}} a^\top Ab. \tag{6}$$

A Nash equilibrium of the bilinear zero-sum game above is a pair of points $(a^*, b^*) \in \mathcal{A}\times\mathcal{B}$ which satisfies the saddle-point inequality:

$$(a^*)^\top Ab \le (a^*)^\top Ab^* \le a^\top Ab^*, \qquad \text{for all } \ a \in \mathcal{A}, b \in \mathcal{B}. \tag{7}$$

By von Neumann's min-max theorem (v. Neumann, 1928), (a^*, b^*) always exists, provided that $\mathcal{A}, \mathcal{B}$ are compact. We can measure the convergence to a Nash equilibrium $(a^*, b^*) \in \mathcal{A}\times\mathcal{B}$ via the **duality gap** function $\mathsf{dg} : \mathcal{A}\times\mathcal{B} \to \mathbb{R}_{\ge 0}$ given by

$$\mathsf{dg}\,(a, b) = \max_{\widetilde{b}\in\mathcal{B}} a^\top A\widetilde{b} - \min_{\widetilde{a}\in\mathcal{P}} \widetilde{a}^\top Ab \qquad \text{for } \ (a, b) \in \mathcal{A}\times\mathcal{B}. \tag{8}$$

One can check that $\mathsf{dg}\,(a, b) \ge 0$ for all $(a, b) \in \mathcal{A}\times\mathcal{B}$, and $\mathsf{dg}\,(a^*, b^*) = 0$ if and only if (a^*, b^*) is a Nash equilibrium. Hence, we can use the duality gap to measure the performance of an algorithm to find a Nash equilibrium in min-max games. We also define $\overline{\mathsf{dg}}_K$ to be the duality gap of the average iterates up to K iterations, i.e., for a sequence of strategies $(a_0, b_0), \dots, (a_{K-1}, b_{K-1}) \in \mathcal{A}\times\mathcal{B}$,

$$\overline{\mathsf{dg}}_K \coloneqq \mathsf{dg}\left(\frac{1}{K}\sum_{k=0}^{K-1} a_k,\ \frac{1}{K}\sum_{k=0}^{K-1} b_k\right).$$

A natural strategy is for each player to follow a greedy method, such as gradient descent, to optimize their own objective function. In the constrained setting when $\mathcal{A}, \mathcal{B} \subsetneq \mathbb{R}^d$, it is natural for each player to follow a constrained greedy method instead of gradient descent, especially considering how the iterations of simultaneous gradient descent diverge from their initial value (Gidel et al., 2019). Hence, we consider the case where each player follows the *mirror descent algorithm* (Nemirovsky and Yudin, 1983) to minimize their loss functions over $\mathcal{A}$ and $\mathcal{B}$, respectively. This gives rise to an algorithm called **Alternating Mirror Descent (AMD)**.

Consider regularizer functions $\alpha : \mathcal{A} \to \mathbb{R}$ and $\beta : \mathcal{B} \to \mathbb{R}$ to keep the strategies of each player in $\mathcal{A}$ and $\mathcal{B}$, respectively, where α and β are each Legendre-type functions. Let

$D_\alpha : \mathcal{A} \times \mathcal{A} \to \mathbb{R}$ and $D_\beta : \mathcal{B} \times \mathcal{B} \to \mathbb{R}$ be the Bregman divergences of α and β, respectively. Then, the AMD algorithm consists of the following updates from a position $(a_k, b_k) \in \mathcal{A} \times \mathcal{B}$ at iteration (or turn) $k \in \mathbb{N}_0$, to the next position $(a_{k+1}, b_{k+1}) \in \mathcal{A} \times \mathcal{B}$ given by:

$$a_{k+1} = \arg\min_{a \in \mathcal{A}} \left\{ a^\top A b_k + \frac{1}{\eta} D_\alpha (a, a_k) \right\}, \tag{9a}$$

$$b_{k+1} = \arg\min_{b \in \mathcal{B}} \left\{ -a_{k+1}^\top A b + \frac{1}{\eta} D_\beta (b, b_k) \right\}, \tag{9b}$$

where $\eta > 0$ is the stepsize. Note, the first player updates their position a based on b_k, and the second player updates their position b based on the updated a_{k+1}. This is the algorithm that was studied in Wibisono et al. (2022). In the unconstrained setting with α, β being quadratic functions, AMD recovers the Alternating Gradient Descent (AGD) algorithm, which was studied in Bailey et al. (2020) and will be revisited later in Sections 8 and 9 of this paper.

A relevant quantity is the **total regret** R_K of the two players after K iterations under AMD. We have the following definition of R_K taken from Bailey et al. (2020); Wibisono et al. (2022):

Definition 1 *The total regret of both players after K iterations of AMD is the best cumulative regret of both players in hindsight:*

$$R_K := \max_{(a,b) \in \mathcal{A} \times \mathcal{B}} \left\{ R_{1,K}(a) + R_{2,K}(b) \right\},$$

where $R_{1,K}$ and $R_{2,k}$ are the regrets of the first and second player, respectively, defined with respect to static actions $a \in \mathcal{A}$ and $b \in \mathcal{B}$ as follows:

$$R_{1,K}(a) := \sum_{k=0}^{K-1} \left(\frac{a_k + a_{k+1}}{2} \right)^\top A b_k - \sum_{k=0}^{K-1} a^\top A b_k,$$

$$R_{2,K}(b) := \sum_{k=0}^{K-1} a_{k+1} A b - \sum_{k=0}^{K-1} a_{k+1}^\top A \left(\frac{b_k + b_{k+1}}{2} \right).$$

The total regret R_K is directly related to $\overline{\mathsf{dg}}_K$ via the following relationship:

Lemma 2 (Wibisono et al. (2022), Lemma 3.1) *Along AMD from any $(a_0, b_0) \in \mathcal{A} \times \mathcal{B}$, for any $K \geq 1$,*

$$\overline{\mathsf{dg}}_K = \frac{1}{K} R_K - \frac{1}{2K} \left(a_0^\top A b_0 - a_K^\top A b_K \right). \tag{10}$$

We note that for the constrained setting where $\mathcal{A}$ and $\mathcal{B}$ are bounded, the last term $\left(a_0^\top A b_0 - a_K^\top A b_K \right)$ in (10) is uniformly bounded in terms of the diameter of the sets, in which case Lemma 2 implies that R_K alone determines the rate of decay of $\overline{\mathsf{dg}}_K$. Hence, the total regret R_K can likewise be used to measure the performance of AMD in the context of min-max games.

Along the AMD algorithm (9), the average iterates $\left(\frac{1}{K}\sum_{k=1}^{K} a_k,\ \frac{1}{K}\sum_{k=1}^{K} b_k\right)$ converges to the Nash equilibrium, as implied by when their duality gap vanishes: $\overline{\mathsf{dg}}_K \to 0$ as $K \to \infty$. Classically, it is known that the duality gap $\overline{\mathsf{dg}}_K$ vanishes at a rate of $\mathcal{O}\left(K^{-1/2}\right)$ for the simultaneous mirror descent algorithm when the stepsize is $\eta = \Theta\left(K^{-1/2}\right)$. In contrast, for AMD, the duality gap $\overline{\mathsf{dg}}_K$ was shown to vanish at a rate of $\mathcal{O}\left(K^{-2/3}\right)$ (Wibisono et al., 2022, Corollary 3.3) when the stepsize is $\eta = \Theta\left(K^{-1/3}\right)$. We will show how to improve the latter bound in Section 9.

5. Alternating Mirror Descent and Symplectic Euler Method

Let $\mathfrak{A} := \nabla\alpha(\mathcal{A}) \subseteq \mathbb{R}^d$ and $\mathfrak{B} := \nabla\beta(\mathcal{B}) \subseteq \mathbb{R}^d$ be the images of $\mathcal{A}$ and $\mathcal{B}$ under $\nabla\alpha$ and $\nabla\beta$, respectively, which we refer to as the *dual spaces*. Let $f := \alpha^* \colon \mathfrak{A} \to \mathbb{R}$ and $g := \beta^* \colon \mathfrak{B} \to \mathbb{R}$ be the convex conjugates (Rockafellar, 1996, Section 26) of α and β, respectively. The following lemma describes how to translate between AMD and an equivalent algorithm in the dual spaces of $\mathcal{A}$ and $\mathcal{B}$:

Lemma 3 *Let $x_k := \nabla\alpha(a_k) \in \mathfrak{A}$ and $y_k := \nabla\beta(b_k) \in \mathfrak{B}$ be the dual variables of $a_k \in \mathcal{A}$, $b_k \in \mathcal{B}$, where $\mathcal{A}, \mathcal{B} \subseteq \mathbb{R}^d$ are closed and convex. If (a_k, b_k) evolves following the AMD algorithm, then (x_k, y_k) evolves via the following updates:*

$$x_{k+1} = x_k - \eta A \nabla g(y_k), \quad y_{k+1} = y_k + \eta A^\top \nabla f(x_{k+1}). \tag{11}$$

Proof Consider the primal updates in $\mathcal{A}$ given by (9a), and define the mapping

$$\varphi_k(a) := a^\top A b_k + \frac{1}{\eta} D_\alpha(a, a_k) \qquad \text{for } a \in \mathcal{A}.$$

Since α is Legendre-type on $\mathcal{A}$ and $D_\alpha(\cdot, a_k)$ consists of α plus an affine term, $D_\alpha(\cdot, a_k)$ is also Legendre-type. Hence, φ_k is strictly convex and essentially smooth, whence it follows that φ_k has a unique minimizer a_{k+1} that lies in the relative interior of $\mathcal{A}$ (Rockafellar, 1996, Sections 26 and 27).

The first-order KKT condition for minimizing φ_k over the closed convex set $\mathcal{A}$ gives

$$0 \in \nabla\varphi_k(a_{k+1}) + N_{\mathcal{A}}(a_{k+1}),$$

where $N_{\mathcal{A}}(a_{k+1})$ is the normal cone of the convex set $\mathcal{A}$ at the point a_{k+1} (Rockafellar, 1996, Definition 2.7 and Theorem 27.4). Because a_{k+1} lies in the relative interior of $\mathcal{A}$, we have $N_{\mathcal{A}}(a_{k+1}) = \{0\}$, whence it follows that $\nabla\varphi_k(a_{k+1}) = 0$.

Finally, since $\nabla_a D_\alpha(a, a_k) = \nabla\alpha(a) - \nabla\alpha(a_k)$, we have

$$\nabla\varphi_k(a_{k+1}) = A b_k + \frac{1}{\eta}\left(\nabla\alpha(a_{k+1}) - \nabla\alpha(a_k)\right) = 0. \tag{12}$$

By an analogous argument applied to the updates in $\mathcal{B}$ given by (9b), we also have

$$-A^\top a_{k+1} + \frac{1}{\eta}\left(\nabla\beta(b_{k+1}) - \nabla\beta(b_k)\right) = 0. \tag{13}$$

Thus, by (12) and (13), respectively, we can write the optimality condition for the AMD algorithm (9) as

$$\nabla\alpha\left(a_{k+1}\right)=\nabla\alpha\left(a_k\right)-\eta A b_k,\quad \nabla\beta\left(b_{k+1}\right)=\nabla\beta\left(b_k\right)+\eta A^\top a_{k+1}. \tag{14}$$

Since $\nabla\phi^*=(\nabla\phi)^{-1}$ for any Legendre-type function $\phi:\mathbb{R}^d\to\mathbb{R}$ (Rockafellar, 1996, Theorem 26.5), we can transform between (a_k,b_k) and (x_k,y_k) as follows:

$$x_k=\nabla\alpha\left(a_k\right)\quad\Longleftrightarrow\quad a_k=\nabla\alpha^*\left(x_k\right)=\nabla f\left(x_k\right), \tag{15a}$$

$$y_k=\nabla\beta\left(b_k\right)\quad\Longleftrightarrow\quad b_k=\nabla\beta^*\left(y_k\right)=\nabla g\left(y_k\right). \tag{15b}$$

Substituting (15) into (14) gives (11). ■

We call the update (11) the **Dual Alternating Mirror Descent (DAMD)** algorithm. DAMD—especially in light of its derivation from AMD—can be interpreted as a version of *Natural Gradient Descent* (Raskutti and Mukherjee, 2015; Gunasekar et al., 2021) on the dual manifold $\mathfrak{A}\times\mathfrak{B}$, where the geometry is defined by the block-diagonal metric tensor

$$G\left(x,y\right):=\begin{pmatrix}\nabla^2 f(x) & 0\\ 0 & \nabla^2 g(y)\end{pmatrix}\qquad\text{for }\ (x,y)\in\mathfrak{A}\times\mathfrak{B},$$

and the update directions are rotated by the skew-symmetric matrix Ω_A, where

$$\Omega_M:=\begin{pmatrix}0 & -M\\ M^\top & 0\end{pmatrix} \tag{16}$$

for any matrix $M\in\mathbb{R}^{d\times d}$. In fact, the equivalence between NGD and mirror descent—as established by Raskutti and Mukherjee (2015)—is structurally analogous to the equivalence we establish between DAMD and AMD, respectively. Moreover, when f and g are quadratic functions, DAMD coincides with the *Alternating Gradient-descent-ascent (AltGDA)* algorithm in min-max optimization and game theory (Lu et al., 2019; Bailey et al., 2020; Zhang et al., 2022).

We now explain the relationship between AMD and the symplectic Euler discretization of a Hamiltonian flow. We recall that we can decompose any real square matrix into a product of real symmetric matrices. We state this fact as the following lemma, and refer the reader to Bosch (1986) for the proof and details on how to compute such a decomposition using the Jordan canonical form.

Lemma 4 (Bosch (1986), Theorem 2) *Every real matrix $A\in\mathbb{R}^{d\times d}$ can be written as a product $A=UV$ of real symmetric matrices $U,V\in\mathbb{R}^{d\times d}$.*

Lemma 4 allows us to transform between DAMD and symplectic Euler, as summarized below.

Theorem 5 *Let $A=UV$ be the symmetric decomposition of the payoff matrix A, and let $\mathcal{P},\mathcal{Q}\subseteq\mathbb{R}^d$. Define the pushforward coordinates and maps*

$$x_k:=Up_k,\qquad y_k:=Vq_k, \tag{17a}$$

$$F\left(p\right):=f\left(Up\right),\quad G\left(q\right):=g\left(Vq\right) \tag{17b}$$

for all $k \in \mathbb{N}_0$ and all $(p,q) \in \mathcal{Z}$. Then, provided the domains satisfy $U\mathcal{P} = \mathfrak{A}$ and $V\mathcal{Q} = \mathfrak{B}$, the iterations (p_k, q_k) of the symplectic Euler discretization (5) *for the Hamiltonian flow generated by $H(p,q) = F(p)+G(q)$ push forward to the iterations (x_k, y_k) of DAMD in* (11).

Proof By the chain rule and the definitions above,

$$\nabla F(p) = U\nabla f(Up) = U\nabla f(x), \tag{18a}$$
$$\nabla G(q) = V\nabla g(Vq) = V\nabla g(y). \tag{18b}$$

Then, starting from the symplectic Euler update (5a) after multiplying both sides by U and using the chain rule (18a), we can write:

$$x_{k+1} = Up_{k+1} = Up_k - \eta U\nabla G(q_k) = x_k - \eta UV\nabla g(y_k) = x_k - \eta A\nabla g(y_k).$$

Similarly, multiplying both sides of (5a) by V, using the chain rule (18b), and using the fact that U and V are symmetric, we get:

$$\begin{aligned} y_{k+1} &= Vq_{k+1} = Vq_k - \eta V\nabla F(p_{k+1}) = y_k - \eta VU\nabla f(x_{k+1}) \\ &= y_k - \eta V^T U^T \nabla f(x_{k+1}) = y_k - \eta A^T \nabla f(x_{k+1}). \end{aligned}$$

This concludes the proof. ∎

Hence, Theorem 5 defines a pushforward mapping from the (p,q)-coordinates for symplectic Euler to the (x,y)-coordinates for DAMD. However, when A is not invertible, at least one of U or V must not be invertible, in which case an inverse transformation from (x,y) to (p,q) need not exist. This will not pose a problem for analyses in the following sections, since these are done with respect to (p,q), F, and G, and thus can always be translated to (x,y), f, and g, respectively.

5.1 Relation between skew-gradient flow and Hamiltonian flow

The equivalence between DAMD and symplectic Euler in discrete time is algebraically analogous with the equivalence between skew-gradient flow (e.g., Abernethy et al., 2021; Wibisono et al., 2022) and Hamiltonian flow in continuous time, respectively, in the sense that both are related via transformations in (17). Skew-gradient flow can be thought of as a generalization of Hamiltonian flow in which the symplectic matrix Ω is replaced by the skew-symmetric matrix Ω_A, as defined earlier in (16).

Note that $\Omega_{I_d} = \Omega$. Furthermore, when A is an involutory matrix, i.e., $A^2 = I_d$, then Ω_A preserves the symplectic form. Namely, we have

$$\Omega_A \Omega \Omega_A^\top = \begin{pmatrix} -A & 0 \\ 0 & -A^\top \end{pmatrix} \begin{pmatrix} 0 & A \\ -A^\top & 0 \end{pmatrix} = \begin{pmatrix} 0 & -A^2 \\ (A^\top)^2 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -I_d \\ I_d & 0 \end{pmatrix} = \Omega.$$

However, for min-max games, the payoff matrix A can be arbitrary.

In skew-gradient flow, the phase space variables $\zeta = (x,y) \in U\mathcal{P} \times V\mathcal{Q}$ obey the differential equation

$$\dot{\zeta}(t) = \Omega_A \nabla \mathsf{E}(\zeta(t))$$

for some differentiable function $\mathsf{E} : U\mathcal{P} \times V\mathcal{Q} \to \mathbb{R}$. Via analogous steps as for showing the relationship between DAMD and symplectic Euler as above, the variables (x, y) for skew-gradient flow are related to the variables (p, q) for Hamiltonian flow via the same relationships (17) at any time: $x = Up$, $y = Vq$, and $H(p, q) := \mathsf{E}(Up, Vq)$. Indeed, starting from Hamiltonian flow (2), multiplying both sides by $\begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix}$, using the chain rule (18), and noting that U and V are symmetric, we have

$$\dot{\zeta} = \begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix} \begin{pmatrix} -\nabla G(q) \\ \nabla F(p) \end{pmatrix} = \begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix} \begin{pmatrix} V & 0 \\ 0 & U \end{pmatrix} \begin{pmatrix} -\nabla g(y) \\ \nabla f(x) \end{pmatrix} \tag{19a}$$

$$= \begin{pmatrix} UV & 0 \\ 0 & VU \end{pmatrix} \begin{pmatrix} -\nabla g(y) \\ \nabla f(x) \end{pmatrix} = \begin{pmatrix} A & 0 \\ 0 & A^\top \end{pmatrix} \begin{pmatrix} -\nabla g(y) \\ \nabla f(x) \end{pmatrix} = \Omega_A \nabla \mathsf{E}(\zeta(t)). \tag{19b}$$

Hence, one can derive the dynamics for any skew-gradient flow given the dynamics of a suitable Hamiltonian flow. Furthermore, when A is invertible, we can reverse the steps above in (19) and thereby also show the converse.

In summary, the discrete-time dynamics of AMD and DAMD are related by duality; furthermore, the dynamics of DAMD and symplectic Euler are the same up to a linear transformation, thereby highlighting the equivalence between these three algorithms. The following diagram visualizes the relationships:

$$\boxed{\begin{array}{c}\text{symplectic Euler (discrete time)}\\ \text{Hamiltonian flow (continuous time)}\end{array}} \quad \begin{array}{c}\xLeftarrow{A \text{ is invertible}}\\ \Longrightarrow\end{array} \quad \boxed{\begin{array}{c}\text{alternating mirror descent (discrete time)}\\ \text{skew-gradient flow (continuous time)}\end{array}}$$

This allows one to apply a framework of analysis in the study of symplectic numerical integrators to analyze AMD. The link between symplectic numerical integrators and AMD has previously been identified in Wibisono et al. (2022), but in this work, we develop and use the links between these fields more explicitly.

6. The modified Hamiltonian

Consider a separable Hamiltonian $H \colon \mathcal{Z} \to \mathbb{R}$, such that $H = F + G$ for some differentiable functions $F \colon \mathcal{P} \to \mathbb{R}$ and $G \colon \mathcal{Q} \to \mathbb{R}$. To ensure that the domains for the Hamiltonian flow and the game-theoretic setting from Section 4 are compatible, $\mathcal{P}$ and $\mathcal{Q}$ should be chosen such that $U\mathcal{P} = \mathfrak{A}$ and $V\mathcal{Q} = \mathfrak{B}$.

The symplectic Euler method (5) is a symplectic integrator which conserves a modified Hamiltonian $\widetilde{H}_\eta$—perturbed in order-by-order corrections in η from H—provided that this perturbation series converges (c.f. Hairer et al., 2006, Sections IX.3.1–IX.3.2; Yoshida, 1993). Thus, the iterations of symplectic Euler lie on level sets of $\widetilde{H}_\eta$, thereby introducing a geometric analysis to predict dynamical properties of the iterates (e.g., boundedness) and inform analyses of both DAMD and AMD by studying these level sets. But to understand and derive $\widetilde{H}_\eta$ for symplectic Euler, we first need to define Poisson brackets.

Poisson brackets. Let $C^n(\mathcal{Z})$ be the space of n-times continuously differentiable real-valued functions over $\mathcal{Z}$ for any $n \in \mathbb{N}_0$ (see Section 2.1 for more details). For any $n \in \mathbb{N}$, we denote the *Poisson bracket*

$$\{\cdot, \cdot\} : C^n(\mathcal{Z}) \times C^n(\mathcal{Z}) \to C^{n-1}(\mathcal{Z})$$

of two functions $\varphi, \psi \in C^n(\mathcal{Z})$ by, for all $(p,q) \in \mathcal{Z}$:

$$\{\varphi, \psi\}(p,q) = -\nabla_p \varphi(p,q)^\top \nabla_q \psi(p,q) + \nabla_q \varphi(p,q)^\top \nabla_p \psi(p,q).$$

For ease of notation, we suppress the dependence on the argument (p,q), and write the above as:

$$\{\varphi, \psi\} = -\nabla_p \varphi^\top \nabla_q \psi + \nabla_q \varphi^\top \nabla_p \psi.$$

Poisson brackets can be iterated—the evaluation starts from the innermost bracket. For example, assuming that $F \in C^2(\mathcal{P})$ and $G \in C^2(\mathcal{Q})$, we have $\{\{F,G\},G\} = \nabla_q G^\top \nabla_p^2 F \nabla_q G$.

Some later definitions involve **Iterated Poisson Brackets (IPBs)**, which we denote using the following compact notation for any $N \in \mathbb{N}$ and $r_1, \ldots, r_N, s_1, \ldots, s_N \in \mathbb{N}_0$:

$$\begin{aligned}&\{F^{s_1} G^{r_1} \cdots F^{s_N} G^{r_N}\} \\ &= \{\cdots \underbrace{\{\{F, \ldots, F\}, F\}}_{s_1}, \underbrace{G\}, \cdots G\}, G\}}_{r_1}, \cdots \underbrace{F\}, \cdots F\}, F\}}_{s_N}, \underbrace{G\}, \cdots\} G\}, G\}}_{r_N}.\end{aligned}$$

For notational ease, we take the convention that single-function brackets return the function itself—i.e. $\{F\} \coloneqq F$ and $\{G\} \coloneqq G$.

N refers to the *rank* of the IPB $\{F^{s_1} G^{r_1} \cdots F^{s_N} G^{r_N}\}$. For a review on Poisson brackets, including how to exponentiate Poisson brackets as operators, please refer to Appendix A.

Formal definition. The dynamics of the symplectic Euler algorithm (5) can be interpolated by a continuous-time Hamiltonian flow according to an infinite series called a **modified Hamiltonian** (Field and Nijhoff, 2003; Alsallami et al., 2018) or **shadow Hamiltonian** (Engle et al., 2005), $\widetilde{H}_\eta$, when that series converges. By using the Dynkin form of the Baker-Campbell-Hausdorff (BCH) formula (see, e.g., Section III.4 in Hairer et al. (2006)), $\widetilde{H}_\eta$ can be expressed as the following formal series in η and IPBs of F and G:

Definition 6 *The modified Hamiltonian $\widetilde{H}_\eta$ associated with the symplectic Euler method* (5) *is defined by the following iterated summation:*

$$\begin{aligned}&\widetilde{H}_\eta(p,q) \\ &= \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} \sum_{r_1+s_1>0 \cdots r_n+s_n>0} \frac{\eta^{r_1+\cdots+r_n+s_1+\cdots+s_n-1} \{G^{r_1} F^{s_1} G^{r_2} F^{s_2} \cdots G^{r_n} F^{s_n}\}(p,q)}{(r_1+\cdots+r_n+s_1+\cdots+s_n) \prod_{i=1}^n r_i! s_i!}.\end{aligned} \tag{21}$$

See Appendix B for details on the derivation of the Dynkin series (21). Here, "formal" means that the convergence of (21), which we call the $\widetilde{H}_\eta$ series or Dynkin series, is not guaranteed. Due to its cumbersome structure, it is difficult to sum the series (21) directly and derive a closed-form. To derive a closed-form for $\widetilde{H}_\eta$ in the quadratic case (Section 8), we propose an equivalent integral formula for $\widetilde{H}_\eta$—see Theorem 18 in Appendix B.2.

We remind the reader that the series (21) is the particular form of the modified Hamiltonian $\widetilde{H}_\eta$ for symplectic Euler. The modified Hamiltonian $\widetilde{H}_\eta$ will have a different form for other symplectic numerical integrators. For more details on how to derive $\widetilde{H}_\eta$ for other symplectic numerical integrators and examples of what $\widetilde{H}_\eta$ looks like in those cases (e.g., for leapfrog integrator), see Yoshida (1993), Skeel and Hardy (2001), or Chapters IX and XV in Hairer et al. (2006).

Convergence of the modified Hamiltonian. There are only a few choices of F and G for which the modified Hamiltonian $\widetilde{H}_\eta$ is known to converge (Suris, 1989; Field and Nijhoff, 2003; Alsallami et al., 2018). The reader may refer to Appendix E in the arXiv version of this paper, Katona et al. (2024), for a detailed review of these cases.

6.1 The interpolating Hamiltonian flow

If the series (21) defining $\tilde{H}_\eta$ converges, then the iterations of symplectic Euler (5) conserve $\widetilde{H}_\eta$. i.e., $\widetilde{H}_\eta$ is constant along the iterations of the algorithm (5). This follows from how the Hamiltonian flow (2) generated by $\widetilde{H}_\eta$, whose ODE is known as the "modified equation" (Skeel and Hardy, 2001; Hairer et al., 2006), interpolates the iterates of the symplectic Euler (5). These properties can be summarized in the theorem below:

Theorem 7 *Let $F \in C^\infty(\mathcal{P})$, $G \in C^\infty(\mathcal{Q})$, and $k \in \mathbb{N}_0$.* **Assume** *that the modified Hamiltonian* (21) *converges when $0 < \eta < \eta_{\max}$ for some $\eta_{\max}$ possibly a function of (p_k, q_k), F, and G. Let (p_{k+1}, q_{k+1}) be one iteration of the symplectic Euler method* (5) *from (p_k, q_k). Then, for all $\eta \in (0, \eta_{\max})$, if we solve the initial-value problem*

$$\dot{\tilde{z}}(t) = \Omega \nabla \widetilde{H}_\eta\left(\tilde{z}(t)\right)$$

for $\tilde{z}(t) \in \mathcal{Z}$ with initial conditions $\tilde{z}(0) = (p_k, q_k)$, then $\tilde{z}(\eta) = (p_{k+1}, q_{k+1})$. In particular, $\widetilde{H}_\eta$ is conserved by one step of the symplectic Euler algorithm (5) *with stepsize η applied to (p_k, q_k):*

$$\widetilde{H}_\eta\left(p_{k+1}, q_{k+1}\right) = \widetilde{H}_\eta\left(p_k, q_k\right).$$

Theorem 7 is a known classical result with rigorous proofs found in several texts, e.g., Chapters III and IX of Hairer et al. (2006). For the sake of completeness, we provide a formal proof of Theorem 7 in Appendix B.2 of the arXiv version of this paper, Katona et al. (2024).

It might seem as if Theorem 7 implies that $\widetilde{H}_\eta(p_k, q_k)$ is conserved for any number of iterations k of symplectic Euler starting from (p_0, q_0) for a fixed stepsize $\eta > 0$. However, the maximum stepsize $\eta_{\max}$ in Theorem 7 is dependent on whether or not the series (21) for $\widetilde{H}_\eta$ converges when evaluated at (p_k, q_k) for all η up to $\eta_{\max}$ at each k. Thus, $\eta_{\max}$ might approach 0 as k grows. We can preclude this by assuming that $\widetilde{H}_\eta$ converges everywhere, as formalized by the following corollary to Theorem 7:

Corollary 8 *Let $F \in C^\infty(\mathcal{P})$ and $G \in C^\infty(\mathcal{Q})$.* **Assume** *that the modified Hamiltonian* (21) *converges pointwise on some subset D of $\mathbb{R}^d \times \mathbb{R}^d$ when $0 < \eta < \eta_{\max}$ for $\eta_{\max}$ possibly a function of D, F, and G. If $(p_k, q_k) \in D$ for all $k \in \mathbb{N}_0$, then $\widetilde{H}_\eta(p_k, q_k) = \widetilde{H}_\eta(p_0, q_0)$ for all $k \in \mathbb{N}_0$.*

Proof If there exists an $\eta_{\max} > 0$ such that $\widetilde{H}_\eta$ converges for all $\eta \in (0, \eta_{\max}]$ on a large enough domain that contains all iterations, then we can apply Theorem 7 with the same

$\eta_{\max}$ to every possible iterate of symplectic Euler. ■

In particular, if we can establish that $\widetilde{H}_\eta$ converges pointwise on all of $D = \mathbb{R}^d \times \mathbb{R}^d$ for some $\eta > 0$, then we can ensure that $\widetilde{H}_\eta\left(p_k, q_k\right) = \widetilde{H}_\eta\left(p_0, q_0\right)$ for all $k \in \mathbb{N}_0$.

7. Truncations of the modified Hamiltonian

When formally rearranged as a power series in η, the first few terms of the series $\tilde{H}_\eta$ (21) are as follows:

$$\begin{aligned}
\widetilde{H}_\eta\left(p, q\right) = F\left(p\right) + G\left(q\right) + \frac{\eta}{2}\left\{F, G\right\} + \frac{\eta^2}{12}\left(\left\{\left\{F, G\right\}, G\right\} + \left\{\left\{G, F\right\}, F\right\}\right) \\
- \frac{\eta^3}{24}\left\{\left\{\left\{F, G\right\}, G\right\}, F\right\} + \mathcal{O}\left(\eta^4\right), \qquad (22) \\
= F + G - \frac{\eta}{2}\langle \nabla_p F, \nabla_q G\rangle + \frac{\eta^2}{12}\nabla_q G^\top \nabla_p^2 F \nabla_q G + \frac{\eta^2}{12}\nabla_p F^\top \nabla_q^2 G \nabla_p F \\
- \frac{\eta^3}{24}\nabla_p F^\top \nabla_q^2 G \nabla_p^2 F \nabla_p F + \mathcal{O}(\eta^4). \qquad (23)
\end{aligned}$$

This is also stated in, e.g., Jacobson (1979), Yoshida (1993), Casas and Murua (2009), and Hall (2015). Hence, perturbation theory motivates the following definition of each order-by-order correction to $\widetilde{H}$ and the sum of all corrections up to a given order in η.

Definition 9 *For $n \in \mathbb{N}_0$, the n^{th}-order correction H_n to the modified Hamiltonian is defined to be*

$$H_n \coloneqq \frac{1}{n+1}\sum_{m=1}^{n+1}\frac{(-1)^{m-1}}{m}\sum_{\substack{r_1+s_1>0 \dots r_m+s_m>0 \\ \sum_{i=1}^m (r_i+s_i)=n+1}}\frac{\left\{G^{r_1}F^{s_1}G^{r_2}F^{s_2}\cdots G^{r_m}F^{s_m}\right\}(p,q)}{\prod_{i=1}^m r_i! s_i!}. \qquad (24)$$

For $N \in \mathbb{N}_0$, the N^{th}-order truncated modified Hamiltonian is defined as follows:

$$\widetilde{H}_\eta^{(N)}(p, q) \coloneqq \sum_{j=0}^{N} \eta^j H_j$$

where $H_0 = H$ is the original Hamiltonian.

We remark that $\widetilde{H}_\eta(p, q) = \lim_{N\to\infty} \widetilde{H}_\eta^{(N)}(p, q)$ only when $\widetilde{H}_\eta(p, q)$ converges absolutely.

Heuristically, one would expect that $\widetilde{H}_\eta^{(N)}$ is approximately conserved by the iterations of symplectic Euler (5), and hence also by DAMD (11) when transforming back to (x, y)-coordinates from (p, q)-coordinates. In particular, one would expect that the conservation in accuracy for $\widetilde{H}_\eta^{(N)}$ increases with N (as the degree of approximation increases) and decreases with η (as the discretization in time becomes less coarse). This is still a conjecture, which we state below to describe how the accuracy of conservation of $\widetilde{H}_\eta^{(N)}$ depends on problem parameters.

Conjecture 10 *If $\mathcal{P}, \mathcal{Q} \subseteq \mathbb{R}^d$ are both closed and convex, and $F : \mathcal{P} \to \mathbb{R}$ and $G : \mathcal{Q} \to \mathbb{R}$ are both L-smooth of orders $1, \ldots, N+2$, then for symplectic Euler (5) with any stepsize $\eta > 0$,*

$$|\widetilde{H}_\eta^{(N)}(z_k) - \widetilde{H}_\eta^{(N)}(z_0)| \leq k\Phi\left(N\right) L^{N+3}\eta^{N+2} \tag{25}$$

holds for each $N \in \mathbb{N}_0$ and some monotonic increasing, bounded function $\Phi : \mathbb{N}_0 \to \mathbb{Q}$.

Proof sketch. The main idea of the framework used to prove Conjecture 10 is as follows. As shown in detail in Appendix D, we notice that $z_{k+1} - z_k = (-\eta\nabla G(q_k), \eta\nabla F(p_{k+1}))$. Thus, we can express the value of $\widetilde{H}_\eta^{(N)}$ at z_{k+1} by expanding each term as a Taylor series about z_k. By choosing the order of expansion judiciously for each term in $\widetilde{H}_\eta^{(N)}$, we show that the sum of all lower-order terms vanish, leaving only $\mathcal{O}\left(\eta^{N+2}\right)$ terms. ■

We have proven Conjecture 10 in the following cases:

(a) For $N \in \{0, 1, 2, 3\}$ in any $d \in \mathbb{N}$ dimensions; and

(b) For $N \in \{0, 1, \ldots, 10\}$ in $d = 1$ dimension.

In Appendix D, we use the framework described in the proof sketch to systematically show that all lower-order terms up to $\mathcal{O}\left(\eta^{N+2}\right)$ (exclusive) vanish for $N \in \{0, 1, 2, 3\}$ to prove (a). We also describe our implementation of this framework in SymPy (Meurer et al., 2017) to show (b). This computer implementation uses the recursive form of the BCH formula—see Section 2.15 in Varadarajan (1984), Casas and Murua (2009), and also (92)—to compute each nth-order correction to $\widetilde{H}_\eta$ and applies symbolic differentiation to show that the necessary higher-order terms vanish. Finally, in Appendix E, we describe what we know thus far about the coefficient function Φ.

Conjecture 10 implies a weaker form of Theorem 4.4 from Wibisono et al. (2022) for $N = 1$, and it generalizes their theorem to at least $N = 3$. We note that Theorem 4.4 from Wibisono et al. (2022) has $1/12$ in place of $\Phi\left(1\right) = 3$, which they achieve by relating the Bregman divergence of H to $\widetilde{H}_\eta^{(1)}$; we could not replicate as tight of a constant for $N > 1$. But unlike how Wibisono et al. (2022) distinguishes the smoothness of the first- and third-derivatives of $H = F + G$, we could choose not to express the dependence of our conjecture (25) on the smoothness of separate higher-order derivatives of F and G. This is done for brevity—the interested reader may refer to Appendix D to see the explicit relationships for each separate derivative. Wibisono et al. (2022) also has a factor of $\sigma_{\max}\left(A\right)^{N+2}$, where $\sigma_{\max}\left(A\right)$ is the largest singular value of A. We do not have this in Conjecture 10 because we are working with F and G as functions of (p, q) rather than f and g as functions of (x, y). By generalizing how a factor of $\sigma_{\max}\left(A\right)^3$ shows up in Theorem 4.4 from Wibisono et al. (2022), we see how Conjecture 10 would change for the trajectories of DAMD instead of symplectic Euler: we would incur a factor of $\sigma_{\max}\left(A\right)^{N+2}$ on the righthand side of (25), F and G would be replaced by f and g, respectively, and the domains $\mathcal{P}$ and $\mathcal{Q}$ would be replaced by $\mathfrak{A}$ and $\mathfrak{B}$, respectively.

The example in Figure 1 does not exhibit a linear growth rate in the error (25) with the number of iterations, k—rather, it demonstrates an error which remains bounded for all k. This suggests that further work may be necessary to refine the dependence on k in

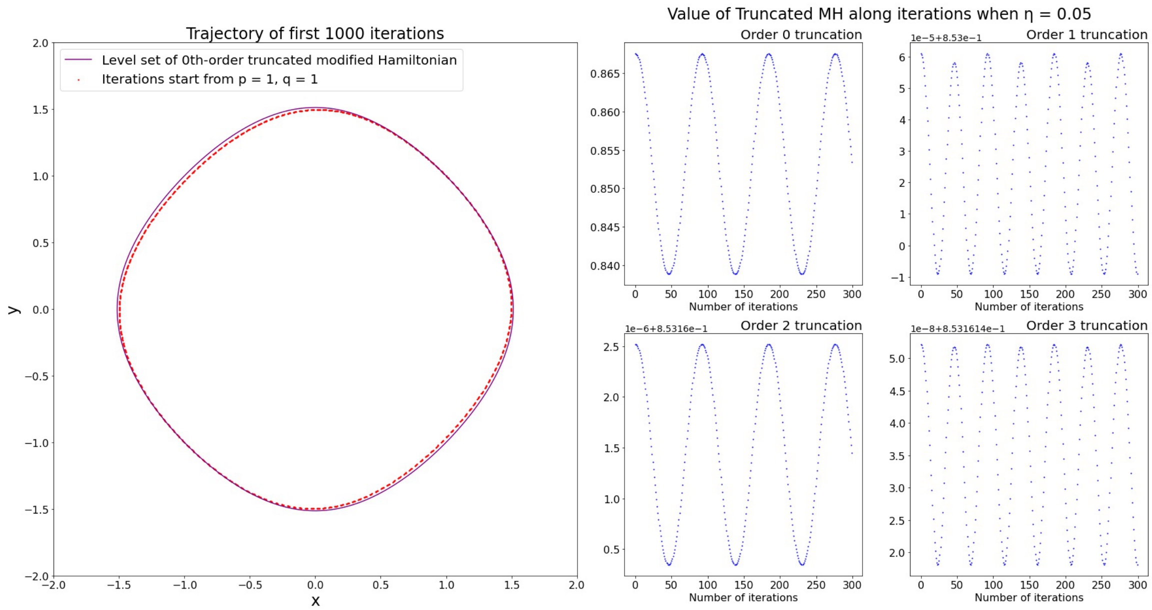

Figure 1: A simulation of the first 1000 iterations of the algorithm in the case where $F = \log(\cosh(p)), G = \log(\cosh(q))$ and $\eta = 0.05$ starting from $p_0 = 1, q_0 = 1$. The larger picture on the left contains the trajectory of the iterations, and the smaller four pictures on the right contain the value of the modified Hamiltonian along the iterations. Note that the small label 1e-6+8.5316e-1 (or analogous notation for the other subplots) is Matplotlib's offset notation: every y-value on that axis actually represents $10^{-6}y + 0.85316$. In other words, the modified Hamiltonian oscillates a few micro-units around 0.85316, and that level of oscillation decreases as the truncation order increases.

Conjecture 10, perhaps upon additional assumptions for F and G, e.g., convexity or closed level sets for $H = F + G$ as in Figure 1. Prior results (Hairer et al., 2006, Theorem 8.1, Chapter IX; Theorem 3.1, Chapter X) feature bounds like (25) with no linear growth rate in k but require either knowing that the trajectories of iterations of symplectic Euler (5) are bounded or that certain resonance and complete integrability conditions are met, both of which we cannot guarantee a priori. Our interest is in showing long-term energy conservation without assuming boundedness of trajectories, since if we already knew that the trajectories of the algorithm were bounded, then the following lemma implies that the total regret for AMD is bounded, as formalized later in Corollary 15:

Lemma 11 (Wibisono et al. (2022), Theorem 4.5) *Suppose that (a_k, b_k) evolve according to the AMD algorithm (9) with stepsize $\eta > 0$. Let $\zeta_k := (x_k, y_k) = (\nabla\alpha(a_k), \nabla\beta(b_k))$ be the dual variables of (a_k, b_k) for $k = 0, 1, \ldots, K$, as defined in Section 5. Then, we have the following formula for the cumulative regret $R_{1,K}(a) + R_{2,K}(b)$ after K iterations of the AMD algorithm for any $(a, b) \in \mathcal{A} \times \mathcal{B}$:*

$$R_{1,K}(a) + R_{2,K}(b) = \frac{1}{\eta}\left(D_H(\zeta_0, \zeta) - D_H(\zeta_K, \zeta) + \widetilde{H}_\eta^{(1)}(\zeta_K) - \widetilde{H}_\eta^{(1)}(\zeta_0)\right), \tag{26}$$

where and D_H is the Bregman divergence of $H = \alpha^* + \beta^*$ (see Section 4 for the definition). Moreover, since H is convex, we can further bound $D_H(\zeta_K, \zeta) \geq 0$, thereby yielding the following upper bound on (26)*:*

$$R_{1,K}(a) + R_{2,K}(b) \leq \frac{1}{\eta}\left(D_H(\zeta_0, \zeta) + \widetilde{H}_\eta^{(1)}(\zeta_K) - \widetilde{H}_\eta^{(1)}(\zeta_0)\right). \tag{27}$$

Proof See Appendix B.3.4 of Wibisono et al. (2022) for the proof. ■

As shown in Section 9, Lemma 11 can also be used to relate H to regret bounds for AMD.

8. Example of a closed-form modified Hamiltonian: quadratic case

We now compute the modified Hamiltonian $\widetilde{H}_\eta$ in closed form for the quadratic case, i.e., when

$$F(p) = p^\top Bp, \qquad G(q) = q^\top Cq$$

for some $B, C \in \mathbb{R}^{d\times d}$. To the best of the authors' knowledge, this is among the first attempts (Field and Nijhoff (2003) is another) to compute $\widetilde{H}_\eta$ by summing the Dynkin series (21) explicitly.

The quadratic case is closely related to when AMD or AGD is applied to unconstrained bilinear zero-sum games, as detailed in Bailey et al. (2020). In particular, when $\mathcal{A} = \mathcal{B} = \mathbb{R}^d$ and α and β are positive-definite quadratic forms, their convex conjugates α^* and β^* are also positive-definite quadratic forms and would thereby fall under this case.

For Bailey et al. (2020), the iterations of DAMD follow the updates (5) with $f = \frac{1}{2}x^\top x$ and $g = \frac{1}{2}y^\top y$. The authors show that these updates conserve the second-order quantity $x^\top x + y^\top y - 2\eta x^\top Ay$ in the original dual variables (x, y). Equivalently, in the (p, q)-coordinates, we have $F = \frac{1}{2}p^\top U^2 p$ and $G = \frac{1}{2}q^\top V^2 q$, and therefore, the conserved quantity is

$$\mathsf{S}(p, q) \coloneqq p^\top U^2 p + q^\top V^2 q - 2\eta p^\top UAVq,$$

where $A = UV$ is the symmetric decomposition of A. Hence, this is an instance of the quadratic case above with $B = U^2$ and $C = V^2$—cf. Theorem 4 from Bailey et al. (2020) for details and a derivation. We demonstrate that $\widetilde{H}_\eta$ and $\mathsf{S}(p, q)$ recover the same conserved quantity (up to a scalar factor which is a function of η) when the dimension d of the strategy space is equal to 1, but recover a different one when $d > 1$.

We have the following theorem regarding the closed form of $\widetilde{H}_\eta$ in the quadratic case.

Theorem 12 *Let $d \in \mathbb{N}$, $p, q \in \mathbb{R}^d, B, C \in \mathbb{R}^{d\times d}$. Let $F(p) = p^\top Bp, G(q) = q^\top Cq$, where we assume without loss of generality that B and C are symmetric (if not, then replace B and C with their symmetrizations $\frac{B+B^\top}{2}$ and $\frac{C+C^\top}{2}$, respectively). Furthermore, assume that BC is diagonalizable and denote its diagonalization as $BC = Q^{-1}\Lambda Q$. Define $T(\eta, \lambda)$ as the following:*

$$T(\eta, \lambda) = \begin{cases} \frac{\arcsin(\sqrt{\lambda\eta^2})}{\sqrt{\lambda\eta^2(1-\lambda\eta^2)}} & \lambda > 0, \\ \frac{\operatorname{arcsinh}(\sqrt{-\lambda\eta^2})}{\sqrt{-\lambda\eta^2(1-\lambda\eta^2)}} & \lambda < 0. \end{cases} \tag{28}$$

When $\sigma_{\max}(BC)\eta^2 < 1$*, the series for the modified Hamiltonian* $\widetilde{H}_\eta$ *conserved by the iterations of symplectic Euler* (5) *converges absolutely to*

$$\widetilde{H}_\eta(p,q) = pQ^{-1}T(\eta,\Lambda)QBp + qCQ^{-1}T(\eta,\Lambda)Qq - 2pQ^{-1}|\Lambda|T(\eta,\Lambda)Qq, \tag{29}$$

where $T(\eta,\Lambda)$ *signifies applying the function* $T(\eta,\cdot)$ *to the entries of* Λ *element-wise, and* $|\Lambda|$ *denotes the element-wise absolute value of the matrix* Λ*. In particular, when* $d = 1$*, the closed-form expression* (29) *simplifies to*

$$\widetilde{H}_\eta(p,q) = T(\eta,BC)(Bp^2 + Cq^2 - 2BCpq) = T(\eta,BC)\mathsf{S}\,(p,q)\,, \tag{30}$$

noting that B *and* C *are scalars when* $d = 1$*.*

We provide the proof of Theorem 12 in Appendix C.2 for the univariate case, and in Appendix C.4 for the multivariate case. The proofs are computational and use the integral form of $\widetilde{H}_\eta$ (Theorem 18).

When B and C are positive semidefinite matrices, their product BC is diagonalizable and has nonnegative eigenvalues (Hong and Horn, 1991). Hence, in particular, the positive-definite quadratic form case mentioned above satisfies the conditions for Theorem 12.

For a plot of $T(\eta, BC)$ vs. η for varying values of B and C, see Figure 2 in Appendix C.2. We can compute the Taylor series expansion for $T(\eta,\lambda)$ centered at $\eta = 0$ as:

$$T(\eta,\lambda) = 1 + \frac{2\lambda\eta^2}{3} + \frac{8\lambda^2\eta^4}{15} + O(\eta^6).$$

As shown by (30), when $d = 1$, we also note that the modified Hamiltonian $\tilde{H}_\eta$ from Theorem 12 is equivalent with (as a scalar multiple of) the known conserved quantity, $\mathsf{S}\,(p,q) = p^\top U^2 p + q^\top V^2 q - 2\eta p^\top UAVq$, introduced in Bailey et al. (2020). However, in general, $\widetilde{H}_\eta$ from Theorem 12 is functionally independent from Q, which is possible since higher-dimensional dynamical systems can have multiple independent conserved quantities. In particular, an autonomous dynamical system in d dimensions can have at most $d-1$ functionally independent constants of motion (since one dimension corresponds to the flow itself), while a Hamiltonian system of dimension $d = 2n$ can have at most n independent integrals in involution—see, e.g., Sections 3.2 and 6.2 in Jose and Saletan (1998), or Section 3.2 and Chapter 5 in Arnold et al. (2006).

We further observe that, for the quadratic case, although $\widetilde{H}_\eta$ is different from Q, the range of η for which the modified Hamiltonian converges lies within the same range as is necessary for the algorithm to have bounded trajectories—i.e., $\sigma_{\max}(B)\,\sigma_{\max}(C)\,\eta^2 < 1$—as can be shown by studying the level sets of Q (Bailey et al., 2020, Theorem 5). This is because $\sigma_{\max}(BC) \le \sigma_{\max}(B)\,\sigma_{\max}(C)$, in which case $\sigma_{\max}(B)\,\sigma_{\max}(C)\,\eta^2 < 1$ implies $\sigma_{\max}(BC)\,\eta^2 < 1$. Thus, the sufficient condition for $\widetilde{H}_\eta$ to converge under Theorem 12 is weaker than that for symplectic Euler to have bounded trajectories, and so $\widetilde{H}_\eta$ could still converge when the trajectories are unbounded.

In Appendix C.3, we also explicitly check the implications from Section 6.1—and in particular, Theorem 7—in the quadratic case for $d = 1$. Namely, we verify explicitly that if $(\tilde{p}(t), \tilde{q}(t))$ follows the Hamiltonian flow in $d = 1$ generated by the modified Hamiltonian (30) with initial conditions $(\tilde{p}(0), \tilde{q}(0)) = (p_k, q_k)$, then the value of the flow at $t = \eta$, $(\tilde{p}(\eta), \tilde{q}(\eta))$, coincides with (p_{k+1}, q_{k+1}).

9. Improved regret analysis of Alternating Mirror Descent

An application of Conjecture 10 lies in studying the algorithmic performance of AMD in the context of zero-sum games introduced in Section 4. Theorem 3.2 from Wibisono et al. (2022) derives a bound on the total regret after K iterations, R_K, in terms of the smoothness of α and β and the sizes of $\mathcal{A}$ and $\mathcal{B}$, and from there concludes that choosing an $\eta = \Theta\left(K^{-1/3}\right)$ stepsize leads to $R_K = \mathcal{O}\left(K^{1/3}\right)$ and a duality gap of the average iterates $\overline{\mathsf{dg}}_K = \mathcal{O}\left(K^{-2/3}\right)$. Using the truncated modified Hamiltonians $\widetilde{H}_\eta^{(N)}$ and Conjecture 10, we generalize Theorem 3.2 from Wibisono et al. (2022) upon assuming higher-order smoothness of α and β.

Firstly, Lemma 11 allows us to prove Theorem 13:

Theorem 13 *Let $\mathcal{A}, \mathcal{B}$ be closed, convex subsets of $\mathbb{R}^d$. Suppose we start from $(a_0, b_0) \in \mathcal{A} \times \mathcal{B}$ such that $D_\alpha(a_k, a_0)$ and $D_\beta(b_k, b_0)$ are uniformly bounded across k—i.e., there exists $M > 0$ with $D_\alpha(a_k, a_0), D_\beta(b_k, b_0) \le M/2$ for all $k \in \mathbb{N}$. Furthermore, suppose that α^* and β^* are L-smooth of orders $1, \dots, N+2$ for some $N \in \mathbb{N}$ over the closed convex hulls of the dual spaces $\mathfrak{A}$ and $\mathfrak{B}$, respectively.*

Then, for all $N \in \mathbb{N}$ such that Conjecture 10 holds, there exists $C > 0$ such that if both players follow AMD with stepsize $\eta = \mathcal{O}(1)$, then the total regret R_K at iteration K is bounded by

$$R_K \le \frac{M}{\eta} + 2\eta \sup_{z \in \mathcal{Z}} \sum_{j=0}^{N-2} \eta^j \left|H_{j+2}(z)\right| + CKL^{N+3}\sigma_{\max}(A)^{N+2}\eta^{N+1}, \tag{31}$$

where the middle summation in (31) vanishes when $N = 1$, $\mathcal{P}, \mathcal{Q} \subseteq \mathbb{R}^d$ are chosen such that $U\mathcal{P} = \mathfrak{A}$ and $V\mathcal{Q} = \mathfrak{B}$, and the H_j's are the jth-order corrections to the modified Hamiltonian $\widetilde{H}_\eta$ generated by $F(\cdot) = \alpha^*(U(\cdot))$ and $G(\cdot) = \beta^*(V(\cdot))$ (see Section 6).

In particular, given a horizon K, if we set the stepsize $\eta = \Theta\left(K^{-1/(N+2)}\right)$, then R_K is $\mathcal{O}\left(K^{1/(N+2)}\right)$. Finally, provided that $||a_k||$ and $||b_k||$ are uniformly bounded for all $k \in \mathbb{N}$, the duality gap of the average iterates $\overline{\mathsf{dg}}_K$ decays like $\mathcal{O}\left(K^{-(N+1)/(N+2)}\right)$.

Proof Any bounds on the kth derivatives for F and G are the same as those for f and g, respectively, except at worst scaled by $\sigma_{\max}(A)^k$. Hence, without loss of generality, we can assume that $A = I$—in which case $p = x$, $q = y$, $f = F$, and $g = G$—and rescale L by $\sigma_{\max}(A)$ wherever it appears at the end of the proof and then divide by $\sigma_{\max}(A)$ once.

Take $\mathcal{A}$ and $\mathcal{B}$ to be closed, convex sets. Then, for any $\zeta, \zeta_0 \in \mathfrak{A} \times \mathfrak{B}$ and all $k \in \mathbb{N}$,

$$D_H(\zeta_0, \zeta_k) = D_f(x_0, x_k) + D_g(y_0, y_k) = D_\alpha(a_k, a_0) + D_\beta(b_k, b_0) \le \frac{M}{2} + \frac{M}{2} = M. \tag{32}$$

Thus, by applying (27) from Lemma 11, (32), and the triangle inequality, we have

$$\begin{aligned} \eta R_K &\le M + \left|\widetilde{H}_\eta^{(1)}(z_K) - \widetilde{H}_\eta^{(1)}(z_0)\right| \\ &\le M + \left|\widetilde{H}_\eta^{(N)}(z_K) - \widetilde{H}_\eta^{(1)}(z_K)\right| + \left|\widetilde{H}_\eta^{(N)}(z_K) - \widetilde{H}_\eta^{(N)}(z_0)\right| + \left|\widetilde{H}_\eta^{(N)}(z_0) - \widetilde{H}_\eta^{(1)}(z_0)\right|. \end{aligned} \tag{33}$$

Then, taking $N \geq 1$, we can use the successive terms $H_2, \ldots, H_N$ in the Dynkin series, order-by-order in η, to express the following:

$$\widetilde{H}_\eta^{(N)} - \widetilde{H}_\eta^{(1)} = \begin{cases} 0, & N = 1, \\ \eta^2 \left(H_2 + \eta H_3 + \cdots + \eta^{N-2} H_N \right), & N > 1. \end{cases} \tag{34}$$

By applying the triangle inequality to (34), we deduce that there exists some $B_{N,\alpha,\beta} \geq 0$ such that $\left| \widetilde{H}_\eta^{(N)} - \widetilde{H}_\eta^{(1)} \right| \leq \eta^2 B_{N,\alpha,\beta}/2$ on all of $\mathcal{Z}$ for any fixed $N \geq 1$, where

$$B_{N,\alpha,\beta} = \begin{cases} 0, & N = 1, \\ 2 \sup_{z \in \mathcal{Z}} \sum_{j=0}^{N-2} \eta^j \left| H_{j+2}(z) \right|, & N > 1. \end{cases}$$

The bound in (33) implies

$$\eta R_K \leq M + \eta^2 B_{N,\alpha,\beta} + \left| \widetilde{H}_\eta^{(N)} (z_K) - \widetilde{H}_\eta^{(N)} (z_0) \right|. \tag{35}$$

Hence, applying Conjecture 10 directly to (35):

$$\left| \widetilde{H}_\eta^{(N)} (z_K) - \widetilde{H}_\eta^{(N)} (z_0) \right| \leq CKL^{N+3} \eta^{N+2}, \tag{36}$$

where $C > 0$ is an upper bound on Φ: $\Phi(n) \leq C$ for all $n \in \mathbb{N}_0$. Substituting (36) into (35), we have derived (31), i.e.,

$$R_K \leq \frac{M}{\eta} + 2\eta \sup_{z \in \mathcal{Z}} \sum_{j=0}^{N-2} \eta^j \left| H_{j+2}(z) \right| + CKL^{N+3} \sigma_{\max}(A)^{N+2} \eta^{N+1} \tag{37}$$

if we take the convention that "$\sum_{j=0}^{-1} = 0$." In particular, given a horizon K, if we further set $\boxed{\eta = \Theta\left(K^{-1/(N+2)}\right)}$, then (37) gives us a total regret $\boxed{R_K = \mathcal{O}\left(K^{1/(N+2)}\right)}$.

It remains to show the iteration complexity of the duality gap of the average iterates, $\overline{\mathsf{dg}}_K$. By assumption, there exists some $B > 0$ such that $||a_k||, ||b_k|| \leq B$ for all $k \in \mathbb{N}$. Thus, after using Lemma 2, the duality gap (10) can be bounded as follows:

$$\begin{aligned} \left| \overline{\mathsf{dg}}_K \right| &\leq \frac{1}{K} \left| R_K \right| + \frac{1}{2K} \left| a_0^\top A b_0 \right| + \frac{1}{2K} \left| a_K^\top A b_K \right| \\ &\leq \frac{1}{K} \left| R_K \right| + \frac{1}{2K} \sigma_{\max}(A) ||a_0|| \, ||b_0|| + \frac{1}{2K} \sigma_{\max}(A) ||a_K|| \, ||b_K|| \\ &\leq \frac{1}{K} \left| R_K \right| + \frac{1}{K} \sigma_{\max}(A) B^2. \end{aligned} \tag{38}$$

In the first inequality above, we use the triangle inequality. Furthermore, since $R_K = \mathcal{O}\left(K^{1/(N+2)}\right)$, it follows that $R_K/K = \mathcal{O}\left(K^{-(N+1)/(N+2)}\right)$. Hence, provided that $||a_k||$ and $||b_k||$ are uniformly bounded for all $k \in \mathbb{N}$, (38) implies the following:

$$\overline{\mathsf{dg}}_K = \mathcal{O}\left(K^{-(N+1)/(N+2)}\right) + \mathcal{O}\left(K^{-1}\right) = \mathcal{O}\left(K^{-(N+1)/(N+2)}\right).$$

Finally, we claim that the series remainder $\sum_{j=0}^{N-2} \eta^j |H_{j+2}(z)|$ in Theorem 13 is bounded; establishing this is necessary to ensure that the algorithmic guarantees claimed therein are actually valid. Note that since f and g are L-smooth of orders $1, \ldots, N+2$, this implies that the operator norms of the derivatives up to order $N+2$ of f and g are bounded by L. We also note from Definition 9 that H_j consists of a linear combination of IPBs whose order (i.e., the number of iterated Poisson brackets) is j. Each of these IPBs of order j consists of an inner product of higher-order derivatives of f and g of at most order j each, and so these are each bounded by L^{j+1}. Thus, since the $\sum_{j=0}^{N-2} \eta^j |H_{j+2}(z)|$ only goes up to order $j+2=N$ (i.e., with respect to H_{j+2}), $\sum_{j=0}^{N-2} \eta^j |H_{j+2}(z)|$ involves a sum of products of higher-order derivatives of at most order N, and so it is bounded. ■

Given what we have proven so far for Conjecture 10, as of now, Theorem 13 confirms that for any dimension d, choosing $\eta = \Theta\left(K^{-1/5}\right)$ gives us $R_K = \mathcal{O}\left(K^{1/5}\right)$ and $\overline{\mathsf{dg}}_K = \mathcal{O}\left(K^{-4/5}\right)$ when all derivatives up to order $N=5$ for F and G are bounded on $\mathcal{P}$ and $\mathcal{Q}$, respectively.

Furthermore, for $d=1$, choosing $\eta = \Theta\left(K^{-1/12}\right)$ gives us $R_K = \mathcal{O}\left(K^{1/12}\right)$ and $\overline{\mathsf{dg}}_K = \mathcal{O}\left(K^{-11/12}\right)$ when all derivatives up to order $N=12$ for F and G are bounded.

But finally, *provided that Conjecture 10 is true for all* $N \in \mathbb{N}_0$, Theorem 13 implies that setting $\eta = \Theta\left(K^{-\varepsilon}\right)$ gives us a total regret $R_K = \mathcal{O}\left(K^{\varepsilon}\right)$ and a duality gap of the average iterates which decays like $\overline{\mathsf{dg}}_K = \mathcal{O}\left(K^{-1+\varepsilon}\right)$ for any fixed $\varepsilon > 0$ when all higher-order derivatives of F and G are bounded.

9.1 Analytic and geometric paths to bounded regret for AMD

Even if Conjecture 10 is true for all $N \in \mathbb{N}_0$, we still *cannot* just take $\varepsilon \to 0$ (or equivalently, $N \to \infty$) in R_K and $\overline{\mathsf{dg}}_K$ unless the $\widetilde{H}_\eta$ series remainder $\sup_{z\in\mathcal{Z}} \sum_{j=0}^{N-2} \eta^j |H_{j+2}(z)|$ remains bounded as $N \to \infty$ and $|\eta L \sigma_{\max}(A)| \leq 1$. The latter condition necessitates a uniform bound L on *all* higher-order derivatives of F and G, which is highly restrictive.

That being said, we can avoid using Theorem 13 and effectively take $\varepsilon \to 0$ when $\widetilde{H}_\eta$ converges under certain conditions, as summarized below:

Theorem 14 *Let $\mathcal{A}, \mathcal{B} \subseteq \mathbb{R}^d$ be closed and convex. Suppose we start from $(a_0, b_0) \in \mathcal{A} \times \mathcal{B}$ such that there exists $M > 0$ with $D_\alpha(a_k, a_0), D_\beta(b_k, b_0) \leq M/2$ for all $k \in \mathbb{N}$, and that $\alpha^* \in C^\infty(\mathfrak{A})$, $\beta^* \in C^\infty(\mathfrak{B})$. Furthermore, suppose that the following hold when $0 < \eta < \eta_{\max}$ for some $\eta_{\max} > 0$ possibly a function of A, $\mathcal{A}$, $\mathcal{B}$, α, and β:*

- *The modified Hamiltonian $\widetilde{H}_\eta$ generated by $F(\cdot) = \alpha^*(U(\cdot))$ and $G(\cdot) = \beta^*(V(\cdot))$ exists (i.e., converges pointwise) at each point in $\mathcal{Z}$, where $\mathcal{P}, \mathcal{Q} \subseteq \mathbb{R}^d$ are chosen such that $U\mathcal{P} = \mathfrak{A}$ and $V\mathcal{Q} = \mathfrak{B}$; and*
- *The second-order tail $\left|\widetilde{H}_\eta(z_k) - \widetilde{H}_\eta^{(1)}(z_k)\right|$ is uniformly bounded for all $k \in \mathbb{N}$ by some $G > 0$, where the H_j's are the jth-order corrections to $\widetilde{H}_\eta$ (see Section 6).*

Then, for all $\eta \in (0, \eta_{\max})$, the total regret R_K for any horizon K is bounded by $\frac{1}{\eta}(2G+M) < \infty$. Thus, $R_K = \mathcal{O}(1)$, and provided that $||a_k||$ and $||b_k||$ are uniformly bounded for all $k \in \mathbb{N}$, the duality gap of the average iterates $\overline{\mathsf{dg}}_K$ decays like $\mathcal{O}\left(K^{-1}\right)$.

Proof As with the proof for Theorem 13, we start by taking $\mathcal{A}$ and $\mathcal{B}$ to be closed, convex sets, in which case (32) holds all the same. Hence, after applying Lemma 11, (32), and the triangle inequality, we have

$$\eta R_K \leq M + \left|\widetilde{H}_\eta(z_K) - \widetilde{H}_\eta^{(1)}(z_K)\right| + \left|\widetilde{H}_\eta(z_K) - \widetilde{H}_\eta(z_0)\right| + \left|\widetilde{H}_\eta(z_0) - \widetilde{H}_\eta^{(1)}(z_0)\right|. \quad (39)$$

We assume that the modified Hamiltonian $\widetilde{H}_\eta$ converges everywhere on $\mathcal{Z}$. Thus, the third term in (39) vanishes since $\widetilde{H}_\eta$ is conserved under the iterations of (5), and hence likewise by DAMD (11) and then AMD (9). Furthermore, by assuming that $\widetilde{H}_\eta(z_k) - \widetilde{H}_\eta^{(1)}(z_k)$ is bounded by $G > 0$ for all $k \in \mathbb{N}$, we bound (39) from above as follows:

$$\eta R_K \leq M + G + 0 + G = M + 2G \qquad \Longleftrightarrow \qquad R_K \leq \frac{1}{\eta}(M + 2G).$$

The rest of the proof follows analogously to the proof for Theorem 13 as it does after (37). ■

Hence, when the modified Hamiltonian $\widetilde{H}_\eta$ converges globally and its second-order correction $\widetilde{H}_\eta - \widetilde{H}_\eta^{(1)}$ remains uniformly bounded along the iterates, Theorem 14 implies that AMD attains the same algorithmic complexity as proximal (implicit) mirror descent (Nemirovski, 2004; Wibisono et al., 2022).

As an application of Theorem 14, Theorem 12 identifies the quadratic setting as a case where absolute convergence of $\widetilde{H}_\eta$ can be established rigorously. This setting is equivalent to Alternating Gradient Descent (AGD) as studied by Bailey et al. (2020), whose Theorem 1 shows bounded regret by a different argument.

In this case, we have $\mathcal{Z} = \mathbb{R}^d \times \mathbb{R}^d$, $F(p) = p^\top Bp$ and $G(q) = q^\top Cq$ with B and C positive-definite, and regularizers with conjugates

$$\alpha(a) = \frac{1}{4}a^\top B^{-1}a \quad \Longleftrightarrow \quad \alpha^*(p) = p^\top Bp, \qquad \beta(b) = \frac{1}{4}b^\top C^{-1}b \quad \Longleftrightarrow \quad \beta^*(q) = q^\top Cq,$$

such that $F = \alpha^*$ and $G = \beta^*$. The modified Hamiltonian $\widetilde{H}_\eta$ converges absolutely whenever $\eta^2\sigma_{\max}(BC) < 1$. Furthermore, when $\eta^2\sigma_{\max}(B)\sigma_{\max}(C) < 1$, Theorem 5 in Bailey et al. (2020) implies that the DAMD (or equivalently, AGD) iterates lie on compact quadratic level sets (ellipsoids). The corresponding Bregman divergences are the weighted Euclidean (Mahalanobis) distances

$$D_\alpha(a_k, a_0) = \frac{1}{4}\|a_k - a_0\|^2_{B^{-1}} = \frac{1}{4}(a_k - a_0)^\top B^{-1}(a_k - a_0),$$
$$D_\beta(b_k, b_0) = \frac{1}{4}\|b_k - b_0\|^2_{C^{-1}} = \frac{1}{4}(b_k - b_0)^\top C^{-1}(b_k - b_0),$$

and hence, these remain uniformly bounded for all iterates $k \in \mathbb{N}$. Moreover, since $\widetilde{H}_\eta$ converges absolutely on these compact ellipsoids by Theorem 12, the remainder $\sum_{j=2}^\infty \eta^j H_j(z_k)$ is uniformly bounded for all $k \in \mathbb{N}$. Hence, by applying Theorem 14, we recover the $\mathcal{O}(K^{-1})$ time-average regret obtained in Bailey et al. (2020).

However, to show that the conditions from Theorem 14 hold in the quadratic case, we used the fact that the symplectic Euler iterates lie on compact ellipsoids. As we noted at the end of Section 7, bounded iterates already imply a bounded regret by Lemma 11:

Corollary 15 *Let $\mathcal{A}, \mathcal{B} \subseteq \mathbb{R}^d$ be closed and convex, and let $z_k \in \mathcal{Z}$ denote the symplectic Euler iterates with stepsize $\eta > 0$. Assume the following:*

1. *For some $C > 0$, the dual iterates $z_k = (p_k, q_k)$ satisfy $||p_k - p_0||, ||q_k - q_0|| \leq C$ for all $k \in \mathbb{N}$, and*

2. *$\nabla F = \nabla \alpha^* (U(\cdot))$ and $\nabla G = \nabla \beta^* (V(\cdot))$ are continuous on $\mathcal{P}$ and $\mathcal{Q}$, respectively.*

Define the closed balls $\mathbb{B}_p := \{p \in \mathcal{P} : ||p - p_0|| \leq C\}$ and $\mathbb{B}_q := \{q \in \mathcal{Q} : ||q - q_0|| \leq C\}$. By continuity, the following suprema are finite:

$$L_F := \sup_{u \in \mathbb{B}_p} ||\nabla F(u)||, \quad L_G := \sup_{v \in \mathbb{B}_q} ||\nabla G(v)||, \quad M_F := \sup_{u \in \mathbb{B}_p} |F(u)|, \quad M_G := \sup_{v \in \mathbb{B}_q} |G(v)|.$$

Let $L := \max\{L_F, L_G\}$ and $M := \max\{M_F, M_G\}$. Then, for every horizon $K \in \mathbb{N}$,

$$R_K \leq \frac{4}{\eta} (M + CL) + L^2.$$

In particular, $R_K = \mathcal{O}(1)$ and $\overline{\mathrm{dg}}_K = \mathcal{O}(K^{-1})$.

Proof We start from (27) in Lemma 11 with $H = F + G$ and $z_k = (p_k, q_k)$:

$$R_K \leq \frac{1}{\eta} \Big(D_H (z_0, z_K) + \widetilde{H}_\eta^{(1)} (z_K) - \widetilde{H}_\eta^{(1)} (z_0) \Big). \tag{40}$$

Note that $D_H (z_0, z_K) = D_F (p_0, p_K) + D_G (q_0, q_K)$. Moreover, by convexity,

$$\begin{aligned} 0 \leq D_F(p_0, p_K) &= F(p_0) - F(p_K) - \langle \nabla F(p_K), p_0 - p_K \rangle \\ &\leq |F(p_0)| + |F(p_K)| + ||\nabla F(p_K)|| \, ||p_K - p_0|| . \end{aligned} \tag{41}$$

By assumption, $|F(p_0)|, |F(p_K)| \leq M$, $||\nabla F(p_K)|| \leq L$, and $||p_K - p_0|| \leq C$. Hence, we can bound (41) as follows:

$$D_F(p_0, p_K) \leq M + M + LC = 2M + LC.$$

The same bound on $D_G (q_0, q_K)$ follows analogously, whence we have

$$D_H (z_0, z_K) \leq 2 (2M + CL) . \tag{42}$$

Next, recall that

$$\widetilde{H}_\eta^{(1)}(p, q) = F(p) + G(q) - \frac{\eta}{2} \langle \nabla F(p), \nabla G(q) \rangle .$$

Thus,

$$\left| \widetilde{H}_\eta^{(1)}(z_K) - \widetilde{H}_\eta^{(1)}(z_0) \right| \leq |F(p_K) - F(p_0)| + |G(q_K) - G(q_0)| \tag{43a}$$

$$+ \frac{\eta}{2} \left| \langle \nabla F(p_K), \nabla G(q_K) \rangle - \langle \nabla F(p_0), \nabla G(q_0) \rangle \right| . \tag{43b}$$

By the mean value inequality,

$$|F(p_K) - F(p_0)| \leq L \, ||p_K - p_0|| \leq LC \quad \text{and} \quad |G(q_K) - G(q_0)| \leq LC, \tag{44}$$

and by applying the triangle and Cauchy-Schwartz inequalities to (43),

$$\begin{aligned} |\langle \nabla F(p_K), \nabla G(q_K)\rangle - \langle \nabla F(p_0), \nabla G(q_0)\rangle| &\leq |\langle \nabla F(p_K), \nabla G(q_K)\rangle| + |\langle \nabla F(p_0), \nabla G(q_0)\rangle| \\ &\leq ||\nabla F(p_K)|| \, ||\nabla G(q_K)|| + ||\nabla F(p_0)|| \, ||\nabla G(q_0)|| \\ &\leq (L)(L) + (L)(L) \\ &= 2L^2. \end{aligned} \tag{45}$$

Hence, using (44) and (45) to bound (43) from above, we have

$$\left| \widetilde{H}_\eta^{(1)}(z_K) - \widetilde{H}_\eta^{(1)}(z_0) \right| \leq LC + LC + \frac{\eta}{2}\left(2L^2\right) = 2LC + \eta L^2. \tag{46}$$

Finally, using (42) and (46) together to bound (40) from above,

$$\boxed{R_K \leq \frac{4}{\eta}\,(M + CL) + L^2.}$$

We then proceed as we did in the proof for Theorem 13 after (37). ■

Corollary 15 shows that when the symplectic Euler iterates remain bounded, the total regret is also bounded under mild assumptions, thereby providing a purely geometric route to proving bounded regret that does not rely on the existence or convergence of the modified Hamiltonian.

Consequently, in the quadratic case, Corollary 15 renders Theorem 14 unnecessary: The presence of bounded iterates already implies bounded regret, even though this case is among the few for which $\widetilde{H}_\eta$ is known to converge absolutely. As mentioned already in Section 6, for other cases, rigorous proofs of convergence for $\widetilde{H}_\eta$ remain scarce, and counterexamples are well documented (Suris, 1989; Field and Nijhoff, 2003; Alsallami et al., 2018). Therefore, identifying general, practically verifiable conditions under which a Hamiltonian H yields a convergent modified Hamiltonian $\widetilde{H}_\eta$ is an important open problem—one whose resolution would make Theorem 14 an analytic foundation for proving bounded regret in AMD.

To conclude, we note that a bounded regret might seem superficially obvious from the start, since if the original Hamiltonian is closed, proper, and coercive convex (e.g., for an ellipsoid), then its level sets are convex and bounded (Rockafellar, 1996, Section 8). However, that is for the Hamiltonian flow generated by the *unperturbed* Hamiltonian H. The implication here is that even the Hamiltonian flow generated by the *modified* Hamiltonian $\widetilde{H}_\eta$ is associated with a bounded regret in discrete time, albeit under assumptions on $\widetilde{H}_\eta$.

10. Conclusion

With regards to energy conservation, there are symplectic integrators which tend to conserve the Hamiltonian better than symplectic Euler for the same stepsize—the leapfrog and implicit midpoint integrators are some examples; see, e.g., Bou-Rabee and Sanz-Serna

(2018) and Chapter I in Hairer et al. (2006) for further details and numerical illustrations. There are also algorithms with better performance than AMD for the setting introduced in Section 4, such as simultaneous proximal mirror descent (Nemirovski, 2004; Wibisono et al., 2022). However, we are mainly interested in studying symplectic Euler because, when the original Hamiltonian is convex, symplectic Euler is the dual space representation of AMD. Moreover, we are interested in studying AMD because it serves as a relatively simple proof of concept for breaching the barrier between the fields of symplectic integration and theoretical computer science. We encourage future studies that use such connections, e.g., by using known symplectic integrators that tend to perform well and seeing if these can be translated to an analogous algorithm in a game-theoretic setting.

Since the field of symplectic integration (let alone Hamiltonian dynamics as a whole) has existed for over three decades, there is no doubt that solid results and rigorous error bounds already exist regarding the trajectories and energy conservation of said methods. And for countless applications in computational physics, chemistry, and biology, these have proved to be sufficient (e.g., Skeel and Hardy, 2001; Engle et al., 2005; Casas and Murua, 2009). However, in the setting of Hamiltonian dynamics as it appears in sampling, optimization, and games, we are more interested in the algorithmic implications and guarantees that such methods might provide. We hope that this paper not only provides a foundation of properties and concepts—a framework—that demonstrates how new and old results in symplectic numerical analysis can be used in such contexts, but also reveals how many more questions remain unanswered.

Acknowledgments

The authors would like to thank John Lazarsfeld, Siddharth Mitra, Yang Cai, Ronald Coifman, and Tiago dos Santos Domingues, amongst several others, for their helpful comments and interest in this work. The authors would also like to thank the Action Editors and anonymous reviewers for their time, diligent attention, and constructive suggestions.

This work was carried out at Yale University with funding from the Yale Graduate School of Arts and Sciences.

Appendix A. Lie algebra and Poisson brackets

Lie algebra. Recall a **Lie algebra** L is a vector space (over $\mathbb{R}$ in our case) with a *Lie bracket* operation $[\,\cdot\,,\,\cdot\,]\colon L \times L \to L$ which is bilinear, anticommutative ($[\ell_1, \ell_2] = -[\ell_2, \ell_1]$ for all $\ell_1, \ell_2 \in L$), and satisfies the Jacobi identity ($[\ell_1, [\ell_2, \ell_3]] + [\ell_2, [\ell_3, \ell_1]] + [\ell_3, [\ell_1, \ell_2]] = 0$ for all $\ell_1, \ell_2, \ell_3 \in L$). For further references on Lie algebra, see, e.g., Jacobson (1979), Varadarajan (1984), and Hall (2015).

Adjoint representation. Recall that for any Lie algebra L equipped with Lie bracket $[\cdot, \cdot]$, for any element $\ell \in L$ we can define the **adjoint representation** or **adjoint action** of ℓ:

$$\mathrm{ad}_\ell\,(\cdot) \coloneqq [\ell, \cdot]$$

as a linear function on L. We also define the k-fold application of adjoint actions ad_ℓ^k by:

$$\begin{aligned}\mathrm{ad}_\ell^1 &= \mathrm{ad}_\ell \\ \mathrm{ad}_\ell^{k+1} &= \mathrm{ad}_\ell^k(\mathrm{ad}_\ell(\cdot)).\end{aligned}$$

With the Poisson bracket as the Lie bracket, this is equivalent to the iterated Poisson bracket, where $\varphi \in C^k(\mathcal{Z})$:

$$\mathrm{ad}_\varphi^k = \{\{\{...\{\cdot, \underbrace{\varphi\}, \varphi\}, \varphi}_{k \text{ times}}\} = \{\cdot, \varphi^k\}.$$

One can check that the Poisson bracket satisfies the axioms for a Lie bracket. Thus, results for Lie algebras also apply to those for a space of C^∞ functions equipped with a Poisson bracket.

Free Lie algebra. Let $F, G \in C^\infty(\mathcal{Z})$. Consider the subspace $\mathcal{L}(F, G)$ of $C^\infty(\mathbb{R}^d)$ of functions generated by the linear combinations of F, G, and their IPBs:

$$\mathcal{L}(F, G) := \mathrm{Span}\left\{\{F^{s_1} G^{r_1} \cdots F^{s_N} G^{r_N}\} : r_1, s_1, \ldots, r_N, s_N \in \mathbb{N}_0, N \in \mathbb{N}\right\}. \tag{48}$$

The vector space $\mathcal{L}(F, G)$ equipped with the Poisson bracket $\{\cdot, \cdot\}$ defines a Lie algebra; this is called the **free Lie algebra** generated by F and G. For a more general definition of free Lie algebra and some properties, see Chapter 3 from Varadarajan (1984).

Exponential map. For some $\eta > 0$, we can formally expand the exponential $\exp(\eta \cdot \mathrm{ad}_\varphi)$ of the adjoint representation ad_φ of some $\varphi \in C^\infty(\mathcal{Z})$ via the repeated application of ad_φ, which can be expressed as IPBs:

$$\exp(\eta \cdot \mathrm{ad}_\varphi) = \sum_{k=0}^\infty \frac{\eta^k \mathrm{ad}_\varphi^k}{k!} = \sum_{k=0}^\infty \frac{\eta^k \{\{\{...\{\cdot, \overbrace{\varphi\}, \varphi\}, \varphi\}, ..., \varphi}^{k}\}}{k!} = \exp(\eta\{\cdot, \varphi\}) \tag{49}$$

See Chapter 2 of Varadarajan (1984) or Chapters 3-5 of Hall (2015) for how the exponential of an adjoint representation (49) generalizes over other Lie algebras.

Appendix B. The modified Hamiltonian and the BCH series

We now discuss how to construct the modified Hamiltonian for the symplectic Euler method via the Baker-Campbell-Hausdorff series. To explain the results, we use the tools from Lie algebra and representation theory introduced in Appendix A.

B.1 Baker-Campbell-Hausdorff series

For any $x, y \in L$, the value of z which solves $\exp(x)\exp(y) = \exp(z)$ can be given via the Dynkin form of the *Baker-Campbell-Hausdorff formula* (Jacobson, 1979):

$$\log(\exp(x)\exp(y)) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} \sum_{\substack{r_1+s_1>0 \\ \vdots \\ r_n+s_n>0}} \frac{\eta^{r_1+\cdots+r_n+s_1+\cdots+s_n-1}\left[x^{r_1}y^{s_1}x^{r_2}y^{s_2}\cdots x^{r_n}y^{s_n}\right]}{(r_1+\cdots+r_n+s_1+\cdots+s_n)\prod_{i=1}^{n} r_i!s_i!}. \tag{50}$$

Here for $r_1, s_1, \ldots, r_n, s_n \in \mathbb{N}_0$ and $n \in \mathbb{N}$, we denote

$$[x^{r_1}y^{s_1}x^{r_2}y^{s_2}\cdots x^{r_n}y^{s_n}] = [\underbrace{x,[x,\cdots[x,}_{r_1}\underbrace{[y,[y,\cdots[y,}_{s_1}\cdots\underbrace{[x,[x,\cdots[x,}_{r_n}\underbrace{[y,[y,\cdots y}_{s_n}]]\cdots]]].$$

See Achilles and Bonfiglioli (2012) and references therein for an overview of different proofs of the BCH formula in various forms.

In particular, when $L = \mathcal{L}(F, G)$ and the Poisson bracket is the Lie bracket, we can define the adjoint representation (Hall, 2015) $\mathrm{ad}_{\varphi}(\cdot) := \{\cdot, \varphi\}$ of *any* function $\varphi \in \mathcal{L}(F, G)$—i.e., the linear map defined by the Poisson bracket of a function in $\mathcal{L}(F, G)$ with φ. Furthermore, instantiating (50) with $\varphi, \psi \in \mathcal{L}(F, G)$ and the Poisson bracket gives the same expression as (50), except with the iterated Lie brackets as IPBs *in reverse order*.[1]

The Dynkin form (21) of the modified Hamiltonian $\widetilde{H}_{\eta}$ is derived as a formal expansion by constructively working from the initial assumptions that there is a Hamiltonian flow (2) generated by some modified Hamiltonian $\widetilde{H}_{\eta}$ which interpolates the iterates of, and is thereby conserved by, the iterates of the symplectic Euler method (5). This aforementioned derivation of (21) is classically known and can be found in several texts—e.g., Yoshida (1990) and Chapters III and IX of Hairer et al. (2006). For sake of completeness, we include a formal derivation of (21) which uses various techniques from Lie algebra in Appendix B.2 of the arXiv version of this paper, Katona et al. (2024).

B.2 The integral form of the BCH series

For the Baker-Campbell-Hausdorff formula on *any* normed Lie algebra $(L, ||\cdot||)$, we propose the following formula equivalent with the Dynkin representation (50) which involves the adjoint representation in a way similar to Theorem 5.5 in Miller (1973) and Theorem 5.3 in Hall (2015):

Lemma 16 *For any $A, B \in L$ sufficiently small in norm—i.e., such that $\left|\left|e^{t\cdot\mathrm{ad}_A}e^{t\cdot\mathrm{ad}_B} - I\right|\right|_{op} < 1$ for all $t \in [0, 1]$—we have*

$$\log\left(e^{A}e^{B}\right) = \int_0^1 \log\left(e^{t\cdot\mathrm{ad}_A}e^{t\cdot\mathrm{ad}_B}\right)\left(e^{t\cdot\mathrm{ad}_A}e^{t\cdot\mathrm{ad}_B} - I\right)^{-1}\left(A + e^{t\cdot\mathrm{ad}_A}B\right)dt. \tag{51}$$

Moreover, the series expansion of (51) converges absolutely.

1. This is because when defining the adjoint representation over a general Lie algebra L vs. $\mathcal{L}(F, G)$, we swap the ordering of the arguments for a general Lie bracket vs. the Poisson bracket, respectively. The derivation of the BCH formula involves interpreting $\exp(x)$ and $\exp(y)$ via their adjoint representations, thereby leading to an expansion in terms of iterated Lie brackets.

Proof We do the computations in reverse order, noting that the steps are reversible after establishing absolute convergence since we did not reorder the series and only rewrote the terms separately using integrals. We start with the BCHD series 50, carefully separate out the sum when $n = 1$ vs. $n \geq 2$, and rewrite the iterated Lie brackets in terms of their adjoint representations:

$$\log\left(e^A e^B\right) = \sum_{r\geq 0} \frac{\mathrm{ad}_A^r B}{(r+1)!} + A + \sum_{n=2}^{\infty} \frac{(-1)^{n-1}}{n} \times \left(\sum_{\substack{r_1+s_1>0 \\ \cdots \\ r_{n-1}+s_{n-1}>0 \\ r_n \geq 0}} \frac{\mathrm{ad}_A^{r_1}\mathrm{ad}_B^{s_1}\cdots\mathrm{ad}_A^{r_{n-1}}\mathrm{ad}_B^{s_{n-1}}\mathrm{ad}_A^{r_n} B}{(r_1+\cdots+r_{n-1}+s_1+\cdots+s_{n-1}+r_n+1)\cdot r_n! \prod_{i=1}^{n-1} r_i! s_i!} + \sum_{\substack{r_1+s_1>0 \\ \cdots \\ r_{n-1}+s_{n-1}>0}} \frac{\mathrm{ad}_A^{r_1}\mathrm{ad}_B^{s_1}\cdots\mathrm{ad}_A^{r_{n-1}}\mathrm{ad}_B^{s_{n-1}} A}{(r_1+\cdots+r_{n-1}+s_1+\cdots+s_{n-1}+1)\cdot \prod_{i=1}^{n-1} r_i! s_i!} \right)$$

This further equals to:

$$\log(e^A e^B) = \sum_{r\geq 0} \int_0^1 \frac{(t\cdot \mathrm{ad}_A)^r B}{r!} dt + A + \sum_{n=2}^{\infty} \frac{(-1)^{n-1}}{n} \times \left(\sum_{\substack{r_1+s_1>0 \\ \cdots \\ r_{n-1}+s_{n-1}>0 \\ r_n \geq 0}} \int_0^1 \frac{(t\cdot \mathrm{ad}_A)^{r_1}}{r_1!} \frac{(t\cdot \mathrm{ad}_B)^{s_1}}{s_1!} \cdots \frac{(t\cdot \mathrm{ad}_A)^{r_{n-1}}}{r_{n-1}!} \frac{(t\cdot \mathrm{ad}_B)^{s_{n-1}}}{s_{n-1}!} \frac{(t\cdot \mathrm{ad}_A)^{r_n}}{r_n!} dt B + \sum_{\substack{r_1+s_1>0 \\ \cdots \\ r_{n-1}+s_{n-1}>0}} \int_0^1 \frac{(t\cdot \mathrm{ad}_A)^{r_1}}{r_1!} \frac{(t\cdot \mathrm{ad}_B)^{s_1}}{s_1!} \cdots \frac{(t\cdot \mathrm{ad}_A)^{r_{n-1}}}{r_{n-1}!} \frac{(t\cdot \mathrm{ad}_B)^{s_{n-1}}}{s_{n-1}!} dt A \right). \tag{52}$$

Under absolute convergence, we can then swap the integrals and summations in (52) and then rewrite the result using the exponential Taylor series $e^X = \sum_{r\geq 0} X^r/r!$ as follows:

$$\begin{aligned}
\log\left(e^A e^B\right) = &\int_0^1 e^{t\cdot \mathrm{ad}_A} B dt + A \\
&+ \sum_{n=2}^{\infty} \frac{(-1)^{n-1}}{n} \left(\int_0^1 \left(\sum_{r+s>0} \frac{(t\cdot \mathrm{ad}_A)^r}{r!} \frac{(t\cdot \mathrm{ad}_B)^s}{s!} \right)^{n-1} \sum_{r_n\geq 0} \frac{(t\cdot \mathrm{ad}_A)^{r_n}}{r_n!} dt B \right. \\
&+ \left. \int_0^1 \left(\sum_{r+s>0} \frac{(t\cdot \mathrm{ad}_A)^r}{r!} \frac{(t\cdot \mathrm{ad}_B)^s}{s!} \right)^{n-1} dt A \right) \\
= &\int_0^1 \left(e^{t\cdot \mathrm{ad}_A} B + A \right) dt \\
&+ \sum_{n=2}^{\infty} \frac{(-1)^{n-1}}{n} \left(\int_0^1 \left(e^{t\cdot \mathrm{ad}_A} e^{t\cdot \mathrm{ad}_B} - I \right)^{n-1} \left(e^{t\cdot \mathrm{ad}_A} B + A \right) dt \right) \\
= &\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} \int_0^1 \left(e^{t\cdot \mathrm{ad}_A} e^{t\cdot \mathrm{ad}_B} - I \right)^{n-1} \left(e^{t\cdot \mathrm{ad}_A} B + A \right) dt. \qquad (53)
\end{aligned}$$

Finally, under absolute convergence again and the series $\log(X+I) = \sum_{r\geq 1} (-1)^{r+1} X^r/r$ (which converges absolutely whenever $\|X\|_{\mathrm{op}} < 1$), (53) becomes

$$\begin{aligned}
&\int_0^1 \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} \left(e^{t\cdot \mathrm{ad}_A} e^{t\cdot \mathrm{ad}_B} - I \right)^{n-1} \left(e^{t\cdot \mathrm{ad}_A} B + A \right) dt = \\
&\boxed{\int_0^1 \log\left(e^{t\cdot \mathrm{ad}_A} e^{t\cdot \mathrm{ad}_B} \right) \left(e^{t\cdot \mathrm{ad}_A} e^{t\cdot \mathrm{ad}_B} - I \right)^{-1} \left(e^{t\cdot \mathrm{ad}_A} B + A \right) dt,} \qquad (54)
\end{aligned}$$

as we wanted to show. Note finally that absolute convergence of the series expansion of (54) is ensured by the bound on the exponentials of the adjoints stated in Lemma 16, thereby ensuring absolute convergence for the original BCH series. ■

We can further express the integrand in (51) via a single series, as demonstrated in the corollary below:

Lemma 17 *Given the same setting as Lemma 16, we have*

$$\log\left(e^A e^B\right) = \int_0^1 \sum_{j=0}^{\infty} \frac{\left(I - e^{t\cdot \mathrm{ad}_A} e^{t\cdot \mathrm{ad}_B}\right)^j}{j+1} \left(A + e^{t\cdot \mathrm{ad}_A} B \right) dt.$$

Proof From p. 161 of Miller (1973), we have the following series expansion for $\log(X)(X-I)^{-1}$ about $X = I$:

$$\log(X)(X-I)^{-1} = \sum_{j=0}^{\infty} \frac{(I-X)^j}{j+1}, \qquad (55)$$

which converges whenever $||X - I||_{\text{op}} < 1$. Then, use (55) to expand the integrand of (51) in Lemma 16. ■

Finally, equip $\mathcal{L}(F, G)$ with some function norm $||\cdot||$ which is continuous over the space. For all $\varphi \in \mathcal{L}(F, G)$, we define the operator norm of φ, $||\varphi||_{\text{op}}$, as the smallest $L > 0$ satisfying $||\{\varphi, \psi\}|| \leq L\,||\psi||$ for all $\psi \in \mathcal{L}(F, G)$. Using this definition, we propose an equivalent integral form of the Dynkin representation (21) of the BCH formula:

Theorem 18 *For any $F \in C^\infty(\mathcal{P})$ and $G \in C^\infty(\mathcal{Q})$ such that $\left|\left|e^{t\{\cdot,G\}}e^{t\{\cdot,F\}} - I\right|\right|_{op} < 1$ for all $t \in [0, 1]$,*

$$\widetilde{H}_\eta(p, q) = \frac{1}{\eta}\int_0^\eta \sum_{j=0}^{\infty} \frac{\left(I - e^{t\{\cdot,G\}}e^{t\{\cdot,F\}}\right)^j}{j+1}\left(G + e^{t\{\cdot,G\}}F\right) dt. \tag{56}$$

Moreover, the series expansion of (56) *converges absolutely.*

Proof Instantiate the Lie bracket in Lemma 17 with the Poisson bracket multiplied by η. ■

The integral form derived above in Theorem 18 will allow us to compute $\widetilde{H}$ directly in the quadratic case, as we will see in Appendix C.

Appendix C. The modified Hamiltonian in the quadratic case

As a test case to demonstrate the potency of the integral form of the MH (Theorem 18), we now study the particular case when $F = ap^2, G = bq^2$, and $d = 1$—the quadratic case. In this case, the function $\mathsf{S}(p, q) = ap^2 + bq^2 - 2\eta abpq$ is a closed-form conserved quantity throughout the iterations of the symplectic Euler method (5), i.e., $\mathsf{S}(p_n, q_n) = \mathsf{S}(p_0, q_0)$ for all $n \in \mathbb{N}_0$ and any stepsize $\eta > 0$. We encourage readers to verify this fact for themselves or see Bailey et al. (2020) and Wibisono et al. (2022) for further details, which demonstrate how the above results also generalize for $d > 1$.

In theory, the function $\mathsf{S}(p, q)$ should also appear in the modified Hamiltonian. And this is indeed the case: By applying symbolic calculation programs to calculate the first terms in the modified Hamiltonian, expressed as the series (21) when $f = ap^2$, $g = bq^2$, we reveal the following pattern in the first few terms of the series:

$$\widetilde{H}_\eta(p, q) = \mathsf{S}(p, q)\left(1 + \frac{2ab\eta^2}{3} + \frac{8(ab)^2\eta^4}{15} + \frac{16(ab)^3\eta^6}{35} + \frac{128(ab)^4\eta^8}{215} + \cdots\right). \tag{57}$$

This pattern encourages us to calculate the closed-form for this series using the integral form of the BCH formula (Theorem 18) directly. To the best of our knowledge, this has not been done in prior literature.

Furthermore, as (57) implies, the modified Hamiltonian in this case is the product of a conserved quantity as a function of p, q, and a power series of η. For η large enough, the power series diverges, yet the conserved quantity still remains conserved along the iterations. Thus, this is a direct example demonstrating that a conserved quantity could persist even when the Dynkin series (21) diverges.

C.1 Functional space of the Iterated Poisson Brackets

We now apply Theorem 18 to compute the modified Hamiltonian in this case. We first have the following lemma regarding the structure of the Iterated Poisson Brackets (IPBs):

Lemma 19 *F, G, and any IPBs generated by F and G has the following quadratic form:*

$$\{\{\{...\{F,G\},F\},G\},...,\} = Ap^2 + Bpq + Cq^2 = \begin{bmatrix} p^2 & pq & q^2 \end{bmatrix} \begin{bmatrix} A \\ B \\ C \end{bmatrix}$$

for some constants $A, B, C \in \mathbb{R}$ dependent on the number of Poisson brackets and the permutation of the functions F, G in the IPB.

Proof This can be shown via induction. Since $\{G,F\} = -\{F,G\} = 4adpq$, the base case holds. Suppose this is true for an IPB I. Then, $\{F,I\} = -\nabla_p F \frac{\partial I}{\partial q} = -(2ap)(2Cq+Bp) = -2aBp^2 + 4aCpq$. Similarly, $\{G,I\} = \nabla_q G \frac{\partial I}{\partial p} = (2dq)(2Ap+Bq) = 2Bdq^2 + 4dApq$. ■

Thus, the functional space endowed by the Poisson bracket in this case is simply the space of pure second-order polynomials of p and q. In this case, there exists a matrix expression of the adjoint representations $\mathrm{ad}_F \coloneqq \{\cdot, F\}$ and $\mathrm{ad}_G \coloneqq \{\cdot, G\}$ on the free Lie algebra $\mathcal{L}(F,G)$ generated by F and G and endowed with the Poisson bracket, as shown in the following lemma.

Lemma 20 *Given the previous representation, we have the matrix representations*

$$\mathrm{ad}_F = \begin{bmatrix} 0 & 0 & 2a \\ 0 & 0 & 0 \\ 0 & 4a & 0 \end{bmatrix}, \quad \mathrm{ad}_G = -\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 2b \\ 4b & 0 & 0 \end{bmatrix}$$

when we choose the basis $\left\{p^2, q^2, pq\right\}$ for $\mathcal{L}(F,G)$.

Proof From Lemma 19, $\varphi \in \mathrm{Span}\left\{p^2, q^2, pq\right\} \Rightarrow \{\varphi, F\}, \{\varphi, G\} \in \mathrm{Span}\left\{p^2, q^2, pq\right\}$, but we desire to express this implication more explicitly. Let $\varphi \in \mathrm{Span}\left\{p^2, q^2, pq\right\}$, in which case $\varphi(p,q) = c_2p^2 + c_1q^2 + c_0pq$ for some constants $c_2, c_1, c_0 \in \mathbb{R}$, which can be represented in $\mathcal{L}(F,G)$ as (c_2, c_1, c_0) when using the basis $\left\{p^2, q^2, pq\right\}$. Routine computations show that

$$\left\{\varphi, p^2\right\} = 2c_0p^2 + 4c_1pq, \quad \left\{\varphi, q^2\right\} = -2c_0q^2 - 4c_2pq, \quad \{\varphi, pq\} = 2c_1q^2 - 2c_2p^2. \tag{58}$$

Since the Poisson bracket is linear in each component, (58) implies that

$$\begin{aligned} \mathrm{ad}_F(\varphi) &= \{\varphi, F\} = \left\{\varphi, ap^2\right\} = (2a)\, c_0p^2 + (4a)\, c_1pq, \\ \mathrm{ad}_G(\varphi) &= \{\varphi, G\} = \left\{\varphi, dq^2\right\} = (-2d)\, c_0q^2 + (-4d)\, c_2pq, \end{aligned}$$

i.e.,

$$\mathrm{ad}_F(c_2, c_1, c_0) = (2ac_0, 0, 4ac_1), \qquad \mathrm{ad}_G(c_2, c_1, c_0) = -(0, 2dc_0, 4dc_2)$$

when expressed in the basis $\left\{p^2, q^2, pq\right\}$. ■

Note that the existence of matrix expressions for ad_F and ad_G should already be expected from Ado's theorem (Hall, 2015), which states that every finite-dimensional real Lie algebra (e.g., $\mathcal{L}(F, G)$ for F and G quadratic) is isomorphic to an algebra of matrices.

Since the adjoint matrix in this case is of finite dimension and nilpotent, we are motivated to compute the MH in this quadratic case by evaluating the formula proposed in Theorem 18 directly. Doing so gives us Theorem 12 and the observations that follow thereafter.

C.2 Proof of Theorem 12

C.2.1 Preliminary work

We start by manipulating Theorem 18 to make calculations a bit easier. Recall that in the last section, we derived the matrix representation for ad_G and ad_F. Hence, if we have access to the diagonalization of $e^{t\{\cdot,G\}}e^{t\{\cdot,F\}} = e^{t\cdot\mathrm{ad}_G}e^{t\cdot\mathrm{ad}_F} = PDP^{-1}$, then we could simplify the integral form of the BCH series (56) into the following form:

$$\widetilde{H}_\eta(p, q) = \frac{1}{\eta}\int_0^\eta P\sum_{j=0}^{\infty}\frac{(I-D)^j}{j+1}P^{-1}\left(G + e^{t\cdot\mathrm{ad}_G}F\right)dt. \tag{59}$$

Motivated by the aforementioned simplification, we start by calculating the matrix exponentials of the matrix expression of the adjoint representations and calculate its diagonalization. With the help of SymPy (Meurer et al., 2017), we have the following:

$$e^{t\cdot\mathrm{ad}_G}e^{t\cdot\mathrm{ad}_F} = \begin{bmatrix} 1 & 8a^2t^2 & 2at \\ 8b^2t^2 & 1 + 8abt^2 + 64a^2b^2t^4 & 2bt + 16a^2bt^3 \\ -4bt & 4at - 32a^2bt^3 & 1 - 8abt^2 \end{bmatrix}. \tag{60}$$

With some help from SymPy (Meurer et al., 2017), we now calculate P and D in the diagonalization $e^{t\cdot\mathrm{ad}_G}e^{t\cdot\mathrm{ad}_F} = PDP^{-1}$:

$$P = \begin{bmatrix} -\frac{1}{2bt} & \frac{-abt-i\sqrt{ab(1-abt^2)}}{2b} & \frac{-abt+i\sqrt{ab(1-abt^2)}}{2b} \\ -\frac{1}{2at} & \frac{-abt+i\sqrt{ab(1-abt^2)}}{2a} & \frac{-abt-i\sqrt{ab(1-abt^2)}}{2a} \\ 1 & 1 & 1 \end{bmatrix}. \tag{61}$$

Here $i = \sqrt{-1}$ is the imaginary unit. We can thereby compute $P^{-1}(G + e^{t\cdot\mathrm{ad}_G}F)$ as follows:

$$P^{-1}(G + e^{t\cdot\mathrm{ad}_G}F) = \begin{bmatrix} \frac{2abt}{abt^2-1} \\ \frac{abt\left(3abt-4a^2b^2t^3+i\left(4abt^2-1\right)\sqrt{ad(1-abt^2)}\right)}{a^2b^2t^3-abt+i(1-abt^2)\sqrt{ab(1-abt^2)}} \\ i\frac{abt\left(2abt-2a^2b^2t^3-i\left(2abt^2-1\right)\sqrt{ab(1-abt^2)}\right)}{\sqrt{ab(1-abt^2)}(1-abt^2)} \end{bmatrix}. \tag{62}$$

Furthermore,

$$D = \mathrm{diag}\left(\begin{bmatrix} 1 \\ 8a^2b^2t^4 - 8abt^2 - 4t\sqrt{ab\left(abt^2-1\right)}\left(2abt^2-1\right) + 1 \\ 8a^2b^2t^4 - 8abt^2 + 4t\sqrt{ab\left(abt^2-1\right)}\left(2abt^2-1\right) + 1 \end{bmatrix}\right). \tag{63}$$

Let $g(t) := \sqrt{ab(1-abt^2)}$. Note that by assumption, $t \leq \eta < 1/\sqrt{ab}$, which implies $t^2 < 1/ab$, or $g(t) \in \mathbb{R}$. Hence, we rewrite (63) and use it to compute $\sum_{j=0}^{\infty} \frac{(I-D)^j}{j+1}$:

$$\sum_{j=0}^{\infty} \frac{(I-D)^j}{j+1} = \operatorname{diag}\left(\begin{bmatrix} 1 \\ \sum_{j=0}^{\infty} \frac{4^j t^j}{j+1}\left(2tg^2(t) + i(2abt^2-1)g(t)\right)^j \\ \sum_{j=0}^{\infty} \frac{4^j t^j}{j+1}(2tg^2(t) - i(2abt^2-1)g(t))^j \end{bmatrix}\right). \tag{64}$$

Using the Taylor series for $-\log(1-x) = \sum_{k=1}^{\infty} \frac{x^k}{k} = \sum_{j=0}^{\infty} \frac{x^{j+1}}{j+1} = x\sum_{j=0}^{\infty} \frac{x^j}{j+1}$, we can then sum (64) as follows:

$$-\operatorname{diag}\left(\begin{bmatrix} -1 \\ \frac{1}{4t(2tg^2(t)+i(2abt^2-1)g(t))} \log\left[1-4t\left(2tg^2(t) + i(2abt^2-1)g(t)\right)\right] \\ \frac{1}{4t(2tg^2(t)-i(2abt^2-1)g(t))} \log\left[1-4t\left(2tg^2(t) - i(2abt^2-1)g(t)\right)\right] \end{bmatrix}\right), \tag{65}$$

Note in particular that this means we are using the principal branch of the logarithm.

C.2.2 Simplification of the integrand

We begin by multiplying (65) and (62) and simplifying to get:

$$\sum_{j=0}^{\infty} \frac{(I-D)^j}{j+1} P^{-1}(G + e^{tab_G}F) = \frac{a^2b^2}{4ig^3(t)} \begin{bmatrix} -8itg(t) \\ -\log\left[1-4t\left(2tg^2(t) + i\left(2abt^2-1\right)g(t)\right)\right] \\ \log\left[1-4t\left(2tg^2(t) - i\left(2abt^2-1\right)g(t)\right)\right] \end{bmatrix}. \tag{66}$$

Let $z(t) := abt + ig(t)$ and $\nu(t) := \log\left[1-4t\left(2tg^2(t) - i\left(2abt^2-1\right)g(t)\right)\right]$. Then, using (61) and (66) to assemble the entire integrand of (59), we have

$$P\sum_{j=0}^{\infty} \frac{(I-D)^j}{j+1} P^{-1}(G + e^{t\cdot \mathrm{ad}_G}F) = \frac{a^2b^2}{2g^3(t)} \begin{bmatrix} \frac{1}{2b}\Im\left\{z(t)\overline{\nu(t)}\right\} + \frac{2g(t)}{b} \\ \frac{1}{2a}\Im\left\{\overline{z(t)\nu(t)}\right\} + \frac{2g(t)}{a} \\ \Im\{\nu(t)\} - 4tg(t) \end{bmatrix}, \tag{67}$$

where $\Im(c)$ is the imaginary component of any $c \in \mathbb{C}$. Noticing the recurrence of the expression in (67), we are motivated to expand out $\nu(t)$ to

$$\nu(t) = \log\left[1 - 8abt^2(1-abt^2) + 4it(2abt^2-1)\sqrt{ab(1-abt^2)}\right]. \tag{68}$$

By routine algebraic manipulations, we show that the complex number within the logarithm of (68) has modulus one:

$$(1-8abt^2(1-abt^2))^2 + (4t(2abt^2-1)\sqrt{ab(1-abt^2)})^2 = 1. \tag{69}$$

Thus, in light of (68) and (69), we can use Euler's formula, double angle formulae, and then trigonometry to rewrite $\nu(t)$ as the following:

$$\nu(t) = i\arcsin(4t(2abt^2-1)\sqrt{ab(1-abt^2)}) = 4i\arcsin(\sqrt{abt^2}). \tag{70}$$

Hence, by (67) and (70), we can simplify the integrand of (59) as follows:

$$\widetilde{H}_\eta(p,q) = \frac{a^2b^2}{\eta}\begin{bmatrix}p^2 & q^2 & pq\end{bmatrix}\int_0^\eta \frac{1}{2g^3(t)}\begin{bmatrix}2at\arcsin(\sqrt{abt^2}) + \frac{1}{b}2g(t)\\ 2bt\arcsin(\sqrt{abt^2}) + \frac{1}{a}2g(t)\\ -4\arcsin(\sqrt{abt^2}) - 4tg(t)\end{bmatrix}dt. \tag{71}$$

Finally, we split the integrand in (71) into the two terms shown in the vector and integrate them separately. Doing so, adding the results, and writing out the dot product gives us the following:

$$\widetilde{H}_\eta(p,q) = \frac{1}{\sqrt{ab\eta^2(1-ab\eta^2)}}\arcsin(\sqrt{ab\eta^2})(ap^2 + bq^2 - 2abpq\eta). \tag{72}$$

When $ab < 0$, we first show the following proper definition of inverse sine functions. Suppose we are interested in finding $y = \arcsin ix$, which we can write equivalently as $y = \arcsin ix = i\operatorname{arcsinh}(-x)$. Then the modified Hamiltonian takes the following form:

$$\frac{\arcsin(\sqrt{ab\eta^2})}{\sqrt{ab\eta^2(1-ab\eta^2)}}(ap^2 + bq^2 - 2abpq\eta) = \frac{\operatorname{arcsinh}(\sqrt{-ab\eta^2})}{\sqrt{-ab\eta^2(1-ab\eta^2)}}(ap^2 + bq^2 - 2abpq\eta).$$

C.2.3 Convergence radius

The argument above is only valid if the conditions of Theorem 18 are actually satisfied. Thus, we need to ensure when $\left|\left|e^{t\cdot\mathrm{ad}_G}e^{t\cdot\mathrm{ad}_F} - I\right|\right|_{\mathrm{op}} = \sigma_{\max}(D - I) < 1$, or equivalently, when

$$\left|-8abt^2(1-abt^2) - 4it(2abt^2-1)\sqrt{ab(1-abt^2)}\right| < 1, \tag{73a}$$

$$\left|-8abt^2(1-abt^2) + 4it(2abt^2-1)\sqrt{ab(1-abt^2)}\right| < 1 \tag{73b}$$

for all $t \in [0,\eta]$, as taken directly from from (63). Noting the equivalence of forms (68) and (70) for $\nu(t)$, (73) can be rewritten as

$$\left|e^{\overline{\nu(t)}} - 1\right| = \left|e^{-4i\arcsin z} - 1\right| < 1,$$
$$\left|e^{\nu(t)} - 1\right| = \left|e^{4i\arcsin z} - 1\right| < 1,$$

respectively, where we have let $z := \sqrt{abt^2}$. From here, we split the work into two cases based on the sign of ab.

Case 1: $ab \geq 0$. In this case, $\arcsin z$ is real, and since arcsin only takes arguments from -1 to 1, we get a contradiction when $z > 1$. Since $e^{-4i\arcsin z} = \overline{e^{4i\arcsin z}}$, $|e^{-4i\arcsin z} - 1| < 1$ if and only if $|e^{4i\arcsin z}| < 1$. When $0 < 4\arcsin z < \frac{\pi}{3}$, we have that $|e^{4i\arcsin z} - 1| < 1$. Similarly, when $\frac{5\pi}{3} < 4\arcsin z < 2\pi$, $|e^{4i\arcsin z} - 1| < 1$ again. Thus, the absolute convergence criteria in Theorem 18 is satisfied whenever $z \in \big(0, \sin(\frac{\pi}{12})\big) \cup \big(\sin(\frac{5\pi}{12}), 1\big)$. We now show that this actually implies that the modified Hamiltonian (21) is absolutely convergent whenever $|z| \in (0,1)$.

For any $\eta \in (0, \sqrt{\frac{1}{ab}})$, we pick $\eta' = \max(\sqrt{\sin\frac{11\pi}{24}}, \eta)$. By construction $\eta' \in \left(\sin(\frac{5\pi}{12}), 1\right)$ and thus the series (21) converges absolutely for η'. We now compare the terms of the two series:

$$\left|\frac{\eta^{r_1+\cdots+r_n+s_1+\cdots+s_n-1}\left\{G^{r_1}F^{s_1}\cdots G^{r_n}F^{s_n}\right\}(p,q)}{n\left(r_1+\cdots+r_n+s_1+\cdots+s_n\right)\prod_{i=1}^n r_i!s_i!}\right| \leq \left|\frac{(\eta')^{r_1+\cdots+r_n+s_1+\cdots+s_n-1}\left\{G^{r_1}F^{s_1}\cdots G^{r_n}F^{s_n}\right\}(p,q)}{n\left(r_1+\cdots+r_n+s_1+\cdots+s_n\right)\prod_{i=1}^n r_i!s_i!}\right|.$$

By the dominated convergence theorem, the modified Hamiltonian (21) converges absolutely for η chosen above.

Case 2: $ab<0$. When $ab < 0$, we let $F' = |a|p^2, G' = |b|q^2$. From Appendix C.1 we know that $\left\{G^{r_1}F^{s_1}\cdots G^{r_n}F^{s_n}\right\}(p,q)$ is always a monomial of p, q and a, b, regardless of the signs of p and q. We have the following equality:

$$\left|\frac{\eta^{r_1+\cdots+r_n+s_1+\cdots+s_n-1}\left\{G^{r_1}F^{s_1}\cdots G^{r_n}F^{s_n}\right\}(p,q)}{n\left(r_1+\cdots+r_n+s_1+\cdots+s_n\right)\prod_{i=1}^n r_i!s_i!}\right| = \left|\frac{\eta^{r_1+\cdots+r_n+s_1+\cdots+s_n-1}\left\{G'^{r_1}F'^{s_1}\cdots G'^{r_n}F'^{s_n}\right\}(p,q)}{n\left(r_1+\cdots+r_n+s_1+\cdots+s_n\right)\prod_{i=1}^n r_i!s_i!}\right|.$$

Since the modified Hamiltonian (21) converges absolutely whenever $\sqrt{|a||b|t^2} < 1$, it converges absolutely for F' and G'. Since the absolute value of the terms is the same, the convergence is also absolute when $ab < 0$, but except when $\sqrt{|ab|t^2} < 1$.

Figure 2 shows the graph of the function $\eta \mapsto \frac{\arcsin\left(\sqrt{ab\eta^2}\right)}{\sqrt{ab\eta^2(1-ab\eta^2)}}$ that appears in the final expression (72) of the modified Hamiltonian. Note that the Taylor expansion of $\eta \mapsto \frac{\arcsin\left(\sqrt{ab\eta^2}\right)}{\sqrt{ab\eta^2(1-ab\eta^2)}}$ is as follows:

$$\frac{\arcsin\left(\sqrt{ab\eta^2}\right)}{\sqrt{ab\eta^2(1-ab\eta^2)}} = 1 + \frac{2(\eta\sqrt{ab})^2}{3} + \frac{8(\eta\sqrt{ab})^4}{15} + \frac{16(\eta\sqrt{ab})^6}{35} + \frac{128(\eta\sqrt{ab})^8}{215} + \cdots,$$

which is precisely what the simplified series formula (57) for $\widetilde{H}_\eta$ suggests. One may further check that the same expansion holds when $ab < 0$.

C.3 Verification: Hamiltonian flow of modified Hamiltonian

In this subsection, we verify Theorem 7 in the single-variable quadratic case. To this end, we have the following lemma:

Lemma 21 *Define the modified Hamiltonian as*

$$\tilde{H}_\eta(p,q) = \frac{\arcsin(\sqrt{ab\eta^2})}{\sqrt{ab\eta^2(1-ab\eta^2)}}(ap^2 + bq^2 - 2abpq\eta).$$

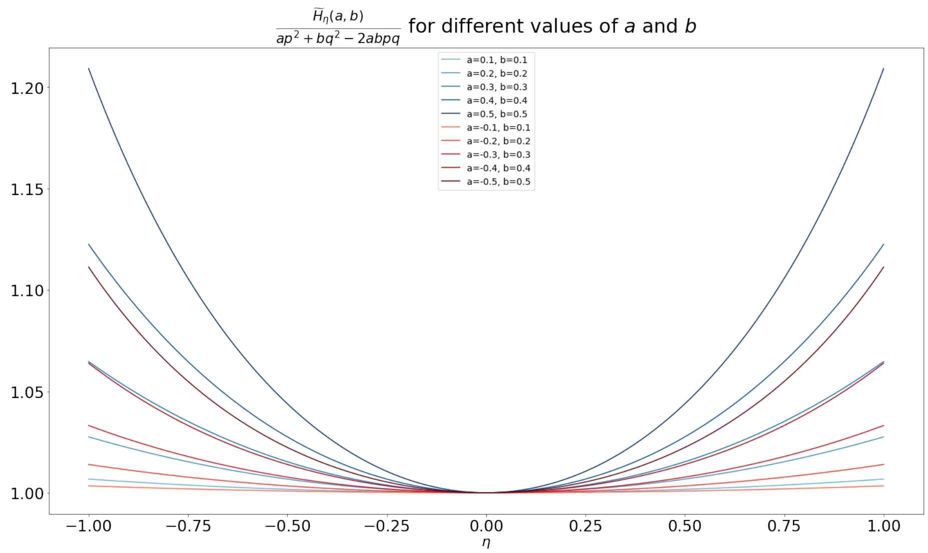

Figure 2: The function $\eta \mapsto \frac{\arcsin\left(\sqrt{ab\eta^2}\right)}{\sqrt{ab\eta^2(1-ab\eta^2)}}$ with different choices of a and b.

If $(p(t), q(t))$ follows the Hamiltonian flow of the modified Hamiltonian:

$$\frac{d}{dt}\begin{pmatrix} p(t) \\ q(t) \end{pmatrix} = \begin{pmatrix} -\frac{\partial}{\partial q}\widetilde{H}_\eta(p(t), q(t)) \\ \frac{\partial}{\partial p}\widetilde{H}_\eta(p(t), q(t)) \end{pmatrix},$$

and starts at

$$\begin{pmatrix} p(0) \\ q(0) \end{pmatrix} = \begin{pmatrix} p_0 \\ q_0 \end{pmatrix},$$

then we have:

$$\begin{pmatrix} p(\eta) \\ q(\eta) \end{pmatrix} = \begin{pmatrix} 1 & -2b\eta \\ 2a\eta & 1-4ab\eta^2 \end{pmatrix}\begin{pmatrix} p_0 \\ q_0 \end{pmatrix},$$

which coincides with the iterates of symplectic Euler (5) *when* $H(p,q) = ap^2 + bq^2$.

Proof For $\eta > 0$, let

$$C_\eta = \frac{\arcsin(\sqrt{ab\eta^2})}{\sqrt{ab\eta^2(1-ab\eta^2)}}$$

so that the modified Hamiltonian above becomes:

$$\tilde{H}_\eta(p,q) = C_\eta(ap^2 + bq^2 - 2abpq\eta).$$

The Hamiltonian flow is:

$$\frac{d}{dt}\begin{pmatrix} p(t) \\ q(t) \end{pmatrix} = \begin{pmatrix} 2C_\eta b(-q(t) + a\eta p(t)) \\ 2C_\eta a(p(t) - b\eta q(t)) \end{pmatrix} = 2C_\eta \begin{pmatrix} ab\eta & -b \\ a & -ab\eta \end{pmatrix}\begin{pmatrix} p(t) \\ q(t) \end{pmatrix}. \tag{74}$$

Using the matrix exponential, the solution to the Hamiltonian flow (74) at time $t = \eta$ is:

$$\begin{pmatrix} p(\eta) \\ q(\eta) \end{pmatrix} = \exp\left(2\eta C_\eta \begin{pmatrix} ab\eta & -b \\ a & -ab\eta \end{pmatrix}\right)\begin{pmatrix} p_0 \\ q_0 \end{pmatrix}. \tag{75}$$

We compute the matrix exponential in (75) as follows:

$$\exp\left(2\eta C_\eta \begin{pmatrix} ab\eta & -b \\ a & -ab\eta \end{pmatrix}\right) = \begin{bmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{bmatrix},$$

where the coefficients are as follows:

$$\begin{aligned}
E_{11} = {} & \frac{\left(ab\eta + \sqrt{ab(ab\eta^2-1)}\right)\exp\left(2\sqrt{ab(ab\eta^2-1)}\eta C_\eta\right)}{2\sqrt{ab(ab\eta^2-1)}} \\
& - \frac{\left(ab\eta - \sqrt{ab(ab\eta^2-1)}\right)\exp\left(-2\sqrt{ab(ab\eta^2-1)}\eta C_\eta\right)}{2\sqrt{ab(ab\eta^2-1)}} \\
E_{12} = {} & \frac{\sqrt{b}\exp\left(-2\sqrt{ab(ab\eta^2-1)}\eta C_\eta\right)}{2\sqrt{a(ab\eta^2-1)}} - \frac{\sqrt{b}\exp\left(2\sqrt{ab(ab\eta^2-1)}\eta C_\eta\right)}{2\sqrt{a(ab\eta^2-1)}} \\
E_{21} = {} & \frac{\sqrt{a}\exp\left(2\sqrt{ab(ab\eta^2-1)}\eta C_\eta\right)}{2\sqrt{b(ab\eta^2-1)}} - \frac{\sqrt{a}\exp\left(-2\sqrt{ab(ab\eta^2-1)}\eta C_\eta\right)}{2\sqrt{b(ab\eta^2-1)}} \\
E_{22} = {} & \frac{\left(-ab\eta + \sqrt{ab(ab\eta^2-1)}\right)\exp\left(2\sqrt{ab(ab\eta^2-1)}\eta C_\eta\right)}{2\sqrt{ab(ab\eta^2-1)}} \\
& - \frac{\left(-ab\eta - \sqrt{ab(ab\eta^2-1)}\right)\exp\left(-2\sqrt{ab(ab\eta^2-1)}\eta C_\eta\right)}{2\sqrt{ab(ab\eta^2-1)}}.
\end{aligned}$$

To simplify the coefficients above, we let $S = \sqrt{ab\eta^2(1-ab\eta^2)}$, $T = ab\eta^2$. We further calculate:

$$\begin{aligned}
\exp\left(2\sqrt{ab(ab\eta^2-1)}\eta C_\eta\right) &= \exp\left(2\sqrt{ab(ab\eta^2-1)}\eta \frac{\arcsin(\sqrt{ab\eta^2})}{\sqrt{ab\eta^2(1-ab\eta^2)}}\right) \\
&= \exp\left(2i\arcsin\left(\sqrt{ab\eta^2}\right)\right) \\
&= 1 - 2T + 2iS
\end{aligned}$$

Similarly, we have:

$$\exp\left(-2\sqrt{ab\left(ab\eta^2-1\right)}\eta C_\eta\right) = 1 - 2T - 2iS.$$

Hence, we could simplify the coefficients in the following way:

$$\begin{pmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{pmatrix} = \begin{pmatrix} 1 & -2b\eta \\ 2a\eta & 1-4ab\eta^2 \end{pmatrix}.$$

The symplectic Euler algorithm, (5), in this case is:

$$\begin{pmatrix} p_{\text{new}} \\ q_{\text{new}} \end{pmatrix} = \begin{pmatrix} p_0 - \eta(2bq_0) \\ q_0 + \eta(2ap_{\text{new}}) \end{pmatrix} = \begin{pmatrix} p_0 - 2b\eta q_0 \\ q_0 + 2a\eta\left(p_0 - 2b\eta q_0\right) \end{pmatrix} = \begin{pmatrix} 1 & -2b\eta \\ 2a\eta & 1-4ab\eta^2 \end{pmatrix}\begin{pmatrix} p_0 \\ q_0 \end{pmatrix}.$$

Since the coefficients match, we confirm that the iterations are indeed interpolated by the Hamiltonian flow generated by the modified Hamiltonian $\widetilde{H}_\eta$. ■

C.4 The multivariate case

In this subsection we derive the closed-form of the modified Hamiltonian in the multivariate quadratic case by utilizing results from monovariate case. The general idea of the proof is the following: we first derive connections between the iterated Poisson brackets in multivariate and univariate cases. We then reorder the series form (21) of the modified Hamiltonian in the multivariate case $\widetilde{H}_{\eta}^{F,G}(p,q)$ and relate the reordered series with the series form of the modified Hamiltonian in the univariate case to get closed-form expression of $\widetilde{H}_{\eta}^{F,G}(p,q)$. Furthermore, this allows us to determine the absolute convergence radius of the reordered form, which then implies equivalence between the original form and the reordered form.

C.4.1 Notations and settings

We assume that $p, q \in \mathbb{R}^d$, where $d \geq 2$. We assume that $B, C \in \mathbb{R}^{d\times d}$ are symmetric matrices such that BC is positive semi-definite. Let $Q^{-1}\Lambda Q$ be the diagonalization of BC. We define $F(p) = p^\top Bp$, $G(q) = q^\top Cq$, and $H^{F,G}(p,q) = F(p) + G(q)$. In using results from the last subsection, we assume that $u, v \in \mathbb{R}$, and define $f(u) = au^2, g(v) = bv^2$, where $ab > 0$, and let $H^{f,g}(u,v) = f(u) + g(v)$.

We want to find the closed-form expression of modified Hamiltonian for the following symplectic Euler updates:

$$p_{k+1} = p_k - 2\eta C q_k, \quad q_{k+1} = q_k + 2\eta B p_{k+1}$$

Throughout the subsection, we let $\star$ denote the tuple $(r_1, s_1, \dots, r_n, s_n)$, i.e.,

$$\star \coloneqq (r_1, s_1, \dots, r_n, s_n).$$

We denote the rank (for its definition, check the review on iterated Poisson bracket in Section 6) of the Poisson bracket corresponding to the tuple $\star$ with $\mathcal{N}(\star)$, denote $\mathcal{M}(\star) \coloneqq \sum_{i=1}^n (r_i + s_i)$, and denote the set of tuples of rank N with $\mathcal{S}(N)$. Finally, we define the following function to express the constant coefficient in the $\widetilde{H}_\eta$ series:

$$\tau(\star) = \tau((r_1, s_1, \cdots, r_n, s_n)) = \frac{(-1)^{n-1}}{n} \frac{1}{(r_1 + \cdots + r_n + s_1 + \cdots + s_n)\prod_{i=1}^n r_i! s_i!}$$

For simplicity of notations, we define

$$I_{p,q}^{F,G} \coloneqq \{G^{r_1}F^{s_1}G^{r_2}F^{s_2}\cdots G^{r_n}F^{s_n}\}(p,q)$$
$$I_{u,v}^{f,g} \coloneqq \{g^{r_1}f^{s_1}g^{r_2}f^{s_2}\cdots g^{r_n}f^{s_n}\}(u,v)$$

Results from Appendix C.1 imply that $I_{u,v}^{f,g}(\star)$ must be a monomial containing a, b, u, v. We denote the constant factor within $I_{u,v}^{f,g}(\star)$ with $J(\star)$, denote the power of a within $I_{u,v}^{f,g}(\star)$ with $L_a(\star)$ and the power of b within $I_{u,v}^{f,g}(\star)$ with $L_b(\star)$. The following proposition is a direct consequence of our work above:

Proposition 22 *$I_{u,v}^{f,g}(\star)$ must be one of the following forms:*

- $J(\star)a^{L_a(\star)}b^{L_b(\star)}u^2$,

- $J(\star)a^{L_a(\star)}b^{L_b(\star)}v^2$, *or*
- $J(\star)a^{L_a(\star)}b^{L_b(\star)}uv$.

We define the following subsets of possible tuples:

$$\mathcal{S}_{pp}(n) \coloneqq \mathcal{S}_{uu}(n) \coloneqq \{\star \in \mathcal{S}(n) \mid I_{u,v}^{f,g}(\star) \text{ contains } uu\}$$
$$\mathcal{S}_{qq}(n) \coloneqq \mathcal{S}_{uu}(n) \coloneqq \{\star \in \mathcal{S}(n) \mid I_{u,v}^{f,g}(\star) \text{ contains } uu\}$$
$$\mathcal{S}_{pq}(n) \coloneqq \mathcal{S}_{uv}(n) \coloneqq \{\star \in \mathcal{S}(n) \mid I_{u,v}^{f,g}(\star) \text{ contains } uv\}$$

Notice that $\star \in \mathcal{S}_{uu}(n) \cup \mathcal{S}_{uu}(n) \cup \mathcal{S}_{uv}(n)$ implies that $I_{u,v}^{f,g}(\star) \neq 0$.
Finally, we define the following functions to connect multivariate and univariate quadratic cases.

$$T_\eta^{\text{pure}}(\lambda) \coloneqq \begin{cases} \frac{\operatorname{arcsin}(\sqrt{\lambda\eta^2})}{\sqrt{\lambda\eta^2(1-\lambda\eta^2)}} & \lambda > 0, \\ \frac{\operatorname{arcsinh}(\sqrt{-\lambda\eta^2})}{\sqrt{-\lambda\eta^2(1-\lambda\eta^2)}} & \lambda < 0. \end{cases}, \quad T_\eta^{\text{cross}}(\lambda) \coloneqq \begin{cases} \frac{-2\sqrt{\lambda}\operatorname{arcsin}(\sqrt{\lambda\eta^2})}{\sqrt{(1-\lambda\eta^2)}} & \lambda > 0, \\ \frac{2\sqrt{-\lambda}\operatorname{arcsinh}(\sqrt{-\lambda\eta^2})}{\sqrt{(1-\lambda\eta^2)}} & \lambda < 0. \end{cases}$$

C.4.2 Connections between the iterated Poisson brackets in two cases

Lemma 23 *If $\star \in \mathcal{S}_{uu}(n)$, then $L_a(\star) = L_b(\star) + 1 = \frac{\mathcal{M}(\star)+1}{2}$. Furthermore, if $\star \in \mathcal{S}_{uu}(n)$, then $L_b(\star) = L_a(\star) + 1 = \frac{\mathcal{M}(\star)+1}{2}$. Finally, if $\star \in \mathcal{S}_{uv}(n)$, then $L_a(\star) = L_b(\star) = \frac{\mathcal{M}(\star)}{2}$.*

Proof This can be verified via an inductive argument based on $\mathcal{M}(\star) = \{1, 2, 3, \cdots\}$. As the base cases, the proposed lemma holds for $M = 1$, where $f = au^2$, $g = bv^2$ and $M = 2$, where $\{f, g\} = -4abuv = -\{g, f\}$. Suppose that the lemma holds for any tuples $\star$ such that $\mathcal{M}(\star) \leq n$, we verify that it also holds for any tuples $\star$ such that $\mathcal{M}(\star) \leq n + 1$. For the sake of simplicity we only verify the case where $\{\cdot, f\}$ is applied to $I_{u,v}^{f,g}(\star)$. We denote the new tuple in this case as $\star'$ and remark that $\mathcal{M}(\star') = \mathcal{M}(\star) + 1$.

- If $I_{u,v}^{f,g}(\star) = J(\star)a^{\frac{\mathcal{M}(\star)+1}{2}}b^{\frac{\mathcal{M}(\star)-1}{2}}u^2$, then $\{I_{u,v}^{f,g}(\star), f\} = \{J(\star)a^{\frac{\mathcal{M}(\star)+1}{2}}b^{\frac{\mathcal{M}(\star)-1}{2}}u^2, au^2\} = 0$.
- If $I_{u,v}^{f,g}(\star) = J(\star)a^{\frac{\mathcal{M}(\star)-1}{2}}b^{\frac{\mathcal{M}(\star)+1}{2}}v^2$, then $\{I_{u,v}^{f,g}(\star), f\} = \{J(\star)a^{\frac{\mathcal{M}(\star)-1}{2}}b^{\frac{\mathcal{M}(\star)+1}{2}}v^2, au^2\}$ $= J(\star')a^{\frac{\mathcal{M}(\star)+1}{2}}b^{\frac{\mathcal{M}(\star)+1}{2}}uv = J(\star')a^{\frac{\mathcal{M}(\star')}{2}}b^{\frac{\mathcal{M}(\star')}{2}}uv$.
- If $I_{u,v}^{f,g}(\star) = J(\star)a^{\frac{\mathcal{M}(\star)}{2}}b^{\frac{\mathcal{M}(\star)}{2}}uv$, then $\{I_{u,v}^{f,g}(\star), f\} = \{J(\star)a^{\frac{\mathcal{M}(\star)}{2}}b^{\frac{\mathcal{M}(\star)}{2}}uv, au^2\}$ $= J(\star')a^{\frac{\mathcal{M}(\star)+2}{2}}b^{\frac{\mathcal{M}(\star)}{2}}u^2 = J(\star')a^{\frac{\mathcal{M}(\star')+1}{2}}b^{\frac{\mathcal{M}(\star')}{2}}u^2$.

Thus, the induction argument is verified in all cases. ∎

Lemma 24 *If $\star \in \mathcal{S}_{uu}(n)$, then $I_{p,q}^{F,G}(\star) = J(\star)p^\top(BC)^{\frac{\mathcal{M}(\star)-1}{2}}Bp$. Furthermore, if $\star \in \mathcal{S}_{uu}(n)$, then $I_{p,q}^{F,G}(\star) = J(\star)q^\top C(BC)^{\frac{\mathcal{M}(\star)-1}{2}}q$. Finally, if $\star \in \mathcal{S}_{uv}(n)$, then $I_{p,q}^{F,G}(\star) = J(\star)p^\top(BC)^{\frac{\mathcal{M}(\star)}{2}}q$.*

Proof This can be verified using an inductive argument based on $\mathcal{M}(\star) = \{1, 2, 3, \cdots\}$. As the base case, the proposed lemma holds for $M = 1$, where $F = p^\top Bp$, $G = q^\top Cq$ and $M = 2$, where $\{F, G\} = -4p^\top BCq = -\{G, F\}$. Suppose the lemma holds for any tuples $\star$ such that $\mathcal{M}(\star) \leq n$, we verify that it also holds for any tuples $\star$ such that $\mathcal{M}(\star) \leq n + 1$. For the sake of simplicity we only verify the case where $\{\cdot, F\}$ is applied to $I_{p,q}^{F,G}(\star)$ and when $\{\cdot, f\}$ is applied to $I_{u,v}^{f,g}(\star)$. To show the inductive argument, we need to compare $\{I_{p,q}^{F,G}(\star), F\}$ and $\{I_{u,v}^{f,g}(\star), f\}$.

If $\star$ ends with $\{\dots, r_n, 0\}$ where $r_n > 0$, the new tuple $\star'$ will be $\{\dots, r_n, 1\}$. On the other hand, if $\star$ ends with $\{\dots, r_n, s_n\}$ where $r_n, s_n > 0$, the new tuple $\star'$ will be $\{\dots, r_n, s_n + 1\}$. In both cases, $I_{p,q}^{F,G}(\star') = I_{p,q}^{F,G}(\star), F$. Note that $\mathcal{M}(\star') = \mathcal{M}(\star) + 1$.

When $\star \in \mathcal{S}_{uu}(n)$, $I_{p,q}^{F,G} = J(\star)p^\top (BC)^{\frac{\mathcal{M}(\star)-1}{2}} Bp$ and $I_{u,v}^{f,g} = J(\star)a^{\frac{\mathcal{M}(\star)+1}{2}} b^{\frac{\mathcal{M}(\star)-1}{2}} u^2$. Thus,

$$\{I_{p,q}^{F,G}(\star), F\} = \{I_{u,v}^{f,g}(\star), f\} = 0.$$

When $\star \in \mathcal{S}_{vv}(n)$, $I_{p,q}^{F,G} = J(\star)q^\top C(BC)^{\frac{\mathcal{M}(\star)-1}{2}} q$ and $I_{u,v}^{f,g} = J(\star)a^{\frac{\mathcal{M}(\star)-1}{2}} b^{\frac{\mathcal{M}(\star)+1}{2}} v^2$. In this case,

$$\{I_{u,v}^{f,g}(\star), f\} = 4J(\star)a^{\frac{\mathcal{M}(\star)+1}{2}} b^{\frac{\mathcal{M}(\star)+1}{2}} uv = 4J(\star)a^{\frac{\mathcal{M}(\star')}{2}} b^{\frac{\mathcal{M}(\star')}{2}} uv.$$

On the other hand,

$$\{I_{p,q}^{F,G}(\star), F\} = 4J(\star)p^\top (BC)^{\frac{\mathcal{M}(\star)+1}{2}} q = 4J(\star)p^\top (BC)^{\frac{\mathcal{M}(\star')}{2}} q.$$

Since the constant term match and the power of a, b matches with that of B and C, the assumption is verified.

When $\star \in \mathcal{S}_{uv}(n)$, $I_{p,q}^{F,G} = J(\star)p^\top (BC)^{\frac{\mathcal{M}(\star)}{2}} q$ and $I_{u,v}^{f,g} = J(\star)a^{\frac{\mathcal{M}(\star)}{2}} b^{\frac{\mathcal{M}(\star)}{2}} uv$. In this case,

$$\{I_{u,v}^{f,g}(\star), f\} = 2J(\star)a^{\frac{\mathcal{M}(\star)+2}{2}} b^{\frac{\mathcal{M}(\star)}{2}} u^2 = 2J(\star)a^{\frac{\mathcal{M}(\star')+1}{2}} b^{\frac{\mathcal{M}(\star')-1}{2}} u^2.$$

On the other hand,

$$\{I_{p,q}^{F,G}(\star), F\} = 2J(\star)p^\top (BC)^{\frac{\mathcal{M}(\star)}{2}} Bp = 2J(\star)p^\top (BC)^{\frac{\mathcal{M}(\star')-1}{2}} Bp.$$

Since the constant term match and the power of a, b matches with that of B and C, the assumption is verified. ■

Corollary 25 *If $\star \in \mathcal{S}_{uu}(n)$, then $I_{p,q}^{F,G}(\star) = J(\star)(Qp)^\top \Lambda^{\frac{\mathcal{M}(\star)-1}{2}} QBp$. If $\star \in \mathcal{S}_{vv}(n)$, then $I_{p,q}^{F,G}(\star) = J(\star)(QC^\top q)^\top \Lambda^{\frac{\mathcal{M}(\star)-1}{2}} Qq$. If $\star \in \mathcal{S}_{uv}(n)$, then $I_{p,q}^{F,G}(\star) = J(\star)(Qp)^\top \Lambda^{\frac{\mathcal{M}(\star)}{2}} Qq$.*

Proof We note that $(BC)^{\frac{\mathcal{M}(\star)-1}{2}} B = (Q^{-1}\Lambda Q)^{\frac{\mathcal{M}(\star)-1}{2}} B = Q^{-1}\Lambda^{\frac{\mathcal{M}(\star)-1}{2}} QB$, $C(BC)^{\frac{\mathcal{M}(\star)-1}{2}} = C(Q^{-1}\Lambda Q)^{\frac{\mathcal{M}(\star)-1}{2}} = CQ^{-1}\Lambda^{\frac{\mathcal{M}(\star)-1}{2}} Q$, and $(BC)^{\frac{\mathcal{M}(\star)}{2}} = (Q^{-1}\Lambda Q)^{\frac{M(\star)}{2}} = Q^{-1}\Lambda^{\frac{M(\star)}{2}} Q$. ■

Hence, if $\star \in \mathcal{S}_{pp}$, then $I_{u,v}^{f,g}(\star)$ is a bilinear product involving p with itself, and analogous statements hold for $\mathcal{S}_{qq}$ and $\mathcal{S}_{pq}$.

C.4.3 Reordering the series: finding the closed-form and absolute convergence radius

Lemma 26 *The Dynkin form (21) of the modified Hamiltonian in multivariate quadratic case $\widetilde{H}_\eta^{F,G}(p,q)$ can be rearranged as follows:*

$$\begin{aligned}\hat{H}_\eta^{F,G}(p,q) = (Qp)^\top &\left(\sum_{N=1}^{\infty}\frac{(-1)^{N-1}}{N}\sum_{\star\in\mathcal{S}_{pp}(N)}\eta^{\mathcal{M}(\star)-1}\tau(\star)J(\star)\Lambda^{\frac{\mathcal{M}(\star)-1}{2}}\right)QBp\\ &+ (QC^\top q)^\top\left(\sum_{N=1}^{\infty}\frac{(-1)^{N-1}}{N}\sum_{\star\in\mathcal{S}_{qq}(N)}\eta^{\mathcal{M}(\star)-1}\tau(\star)J(\star)\Lambda^{\frac{\mathcal{M}(\star)-1}{2}}\right)Qq\\ &+ (Qp)^\top\left(\sum_{N=1}^{\infty}\frac{(-1)^{N-1}}{N}\sum_{\star\in\mathcal{S}_{pq}(N)}\eta^{\mathcal{M}(\star)-1}\tau(\star)J(\star)\Lambda^{\frac{\mathcal{M}(\star)}{2}}\right)Qq \end{aligned}\tag{79}$$

Note that the equivalence between $\widetilde{H}_\eta^{F,G}(p,q)$ and $\hat{H}_\eta^{F,G}(p,q)$ is not guaranteed, because we have not proven the absolute convergence yet.

Proof Let us start from the Dynkin form (21) of the modified Hamiltonian:

$$\widetilde{H}_\eta^{F,G}(p,q) = \sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{n}\sum_{\substack{r_1+s_1>0\\ \vdots\\ r_n+s_n>0}}\frac{\eta^{r_1+\cdots+r_n+s_1+\cdots+s_n-1}\left\{G^{r_1}F^{s_1}G^{r_2}F^{s_2}\cdots G^{r_n}F^{s_n}\right\}(p,q)}{(r_1+\cdots+r_n+s_1+\cdots+s_n)\prod_{i=1}^n r_i!s_i!}.$$

The summation can be reordered as:

$$\hat{H}_\eta^{F,G}(p,q) = \sum_{n=1}^{\infty}\sum_{\star\in\mathcal{S}_{pp}(n)\cup\mathcal{S}_{pq}(n)\cup\mathcal{S}_{pq}(n)}\tau(\star)\eta^{\mathcal{M}(\star)}I_{p,q}^{F,G}(\star).$$

We compute the summation for $\mathcal{S}_{pp}(n), \mathcal{S}_{qq}(n)$ and $\mathcal{S}_{pq}(n)$ separately, and apply the results in Lemma 24 to conclude. ■

Theorem 27 *The reordered $\widetilde{H}_\eta$ series in Lemma 26 has the following closed form:*

$$\hat{H}_\eta^{F,G}(p,q) = (Qp)^\top T_\eta^{pure}(\Lambda)QBp + (QC^\top q)^\top T_\eta^{pure}(\Lambda)Qq + (Qp)^\top T_\eta^{cross}(\Lambda)Qq.$$

Whenever each eigenvalue λ in Λ satisfies that $|\lambda\eta^2|\le 1$, the original BCH series $\widetilde{H}_\eta^{F,G}(p,q)$ converges absolutely and reordering of summation is allowed. Since BC is assumed to be diagonalizable, if the maximal singular value of BC is smaller than $\frac{1}{\eta^2}$, then $\widetilde{H}_\eta^{F,G}(p,q) = \hat{H}_\eta^{F,G}(p,q)$.

Proof Theorem 12 states that in the univariate case, BCH series $\widetilde{H}_\eta^{f,g}(u,v)$ converges to $\frac{1}{\sqrt{ab\eta^2(1-ab\eta^2)}}\arcsin(\sqrt{ab\eta^2})(au^2+bv^2-2abuv\eta)$. Furthermore, the convergence is absolute whenever $|ab\eta^2|<1$.

Assume without loss of generality that $ab > 0$, for the other case is similar. Absolute convergence allows for the following reordering of the summation:

$$
\begin{aligned}
\widetilde{H}_\eta^{f,g}(u,v) &= \left(\sum_{N=1}^{\infty} \frac{(-1)^{N-1}}{N} \sum_{\star \in \mathcal{S}_{uu}(N)} \eta^{\mathcal{M}(\star)-1} \tau(\star) J(\star) (ab)^{\frac{\mathcal{M}(\star)-1}{2}}\right) au^2 \\
&\quad + \left(\sum_{N=1}^{\infty} \frac{(-1)^{N-1}}{N} \sum_{\star \in \mathcal{S}_{vv}(N)} \eta^{\mathcal{M}(\star)-1} \tau(\star) J(\star) (ab)^{\frac{\mathcal{M}(\star)-1}{2}}\right) bv^2 \\
&\quad + \left(\sum_{N=1}^{\infty} \frac{(-1)^{N-1}}{N} \sum_{\star \in \mathcal{S}_{uv}(N)} \eta^{\mathcal{M}(\star)-1} \tau(\star) J(\star) (ab)^{\frac{\mathcal{M}(\star)}{2}}\right) uv \\
&= \frac{1}{\sqrt{ab\eta^2(1-ab\eta^2)}} \arcsin(\sqrt{ab\eta^2})(au^2 + bu^2 - 2abuv\eta). \qquad (80)
\end{aligned}
$$

We now substitute the variable ab with λ. Equation (80) above implies the following:

$$
\sum_{N=1}^{\infty} \frac{(-1)^{N-1}}{N} \sum_{\star \in \mathcal{S}_{uu}(N)} \eta^{\mathcal{M}(\star)-1} \tau(\star) J(\star) \lambda^{\frac{\mathcal{M}(\star)-1}{2}} = \frac{\arcsin(\sqrt{\lambda\eta^2})}{\sqrt{\lambda\eta^2(1-\lambda\eta^2)}} = T_\eta^{\text{pure}}(\lambda) \qquad (81a)
$$

$$
\sum_{N=1}^{\infty} \frac{(-1)^{N-1}}{N} \sum_{\star \in \mathcal{S}_{vv}(N)} \eta^{\mathcal{M}(\star)-1} \tau(\star) J(\star) \lambda^{\frac{\mathcal{M}(\star)-1}{2}} = \frac{\arcsin(\sqrt{\lambda\eta^2})}{\sqrt{\lambda\eta^2(1-\lambda\eta^2)}} = T_\eta^{\text{pure}}(\lambda) \qquad (81b)
$$

$$
\sum_{N=1}^{\infty} \frac{(-1)^{N-1}}{N} \sum_{\star \in \mathcal{S}_{uv}(N)} \frac{\eta^{\mathcal{M}(\star)-1}}{\mathcal{M}(\star)} \tau(\star) J(\star) \lambda^{\frac{\mathcal{M}(\star)}{2}} = \frac{-2\sqrt{\lambda}\arcsin(\sqrt{\lambda\eta^2})}{\sqrt{(1-\lambda\eta^2)}} = T_\eta^{\text{cross}}(\lambda). \qquad (81c)
$$

Furthermore, the absolute convergence of original Dynkin series (21) when $|\lambda\eta^2| < 1$ implies absolute convergence of $T_\eta^{\text{pure}}(\lambda)$ and $T_\eta^{\text{cross}}(\lambda)$ when $|\lambda\eta^2| < 1$.

Since Λ is a diagonal matrix, taking powers of Λ is equivalent with taking the same powers element-wise. Thus, the summation of matrices in (79) is equivalent with summing each element along the diagonal separately. Furthermore, since the terms in (79) are precisely those in (80) and thereby equal to the respective ones in (81), the closed form should equal to the application of T_η^{pure} and T_η^{cross} on the diagonal elements separately. Finally, note that since $T_\eta^{\text{cross}}(\lambda) = |\lambda| T_\eta^{\text{pure}}(\lambda)$ holds for both $\lambda > 0$ and $\lambda < 0$, we could simplify $T_\eta^{\text{cross}}(\Lambda)$ as $|\Lambda| T_\eta^{\text{pure}}(\Lambda)$. Thus, the closed-form is proven.

Finally, following the convergence criteria in Theorem 12, the convergence criterion here should be $\lambda_{\max}(|\Lambda|)\eta^2 = \sigma_{\max}(BC) < 1$. Hence, since $\sigma_{\max}(BC) \le \sigma_{\max}(B)\sigma_{\max}(C)$, taking $\sigma_{\max}(B)\sigma_{\max}(C)\eta^2 < 1$ suffices to ensure convergence. ■

Appendix D. Partial proof of Conjecture 10

We start by taking the Taylor series of the Nth modification H_N:

$$
\begin{aligned}
&H_N(p_{n+1}, q_{n+1}) \\
&= H_N(p_n, q_n) + \sum_{i=1}^{\infty} \frac{1}{i!} \Big((p_{n+1} - p_n) \cdot \nabla_p + (q_{n+1} - q_n) \cdot \nabla_q\Big)^i H_N(p_n, q_n), \qquad (82)
\end{aligned}
$$

where the gradients are to be applied to H_N. We proceed by plugging the iterates (5) of symplectic Euler into (82):

$$\begin{aligned}&H_N(p_{n+1}, q_{n+1}) - H_N(p_n, q_n)\\&= \sum_{i=1}^{\infty} \frac{\eta^i}{i!}\Big(- \nabla_q G\Big|_{q=q_n} \cdot \nabla_p + \nabla_p F\Big|_{p=p_{n+1}} \cdot \nabla_q\Big)^i H_N(p_n, q_n).\end{aligned} \tag{83}$$

The fact that $\nabla_p F$ is evaluated at p_{n+1} requires us to also approximate $F(p_{n+1})$ with Taylor series:

$$\begin{aligned}F(p_{n+1}) &= F\left(p_n - \eta \nabla_q G(q_n)\right)\\&= F\left(p_n\right) - \eta \left[\nabla_p F, \nabla_q G\right]\big|_{p_n,q_n}\\&\quad+ \frac{\eta^2}{2!}\left[\nabla_p^2 F, \nabla_q G^{\otimes 2}\right]\big|_{p_n,q_n} - \frac{\eta^3}{3!}\left[\nabla_p^3 F, \nabla_q G^{\otimes 3}\right]\Big|_{p_n,q_n} + \cdots\\&= F\left(p_n\right) + \eta\left\{F, G\right\}\big|_{p_n,q_n} + \frac{\eta^2}{2!}\left\{\left\{F, G\right\}, G\right\}\big|_{p_n,q_n} + \frac{\eta^3}{3!}\left\{\left\{\left\{F, G\right\}, G\right\}, G\right\}\big|_{p_n,q_n} + \cdots\\&= e^{\eta\{\cdot, G\}} F\Big|_{p_n,q_n}.\end{aligned} \tag{84}$$

We proceed by plugging (84) into (83):

$$\begin{aligned}&H_N(p_{n+1}, q_{n+1}) - H_N(p_n, q_n)\\&= \sum_{i=1}^{\infty} \frac{\eta^i}{i!}\left[-\nabla_q G\left(q_n\right) \cdot \nabla_p + \nabla_p \left(e^{\eta\{\cdot, G\}} F\right)\Big|_{p_n,q_n} \cdot \nabla_q\right]^i H_N\Bigg|_{p_n,q_n}.\end{aligned} \tag{85}$$

For (85), we note that the evaluations at p_n, q_n both inside and outside the brackets are intentional. The reader should interpret this notation in the sense that, when the differential operator inside the brackets is repeatedly applied to H_N, the derivatives with respect to p and q do *not* apply to $\nabla_q G\left(q_n\right)$ or $\nabla_p \left(e^{\eta\{\cdot, G\}} F\right)\big|_{p_n,q_n}$, as one would normally expect from the product rule.

We proceed by rewriting the terms contained in $\widetilde{H}_N(p_{n+1}, q_{n+1}) - \widetilde{H}_N(p_n, q_n)$ as the sum of $N+1$ Taylor series. Using the exponential operator series expansion, we can further rewrite (85) in the following form:

$$\begin{aligned}&H_N(p_{n+1}, q_{n+1}) - H_N(p_n, q_n)\\&= \Big\{\exp\Big[\eta\Big(-\nabla_q G\left(q_n\right) \cdot \nabla_p + \nabla_p \left(\exp\left(\eta\left\{\cdot, G\right\}\right) F\right)\big|_{p_n,q_n} \cdot \nabla_q\Big)\Big] - 1\Big\} H_N\Big|_{p_n,q_n}.\end{aligned} \tag{86}$$

Then, by taking a Taylor series of (86) about $\eta = 0$, the $\mathcal{O}\left(\eta^j\right)$ term in (86) is $C_{j,k} \coloneqq \Omega_j(H_k)$ for all $j \in \mathbb{N}$ and $k \in \mathbb{N}_0$, where $\Omega_j : C^{r+j}\left(\mathcal{Z}, \mathbb{R}\right) \to C^r\left(\mathcal{Z}, \mathbb{R}\right)$ for any $r \in \mathbb{N}_0$ is the differential operator

$$\Omega_j\left(f\right) \coloneqq \frac{1}{j!}\frac{\partial^{(j)}}{\partial \eta^{(j)}}\Big\{\exp\Big[\eta\Big(-\nabla_q G\left(q_n\right) \cdot \nabla_p + \nabla_p \left(\exp\left(\eta\left\{\cdot, G\right\}\right) F\right)\big|_{p_n,q_n} \cdot \nabla_q\Big)\Big] - 1\Big\}\Bigg|_{\eta=0} f \tag{87}$$

for all $f \in C^{r+j}(\mathcal{Z}, \mathbb{R})$ and $j \in \mathbb{N}$.

To get the conservation error of the kth-order truncated modified Hamiltonian, $\widetilde{H}_\eta^{(k)}$, we combine the first k modifications and truncate the Taylor series at $\mathcal{O}\left(\eta^{k+1}\right)$ using the Lagrange form of the remainder from single-variable Taylor's theorem. Hence, we have the following:

Lemma 28 *Suppose that the domains $\mathcal{P}$ and $\mathcal{Q}$ are convex and closed, and assume that F and G are sufficiently smooth. If the conditions*

$$\sum_{j=0}^{m-1} C_{m-j,j}\big|_{p,q} = 0, \qquad \text{for all} \;\; 1 \le m \le N+1, \tag{88}$$

and $|C_{m-j+1,j}| \le L$ hold for all $(p,q) \in \mathcal{Z}$, then we have that $\widetilde{H}_\eta^{(N)}(p_{n+1}, q_{n+1}) - \widetilde{H}_\eta^{(N)}(p_n, q_n) \le \eta^{N+2}(N+1)L$.

Proof Since $\mathcal{P}$ and $\mathcal{Q}$ are convex and closed, (p_n, q_n) remains within $\mathcal{Z}$ for all n, and we can apply the single-variable Taylor's theorem with Lagrange remainder (Ross, 2013, Theorem 31.3) to expand (86) in the stepsize η about $\eta = 0$, for any fixed truncation order $j + k = N$. Doing so gives

$$\begin{aligned}
&(H_0(p_{n+1}, q_{n+1}) - H_0(p_n, q_n)) \\
&\quad = \left[\eta C_{1,0} + \eta^2 C_{2,0} + \cdots + \eta^{N+1} C_{N+1,0}\right]_{z=z_n} + \eta^{N+2}\, C_{N+2,0}\big|_{z=\zeta_0} \\
&\eta^1 \left(H_1(p_{n+1}, q_{n+1}) - H_1(p_n, q_n)\right) \\
&\quad = \left[\eta^2 C_{1,1} + \eta^3 C_{2,1} + \cdots + \eta^{N+1} C_{N,1}\right]_{z=z_n} + \eta^{N+2}\, C_{N+1,1}\big|_{z=\zeta_1} \\
&\eta^2 \left(H_2(p_{n+1}, q_{n+1}) - H_2(p_n, q_n)\right) \\
&\quad = \left[\eta^3 C_{1,2} + \eta^4 C_{2,2} + \cdots + \eta^{N+1} C_{N-1,2}\right]_{z=z_n} + \eta^{N+2}\, C_{N,2}\big|_{z=\zeta_2} \\
&\vdots \quad = \\
&\eta^N \left(H_N(p_{n+1}, q_{n+1}) - H_N(p_n, q_n)\right) \\
&\quad = \eta^{N+1}\, C_{1,N}\big|_{z=z_n} + \eta^{N+2}\, C_{2,N}\big|_{z=\zeta_N}
\end{aligned} \tag{89}$$

for the order-N truncated modified Hamiltonian, where $\zeta_0, \zeta_1, \ldots, \zeta_N$ lie on the convex line segment $L_n \coloneqq \left\{z \in \mathbb{R}^d \times \mathbb{R}^d : z = (1-\lambda)\, z_{n+1} + \lambda z_n, 0 \le \lambda \le 1\right\}$ between z_n and z_{n+1}. Thus,

$$\begin{aligned}
&\left|\widetilde{H}_\eta^{(N)}(p_{n+1}, q_{n+1}) - \widetilde{H}_\eta^{(N)}(p_n, q_n)\right| \\
&= \left|\sum_{m=1}^{N+1} \eta^m \sum_{j=0}^{m-1} C_{m-j,j}\big|_{p_n,q_n} + \eta^{N+2} \sum_{m=0}^{N} C_{N+2-m,m}|_{z=\zeta_m}\right| \\
&\le \eta^{N+2} \sum_{m=0}^{N} |C_{N+2-m,m}|_{z=\zeta_m}| \\
&\le \eta^{N+2}(N+1)L.
\end{aligned} \tag{90}$$

■

It now remains to show that the sum along each diagonal in (89) actually cancels to zero except for that which corresponds to the last coefficient at order η^{N+2}. To partially answer this question, in the next two subsections, we show that the following lemma holds:

Lemma 29 *For $N = 0, 1, 2, 3$, if $F : \mathcal{P} \times \mathbb{R}$ and $G : \mathcal{Q} \times \mathbb{R}$ have derivatives up to order $N+2$ on their respective domains, then we have that $\sum_{j=0}^{i} C_{i+1-j,j} = 0$ for all $i \leq N$ on all of $\mathcal{Z}$. Moreover, for $N = 0, 1, \ldots, 10$, when $d = 1$ and F, G have derivatives up to order $N+2$ on their respective domains, then we have $\sum_{j=0}^{i} C_{i+1-j,j} = 0$ for all $i \leq N$ on $\mathcal{Z}$.*

Once Lemma 29 has been proven, we could use that and Lemma 28 to show that $\widetilde{H}_\eta^{(N)}(p_{n+1}, q_{n+1}) - \widetilde{H}_\eta^{(N)}(p_n, q_n) = \mathcal{O}(\eta^{N+2})$. However, at least up to $j + k = 5$ (i.e., up to $N = 3$), we will show that each $C_{j,k}$ can be expressed as a linear combination of products of higher-order derivatives of F and G such that, for each product in this linear combination, the sum of the orders of derivatives in the product adds up to $2\,(j+k) = 2N+4$, no individual derivative is of higher order than $j + k = N + 2$, and there are $N + 3$ of such terms. Thus, we can show Conjecture 10 for (in theory) arbitrarily high $N \in \mathbb{N}_0$ by assuming that F and G are both L-smooth of orders $1, \ldots, N+2$, since in that case, (90) can be simplified to

$$\left|\widetilde{H}_\eta^{(N)}\left(p_{n+1}, q_{n+1}\right) - \widetilde{H}_\eta^{(N)}\left(p_n, q_n\right)\right| \leq \Phi\left(N\right) L^{N+3} \eta^{N+2} \tag{91}$$

for some function $\Phi : \mathbb{N}_0 \to \mathbb{Q}$ to be determined.

Using the aforementioned steps, we can verify (91) by routine direct computation and thereby prove Lemma 29 for orders $N \in \{0, 1, 2, 3\}$ and for arbitrary d dimensions. For the sake of brevity, we omit the details here; the explicit computations can be found in Appendix F.1 of the arXiv version of this paper, Katona et al. (2024).

However, for $d = 1$ dimensions, (91) can be verified by computer for orders $N \in \{0, 1, \ldots, 10\}$. The reason why this approach only works for $d = 1$ since SymPy (Meurer et al., 2017) does not currently have support for higher-order derivative tensors. To compute the successive terms of the modified Hamiltonian in SymPy, we implement the recursive form of the BCH formula as stated in Casas and Murua (2009) and Section 2.15 in Varadarajan (1984), and proved as Lemma 2.15.3 in Varadarajan (1984). When applied to the setting of symplectic Euler and Alternating Mirror Descent, we recover the following for any $N \in \mathbb{N}$:

$$H_0 = F + G, \tag{92a}$$

$$N H_N = \frac{1}{2}\left\{H_{N-1}, G - F\right\} + \sum_{p=1}^{\lfloor (N-1)/2 \rfloor} \frac{B_{2p}}{(2p)!}\left(\operatorname{ad}_{\widetilde{H}}^{2p}\left(F+G\right)\right)_N, \tag{92b}$$

where

$$\left(\operatorname{ad}_{\widetilde{H}}^{2p}\left(F+G\right)\right)_N = \sum_{\substack{k_1+\cdots+k_{2p}=N-1 \\ k_1 \geq 1, \ldots, k_{2p} \geq 1}} \left\{\left(F+G\right) H_{k_{2p}-1} \cdots H_{k_1-1}\right\} \tag{93}$$

denotes the projection of $\mathrm{ad}^{2p}_{\widetilde{H}}(F+G)$ onto the homogeneous subspace $\mathcal{L}(F,G)_{N+1}$ consisting of degree $N+1$ elements from $\mathcal{L}(F,G)$, as defined in (48).

In SymPy (Meurer et al., 2017), we use the recursive formula above to compute the order-by-order corrections to the modified Hamiltonian H_k. These are then used to compute the values of $C_{j,k} = \Omega_j(H_k)$ for each fixed value of $j+k-1 = N$ to check Conjecture 10 at each order in N. These are also used to compute the remainders for fixed values of $j+k = N+1$.

We use the formula (87) to compute the action of Ω_j on each H_k as needed. However, we do this by expanding both exponentials as power series in η up to sufficiently high order, from which we can compute the derivatives symbolically with respect to η and then evaluate the result at $\eta = 0$.

D.1 Verification of Lemma 29 for orders $N \in \{0,1,2,3\}$ in any $d > 1$

Calculation of Ω_j and H_n. We start by recalling the first few terms in the modified Hamiltonian series:

$$\begin{aligned}\widetilde{H}_\eta = F + G - \frac{\eta}{2}\langle \nabla_p F, \nabla_q G\rangle + \frac{\eta^2}{12}\nabla_q G^\top \nabla_p^2 F \nabla_q G + \frac{\eta^2}{12}\nabla_p F^\top \nabla_q^2 G \nabla_p F \\ - \frac{\eta^3}{24}\nabla_p F^\top \nabla_q^2 G \nabla_p^2 F \nabla_p F + \mathcal{O}(\eta^4).\end{aligned}$$

Thus, we know that

$$\begin{aligned}&H_0 = F+G, \quad H_1 = -\frac{1}{2}\nabla_p F^\top \nabla_q G, \quad H_2 = \frac{1}{12}(\nabla_q G^\top \nabla_p^2 F \nabla_q G^\top + \nabla_p F \nabla_q^2 G \nabla_p F), \\ &H_3 = -\frac{1}{24}\nabla_q G^\top \nabla_p^2 F \nabla_q^2 G \nabla_p F. \end{aligned} \tag{94}$$

On the other hand, recall the expression of Ω_j from equation (87):

$$\Omega_j(f) \coloneqq \frac{1}{j!}\frac{\partial^{(j)}}{\partial \eta^{(j)}}\left\{\exp\left[\eta\left(-\nabla_q G(q_n)\cdot\nabla_p + \nabla_p\left(\exp\left(\eta\{\cdot,G\}\right)F\right)\big|_{p_n,q_n}\cdot\nabla_q\right)\right] - 1\right\}\Bigg|_{\eta=0} f,$$

we could calculate Ω_j for $j \in \{0, 1, 2, 3, 4\}$:

$$
\begin{aligned}
\Omega_1(\cdot) &= -\nabla_p(\cdot)\nabla_q G + \nabla_q(\cdot)\nabla_p F && (95)\\
\Omega_2(\cdot) &= \nabla_q(\cdot)\nabla_p\{F, G\} + \frac{1}{2}\nabla_{pp}(\cdot)\left[\nabla_q G^{\otimes 2}\right] - \nabla_{pq}(\cdot)[\nabla_q G, \nabla_p F] + \frac{1}{2}\nabla_{qq}(\cdot)\left[\nabla_p F^{\otimes 2}\right] && (96)\\
\Omega_3(\cdot) &= \frac{1}{2}\nabla_q(\cdot)\nabla_{ppp}F[\nabla_q G^{\otimes 2}] + \nabla_{pq}(\cdot)[\nabla_q G, \nabla_{pp}F\nabla_q G] - \nabla_{qq}(\cdot)[\nabla_p F, \nabla_{pp}F\nabla_q G]\\
&\quad - \frac{1}{6}\nabla_{ppp}(\cdot)[\nabla_q(G)^{\otimes 3}] + \frac{1}{2}\nabla_{ppq}(\cdot)[\nabla_q G^{\otimes 2}, \nabla_p F]\\
&\quad - \frac{1}{2}\nabla_{pqq}(\cdot)[\nabla_q G, \nabla_p F^{\otimes 2}] + \frac{1}{6}\nabla_{qqq}(\cdot)[\nabla_p(F)^{\otimes 3}] && (97)\\
\Omega_4(\cdot) &= \frac{1}{24}\nabla_{pppp}(\cdot)[\nabla_q G^{\otimes 4}] - \frac{1}{6}\nabla_{pppq}(\cdot)[\nabla_q G^{\otimes 3}, \nabla_p F] - \frac{1}{2}\nabla_{ppq}(\cdot)[\nabla_q G^{\otimes 2}, \nabla_{pp}F\nabla_q G]\\
&\quad + \frac{1}{4}\nabla_{ppqq}(\cdot)[\nabla_q G^{\otimes 2}, \nabla_p F^{\otimes 2}] - \frac{1}{2}\nabla_{pq}(\cdot)[\nabla_q G, \nabla_{ppp}F[\nabla_q G^{\otimes 2}]]\\
&\quad + \nabla_{pqq}(\cdot)[\nabla_q G, \nabla_p F, \nabla_{pp}F\nabla_q G] - \frac{1}{6}\nabla_{pqqq}(\cdot)[\nabla_q G, \nabla_p F^{\otimes 3}]\\
&\quad - \frac{1}{6}\nabla_q(\cdot)\nabla_{pppp}F[\nabla_q G^{\otimes 3}] + \frac{1}{2}\nabla_{qq}(\cdot)[\nabla_p F, \nabla_{ppp}F[\nabla_q G^{\otimes 2}]]\\
&\quad - \frac{1}{2}\nabla_{qqq}(\cdot)[\nabla_p F^{\otimes 2}, \nabla_{pp}F\nabla_q G] + \frac{1}{2}\nabla_{qq}(\cdot)[(\nabla_{pp}F\nabla_q G)^{\otimes 2}] + \frac{1}{24}\nabla_{qqqq}(\cdot)[\nabla_p F^{\otimes 4}]\\
& && (98)
\end{aligned}
$$

Verification of Lemma 29 for $N = 0, d > 1$.

$$
\begin{aligned}
C_{1,0} &= \Omega_1(H_0)\\
&= -\nabla_q G\nabla_p(H_0) + \nabla_p F\nabla_q(H_0)\\
&= -\nabla_q G\nabla_p F + \nabla_p F\nabla_q G\\
&= 0. && (99)
\end{aligned}
$$

In the next paragraph, using the calculations for Lemma 28, particular inequality (90), we verify the following inequality in equation (104):

$$
\left|\widetilde{H}_\eta^{(0)}(p_{n+1}, q_{n+1}) - \widetilde{H}_\eta^{(0)}(p_n, q_n)\right| \le L^3\eta^2, \tag{100}
$$

where we show the last inequality.

Verification of Lemma 29 for $N = 1, d > 1$. We first calculate $\Omega_2(H_0)$ and $\Omega_1(H_1)$:

$$
\begin{aligned}
\Omega_2(H_0) &= \nabla_q(H_0)\nabla_p\{F, G\} + \frac{1}{2}\nabla_{pp}(H_0)\left[\nabla_q G^{\otimes 2}\right] - \nabla_{pq}(H_0)[\nabla_q G, \nabla_p F] + \frac{1}{2}\nabla_{qq}(H_0)\left[\nabla_p F^{\otimes 2}\right]\\
&= -\frac{1}{2}\nabla_{pp}F\left[\nabla_q G^{\otimes 2}\right] + \frac{1}{2}\nabla_{qq}G\left[\nabla_p F^{\otimes 2}\right], && (101)\\
\Omega_1(H_1) &= -\nabla_p(H_1)\nabla_q G + \nabla_q(H_1)\nabla_p F = \frac{1}{2}\nabla_{pp}F[\nabla_q G^{\otimes 2}] - \frac{1}{2}\nabla_{qq}G[\nabla_p F^{\otimes 2}]. && (102)
\end{aligned}
$$

We now sum (101) and (102):

$$\begin{aligned}\Omega_2(H_0)+\Omega_1(H_1) &= -\frac{1}{2}\nabla_{pp}F\left[\nabla_q G^{\otimes 2}\right]+\frac{1}{2}\nabla_{qq}G\left[\nabla_p F^{\otimes 2}\right]\\ &\quad+\frac{1}{2}\nabla_{pp}F[\nabla_q G^{\otimes 2}]-\frac{1}{2}\nabla_{qq}G[\nabla_p F^{\otimes 2}]\\ &= 0. \end{aligned} \tag{103}$$

Thus, the case where $N=1$ is verified. In the next paragraph, using the calculations for Lemma 28, particular inequality (90), we verify the following inequality in equation (109):

$$\left|\widetilde{H}_\eta^{(1)}\left(p_{n+1},q_{n+1}\right)-\widetilde{H}_\eta^{(1)}\left(p_n,q_n\right)\right|\le 3L^4\eta^3.$$

Furthermore, to get an upper bound on the $N=0$ case, assuming that F and G are L-smooth of orders 1 and 2, (90) gives us

$$\begin{aligned}\left|\widetilde{H}_\eta^{(0)}\left(p_{n+1},q_{n+1}\right)-\widetilde{H}_\eta^{(0)}\left(p_n,q_n\right)\right| &\le \eta^2\max_{z\in L_n}\left|\Omega_2\left(H_0\right)(z)\right|\\ &\le \frac{\eta^2}{2}\left(\max\left|\nabla_{pp}F\left[\nabla_q G^{\otimes 2}\right]\right|+\max\left|\nabla_{qq}G\left[\nabla_p F^{\otimes 2}\right]\right|\right)\\ &\le L^3\eta^2 \end{aligned}\tag{104}$$

The second inequality above follows from triangle inequality, and the last from Cauchy-Schwarz inequality.

Verification of Lemma 29 for $N=2, d>1$. We first calculate $\Omega_3(H_0)$, $\Omega_2(H_1)$ and $\Omega_1(H_2)$:

$$\begin{aligned}\Omega_3(H_0) &= \frac{1}{2}\nabla_q(H_0)\nabla_{ppp}F[\nabla_q G^{\otimes 2}]+\nabla_{pq}(H_0)[\nabla_q G,\nabla_{pp}F\nabla_q G]\\ &\quad-\nabla_{qq}(H_0)[\nabla_p F,\nabla_{pp}F\nabla_q G]-\frac{1}{6}\nabla_{ppp}(H_0)[\nabla_q(G)^{\otimes 3}]\\ &\quad+\frac{1}{2}\nabla_{ppq}(H_0)[\nabla_q G^{\otimes 2},\nabla_p F]-\frac{1}{2}\nabla_{pqq}(H_0)[\nabla_p G,\nabla_p F^{\otimes 2}]+\frac{1}{6}\nabla_{qqq}(H_0)[\nabla_p(F)^{\otimes 3}]\\ &=\frac{1}{2}\nabla_q G\nabla_{ppp}F[\nabla_q G^{\otimes 2}]-0-\nabla_{qq}G[\nabla_p F,\nabla_{pp}F\nabla_q G]-\frac{1}{6}\nabla_{ppp}F[\nabla_q(G)^{\otimes 3}]\\ &\quad+0+0+\frac{1}{6}\nabla_{qqq}G[\nabla_p(F)^{\otimes 3}]\\ &=\frac{1}{3}\nabla_{ppp}F[\nabla_q G^{\otimes 3}]-\nabla_p F\nabla_{qq}G\nabla_{pp}F\nabla_q G+\frac{1}{6}\nabla_{qqq}G[\nabla_p(F)^{\otimes 3}], \end{aligned}\tag{105}$$

$$\begin{aligned}\Omega_2(H_1) &= \nabla_q(H_1)\nabla_p\{F,G\}+\frac{1}{2}\nabla_{pp}(H_1)\left[\nabla_q G^{\otimes 2}\right]\\ &\quad-\nabla_{pq}(H_1)[\nabla_q G,\nabla_p F]+\frac{1}{2}\nabla_{qq}(H_1)\left[\nabla_p F^{\otimes 2}\right]\\ &=\frac{1}{2}\nabla_p F\nabla_{qq}G\nabla_{pp}F\nabla_q G-\frac{1}{4}\nabla_{ppp}F[\nabla_q G^{\otimes 3}]\\ &\quad+\frac{1}{2}\nabla_q G\nabla_{pp}F\nabla_{qq}G\nabla_p F-\frac{1}{4}\nabla_{qqq}G[\nabla_p F^{\otimes 3}]\\ &=-\frac{1}{4}\nabla_{ppp}F[\nabla_q G^{\otimes 3}]+\nabla_q G\nabla_{pp}F\nabla_{qq}G\nabla_p F-\frac{1}{4}\nabla_{qqq}G[\nabla_p F^{\otimes 3}], \end{aligned}\tag{106}$$

$$
\begin{aligned}
\Omega_1(H_2) &= -\nabla_p(H_2)\nabla_q G + \nabla_q(H_2)\nabla_p F \\
&= -\frac{1}{12}\nabla_{ppp}F[\nabla_q G^{\otimes 3}] - \frac{1}{6}\nabla_p F\nabla_{qq}G\nabla_{pp}F\nabla_q G + \frac{1}{6}\nabla_p F\nabla_{qq}G\nabla_{pp}F\nabla_q G + \frac{1}{12}\nabla_{qqq}G[\nabla_p F^{\otimes 3}] \\
&= -\frac{1}{12}\nabla_{ppp}F[\nabla_q G^{\otimes 3}] + \frac{1}{12}\nabla_{qqq}G[\nabla_p F^{\otimes 3}]. \qquad (107)
\end{aligned}
$$

We now sum (105), (106) and (107):

$$
\begin{aligned}
\Omega_3(H_0) + \Omega_2(H_1) + \Omega_1(H_2) &= \frac{1}{3}\nabla_{ppp}F[\nabla_q G^{\otimes 3}] - \nabla_p F\nabla_{qq}G\nabla_{pp}F\nabla_q G + \frac{1}{6}\nabla_{qqq}G[\nabla_p(F)^{\otimes 3}] \\
&\quad - \frac{1}{4}\nabla_{ppp}F[\nabla_q G^{\otimes 3}] + \nabla_q G\nabla_{pp}F\nabla_{qq}G\nabla_p F - \frac{1}{4}\nabla_{qqq}G[\nabla_p F^{\otimes 3}] \\
&\quad - \frac{1}{12}\nabla_{ppp}F[\nabla_q G^{\otimes 3}] + \frac{1}{12}\nabla_{qqq}G[\nabla_p F^{\otimes 3}] \\
&= 0 \qquad (108)
\end{aligned}
$$

Thus, the case where $N = 2$ is verified. In the next paragraph, using the calculations for Lemma 28, particular inequality (90), we verify the following inequality in equation (115):

$$
\left|\widetilde{H}_\eta^{(2)}(p_{n+1}, q_{n+1}) - \widetilde{H}_\eta^{(2)}(p_n, q_n)\right| \le \frac{49}{12}L^5\eta^4
$$

Moreover, to get an upper bound on the $N = 1$ case, assuming that F and G are L-smooth of orders 1, 2, and 3, (90) gives us

$$
\begin{aligned}
\left|\widetilde{H}_\eta^{(1)}(p_{n+1}, q_{n+1}) - \widetilde{H}_\eta^{(1)}(p_n, q_n)\right| &\le \eta^3\left(\max_{z\in L_n}|\Omega_3(H_0)(z)| + \max_{z\in L_n}|\Omega_2(H_1)(z)|\right) \\
&\le \eta^3 L^4\left(\frac{1}{3} + 1 + \frac{1}{6} + \frac{1}{4} + 1 + \frac{1}{4}\right) \\
&\le 3L^4\eta^3. \qquad (109)
\end{aligned}
$$

Verification of Lemma 29 for $N=3, d>1$. We first calculate $\Omega_4(H_0)$, $\Omega_3(H_1)$, $\Omega_2(H_2)$ and $\Omega_1(H_3)$:

$$\begin{aligned}
\Omega_4(H_0) &= \frac{1}{24}\nabla_{pppp}(H_0)[\nabla_q G^{\otimes 4}] - \frac{1}{6}\nabla_{pppq}(H_0)[\nabla_q G^{\otimes 3}, \nabla_p F] - \frac{1}{2}\nabla_{ppq}(H_0)[\nabla_q G^{\otimes 2}, \nabla_{pp} F \nabla_q G] \\
&\quad + \frac{1}{4}\nabla_{ppqq}(H_0)[\nabla_q G^{\otimes 2}, \nabla_p F^{\otimes 2}] - \frac{1}{2}\nabla_{pq}(H_0)[\nabla_q G, \nabla_{ppp} F[\nabla_q G^{\otimes 2}]] \\
&\quad + \nabla_{pqq}(H_0)[\nabla_q G, \nabla_p F, \nabla_{pp} F \nabla_q G] - \frac{1}{6}\nabla_{pqqq}(H_0)[\nabla_q G, \nabla_p F^{\otimes 3}] \\
&\quad - \frac{1}{6}\nabla_q(H_0)\nabla_{pppp} F[\nabla_q G^{\otimes 3}] + \frac{1}{2}\nabla_{qq}(H_0)[\nabla_p F, \nabla_{ppp} F[\nabla_q G^{\otimes 2}]] \\
&\quad - \frac{1}{2}\nabla_{qqq}(H_0)[\nabla_p F^{\otimes 2}, \nabla_{pp} F \nabla_q G] + \frac{1}{2}\nabla_{qq}(H_0)[(\nabla_{pp} F \nabla_q G)^{\otimes 2}] + \frac{1}{24}\nabla_{qqqq}(H_0)[\nabla_p F^{\otimes 4}] \\
&= \frac{1}{24}\nabla_{pppp} F[\nabla_q G^{\otimes 4}] + 0 - 0 + 0 - 0 + 0 - 0 - \frac{1}{6}\nabla_q G \nabla_{pppp} F[\nabla_q G^{\otimes 3}] \\
&\quad + \frac{1}{2}\nabla_{qq} G[\nabla_p F, \nabla_{ppp} F[\nabla_q G^{\otimes 2}]] - \frac{1}{2}\nabla_{qqq} G[\nabla_p F^{\otimes 2}, \nabla_{pp} F \nabla_q G] \\
&\quad + \frac{1}{2}\nabla_{qq} G[(\nabla_{pp} F \nabla_q G)^{\otimes 2}] + \frac{1}{24}\nabla_{qqqq} G[\nabla_p F^{\otimes 4}] \\
&= -\frac{1}{8}\nabla_{pppp} F[\nabla_q G^{\otimes 4}] - \frac{1}{2}\nabla_{qqq} G[\nabla_p F^{\otimes 2}, \nabla_{pp} F \nabla_q G] + \frac{1}{24}\nabla_{qqqq} G[\nabla_p F^{\otimes 4}] \\
&\quad + \frac{1}{2}\nabla_{qq} G[(\nabla_{pp} F \nabla_q G)^{\otimes 2}] + \frac{1}{2}\nabla_{qq} G[\nabla_p F, \nabla_{ppp} F[\nabla_q G^{\otimes 2}]],
\end{aligned} \tag{110}$$

$$\begin{aligned}
\Omega_3(H_1) &= \frac{1}{2}\nabla_q(H_1)\nabla_{ppp} F[\nabla_q G^{\otimes 2}] + \nabla_{pq}(H_1)[\nabla_q G, \nabla_{pp} F \nabla_q G] - \nabla_{qq}(H_1)[\nabla_p F, \nabla_{pp} F \nabla_q G] \\
&\quad - \frac{1}{6}\nabla_{ppp}(H_1)[\nabla_q G^{\otimes 3}] + \frac{1}{2}\nabla_{ppq}(H_1)[\nabla_q G^{\otimes 2}, \nabla_p F] \\
&\quad - \frac{1}{2}\nabla_{pqq}(H_1)[\nabla_q G, \nabla_p F^{\otimes 2}] + \frac{1}{6}\nabla_{qqq}(H_1)[\nabla_p F^{\otimes 3}] \\
&= -\frac{1}{4}\nabla_p F \nabla_{qq} G \nabla_{ppp} F[\nabla_q G^{\otimes 2}] - \frac{1}{2}\nabla_q G \nabla_{pp} F \nabla_{qq} G \nabla_{pp} F \nabla_q G + \frac{1}{2}[\nabla_p F^{\otimes 2}]\nabla_{qqq} G \nabla_{pp} F \nabla_q G \\
&\quad + \frac{1}{12}\nabla_{pppp} F[\nabla_q G^{\otimes 4}] - \frac{1}{4}[\nabla_q G^{\otimes 2}]\nabla_{ppp} F \nabla_{qq} G \nabla_p F \\
&\quad + \frac{1}{4}\nabla_q G \nabla_{pp} F \nabla_{qqq} G[\nabla_p F^{\otimes 2}] - \frac{1}{12}\nabla_{qqqq} G[\nabla_p F^{\otimes 4}] \\
&= \frac{1}{12}\nabla_{pppp} F[\nabla_q G^{\otimes 4}] - \frac{1}{2}[\nabla_q G^{\otimes 2}]\nabla_{ppp} F \nabla_{qq} G \nabla_p F - \frac{1}{2}\nabla_q G \nabla_{pp} F \nabla_{qq} G \nabla_{pp} F \nabla_q G \\
&\quad + \frac{3}{4}[\nabla_p F^{\otimes 2}]\nabla_{qqq} G \nabla_{pp} F \nabla_q G - \frac{1}{12}\nabla_{qqqq} G[\nabla_p F^{\otimes 4}],
\end{aligned} \tag{111}$$

$$\begin{aligned}
\Omega_2(H_2) &= -\nabla_q(H_2)\nabla_{pp}F\nabla_qG \\
&\quad + \frac{1}{2}\nabla_{pp}(H_2)\left[\nabla_qG^{\otimes 2}\right] - \nabla_{pq}(H_2)[\nabla_qG, \nabla_pF] + \frac{1}{2}\nabla_{qq}(H_2)\left[\nabla_pF^{\otimes 2}\right] \\
&= -\frac{1}{6}\nabla_qG\nabla_{pp}F\nabla_{qq}G\nabla_{pp}F\nabla_qG - \frac{1}{12}[\nabla_pF^{\otimes 2}]\nabla_{qqq}G\nabla_{pp}F\nabla_qG \\
&\quad + \frac{1}{24}\nabla_{pppp}F[\nabla_qG^{\otimes 4}] + \frac{1}{12}\nabla_pF\nabla_{qq}G\nabla_{ppp}F[\nabla_qG^{\otimes 2}] + \frac{1}{12}\nabla_qG\nabla_{pp}F\nabla_{qq}G\nabla_{pp}F\nabla_qG \\
&\quad - \frac{1}{6}[\nabla_qG^{\otimes 2}]\nabla_{ppp}F\nabla_{qq}G\nabla_pF - \frac{1}{6}[\nabla_pF^{\otimes 2}]\nabla_{qqq}G\nabla_{pp}F\nabla_qG \\
&\quad + \frac{1}{12}\nabla_qG\nabla_{pp}F\nabla_{qqq}G[\nabla_pF^{\otimes 2}] + \frac{1}{12}\nabla_pF\nabla_{qq}G\nabla_{pp}F\nabla_{qq}G\nabla_pF + \frac{1}{24}\nabla_{qqqq}G[\nabla_pF^{\otimes 4}] \\
&= \frac{1}{24}\nabla_{pppp}F[\nabla_qG^{\otimes 4}] - \frac{1}{12}\nabla_pF\nabla_{qq}G\nabla_{ppp}F[\nabla_qG^{\otimes 2}] + \frac{1}{12}\nabla_pF\nabla_{qq}G\nabla_{pp}F\nabla_{qq}G\nabla_pF \\
&\quad - \frac{1}{12}\nabla_qG\nabla_{pp}F\nabla_{qq}G\nabla_{pp}F\nabla_qG - \frac{1}{6}[\nabla_pF^{\otimes 2}]\nabla_{qqq}G\nabla_{pp}F\nabla_qG + \frac{1}{24}\nabla_{qqqq}G[\nabla_pF^{\otimes 4}],
\end{aligned} \tag{112}$$

$$\begin{aligned}
\Omega_1(H_3) &= -\nabla_qG\nabla_p(H_3) + \nabla_pF\nabla_q(H_3) \\
&= \frac{1}{12}\nabla_qG\nabla_{pp}F\nabla_{qq}G\nabla_{pp}F\nabla_qG + \frac{1}{12}[\nabla_qG^{\otimes 2}]\nabla_{ppp}F\nabla_{qq}G\nabla_pF \\
&\quad - \frac{1}{12}\nabla_pF\nabla_{qq}G\nabla_{pp}F\nabla_{qq}G\nabla_pF - \frac{1}{12}\nabla_qG\nabla_{pp}F\nabla_{qqq}G[\nabla_pF^{\otimes 2}] \\
&= \frac{1}{12}[\nabla_qG^{\otimes 2}]\nabla_{ppp}F\nabla_{qq}G\nabla_pF + \frac{1}{12}\nabla_qG\nabla_{pp}F\nabla_{qq}G\nabla_{pp}F\nabla_q \\
&\quad - \frac{1}{12}\nabla_pF\nabla_{qq}G\nabla_{pp}F\nabla_{qq}G\nabla_pF - \frac{1}{12}\nabla_qG\nabla_{pp}F\nabla_{qqq}G[\nabla_pF^{\otimes 2}].
\end{aligned} \tag{113}$$

We now sum (110), (111), (112) and (113):

$$\begin{aligned}
&\Omega_4(H_0) + \Omega_3(H_1) + \Omega_2(H_2) + \Omega_1(H_3) = \\
&= \left(-\frac{1}{8} + \frac{1}{12} + \frac{1}{24}\right)\nabla_{pppp}F[\nabla_qG^{\otimes 4}] + \left(\frac{1}{2} - \frac{1}{2} + \frac{1}{12} - \frac{1}{12}\right)\nabla_pF\nabla_{qq}G\nabla_{ppp}F[\nabla_qG^{\otimes 2}] \\
&\quad + \left(\frac{1}{2} - \frac{1}{2} - \frac{1}{12} + \frac{1}{12}\right)\nabla_qG\nabla_{pp}F\nabla_{qq}G\nabla_{pp}F\nabla_qG + \left(\frac{1}{12} - \frac{1}{12}\right)\nabla_pF\nabla_{qq}G\nabla_{pp}F\nabla_{qq}G\nabla_pF \\
&\quad + \left(-\frac{1}{2} + \frac{3}{4} - \frac{1}{6} - \frac{1}{12}\right)[\nabla_pF^{\otimes 2}]\nabla_{qqq}G\nabla_{pp}F\nabla_qG + \left(\frac{1}{24} - \frac{1}{12} + \frac{1}{24}\right)\nabla_{qqqq}G[\nabla_pF^{\otimes 4}] \\
&= 0.
\end{aligned} \tag{114}$$

Thus, the case where $N = 3$ is verified. To get an upper bound on the $N = 2$ case, assuming that F and G are L-smooth of orders $1, \ldots, 4$, (90) implies

$$\begin{aligned}
\left|\widetilde{H}_\eta^{(2)}(p_{n+1}, q_{n+1}) - \widetilde{H}_\eta^{(2)}(p_n, q_n)\right| &\le \eta^4\left(\max_{z\in L_n}|\Omega_4(H_0)(z)| + \max_{z\in L_n}|\Omega_3(H_1)(z)| + \max_{z\in L_n}|\Omega_2(H_2)(z)|\right) \\
&\le \frac{49}{12}L^5\eta^4.
\end{aligned} \tag{115}$$

We refer the readers to Appendix F of Katona et al. (2024) for a thorough derivation of the upper bound on the $N = 3$ case using calculations for the $N = 4$ case, where we drew the following conclusion:

$$\begin{aligned}
&\left|\widetilde{H}_\eta^{(3)}\left(p_{n+1}, q_{n+1}\right) - \widetilde{H}_\eta^{(3)}\left(p_n, q_n\right)\right| \\
&\leq \eta^5 \left(\max_{z \in L_n} \left|\Omega_5\left(H_0\right)(z)\right| + \max_{z \in L_n} \left|\Omega_4\left(H_1\right)(z)\right| + \max_{z \in L_n} \left|\Omega_3\left(H_2\right)(z)\right| + \max_{z \in L_n} \left|\Omega_2\left(H_3\right)(z)\right|\right) \\
&\leq \eta^5 L^6 \left(\frac{13}{8} + \frac{13}{6} + \frac{61}{72} + \frac{5}{6}\right) = \frac{197}{36} \eta^5 L^6.
\end{aligned}$$

D.2 Verification of Lemma 29 for orders $N \leq 10$ and $d = 1$

When $d = 1$, the iterated Poisson brackets become scalar multiplications, and in this case, we could easily verify the conjecture with SymPy, a Python package for symbolic calculation. We have uploaded our code onto a repository, Katona and Wang (2025), and we briefly explain a few key lines of code below:

```
def P_bracket(eta, F, G):
    return eta*(diff(F,q)*diff(G,p) - diff(F,p)*diff(G,q))
```

Here, we define the symbolic expression of the Poisson bracket $\{F, G\}$.

```
def H_m(m, eta, F, G):
    if m == 1: return F + G
    func = P_bracket(eta, H_m(m-1,eta,F,G), G - F) / 2
    for r in range(1, (m-1)//2 + 1):
        ad = 0
        for part in ordered_partitions(m-1, 2*r):
            for perm in multiset_permutations(part):
                B = F + G
                for i in perm:
                    B = P_bracket(eta, B, H_m(i,eta,F,G))
                ad += B
        func += bernoulli(2*r) * ad / factorial(2*r)
    return simplify(nsimplify(func / m))
```

Here, we calculate the m-th modification (see Definition 9) of the modified Hamiltonian series by summing the Poisson brackets.

```
def op_exp(eta, n, F, G):
    inner_operator = F
    for i in (np.arange(n) + 1):
        bracket = F
        for k in np.arange(i):
            bracket = P_bracket(eta, bracket, G)
```

```
        inner_operator += bracket / factorial(i)
    return inner_operator

def nested_op(eta, n, F, G, func):
    series = 0
    for i in np.arange(n + 1):
        for j in np.arange(n - i + 1):
            series += (-eta * diff(G, q)) ** i \
                * (eta * diff(op_exp(eta, n, F, G), p)) ** j \
                * diff(diff(func, p, i), q, j) / (factorial(i) * factorial(j))
    return series
```

Here, we define the exponential series of operators and the symbolic expression of the operator $\Omega_j(\cdot)$.

```
def C_kn(eta, k, n, F, G):
    return expand(diff(nested_op(eta, n, F, G,\
        H_k(k, eta, f(p), g(q))), eta, n).subs(eta, 0) / factorial(n))

N = 30 # check Nth order in eta
for i in (np.arange(N) + 1):
    print('Diagonal ', i, ':')
    sum = 0
    for k in np.arange(i):
        res = C_kn(eta, k, i - k, f(p), g(q))
        sum += res
    simplify(sum)
    print(sum)
```

Finally, we define the symbolic expression of the terms $\Omega_k(H_n)$, and thereby verify that the diagonal sum is indeed 0.

D.3 Concluding the proof

In the prior subsections of Appendix D, we confirmed (91) for orders $N \in \{0, 1, 2, 3\}$ by direct computation for arbitrary dimension d and the cases $N \in \{0, 1, \ldots, 10\}$ by computer for dimension $d = 1$. However, (91) only bounds the difference in $\widetilde{H}_\eta^{(N)}$ after one iteration, while Conjecture 10 gives the same for multiple iterations. To derive a bound after k iterations, we apply (91) recursively and then the triangle inequality as follows:

$$\begin{aligned}\left|\widetilde{H}_\eta^{(N)}(p_k, q_k) - \widetilde{H}_\eta^{(N)}(p_0, q_0)\right| &\leq \sum_{j=0}^{k-1}\left|\widetilde{H}_\eta^{(N)}(p_{j+1}, q_{j+1}) - \widetilde{H}_\eta^{(N)}(p_j, q_j)\right| \\ &\leq k\Phi(N) L^{N+3}\eta^{N+2},\end{aligned}$$

where in the first inequality above, we use the triangle inequality. This concludes the proof. ■

Appendix E. Characterization of Φ

The coefficient function $\Phi : \mathbb{N}_0 \to \mathbb{Q}$ from Conjecture 10 has the following known values for $N = 0, 1, \ldots, 10$:

N	0	1	2	3	4	5	6	7	8	9	10
$\Phi(N)$ (fraction)	1	3	$\frac{49}{12}$	$\frac{197}{36}$	$\frac{139}{20}$	$\frac{1049}{108}$	$\frac{850271}{60480}$	$\frac{10046117}{453600}$	$\frac{16791911}{453600}$	$\frac{73961467}{1088640}$	$\frac{31567569067}{239500800}$
$\Phi(N)$ (rounded)	1.000	3.000	4.083	5.472	6.950	9.713	14.059	22.148	37.019	67.939	131.806

It is not immediately obvious as to why Φ should be bounded. However, up to $N = 10$, we note that $\Phi(N) = 2\frac{\gamma(N)}{\sigma(N+2)}$, where $\gamma : \mathbb{N} \to \mathbb{N}$ returns an unknown sequence of positive integers

N	0	1	2	3	4	5	6	7	8	9	10
$\gamma(N)$	1	18	49	1970	5004	293720	850271	40184468	134335288	16271522740	63135138134

and $\sigma : \mathbb{N} \to \mathbb{N}$ returns the Hirzebruch numbers, which can be defined as follows (OEIS Foundation Inc., 2024):

$$\sigma(n) := \prod_{\substack{2 \le p \le n+1 \\ p \text{ prime}}} p^{\lfloor \frac{n}{p-1} \rfloor} \quad \forall n \in \mathbb{N}.$$

As proven in Corollary 9 from Bedhouche and Farhi (2023), $\sigma(N) = \mathcal{O}\left(N^N\right)$ for N large. Furthermore, $\gamma(N) \le \frac{2}{3}(N+2)^{N+2}$ is true up to $N = 10$ and *only* saturates at $N = 1$. Thus, assuming that $\mathcal{O}(\gamma(N)) = \mathcal{O}\left(N^N\right)$ remains true for $N > 10$, it remains plausible (although hitherto unproven) that $\mathcal{O}(\Phi(N)) = \mathcal{O}(1)$ for large N. Hence, we suspect that Φ is monotonic increasing and bounded as claimed in Conjecture 10.

References

Jacob Abernethy, Kevin A. Lai, and Andre Wibisono. Fast convergence of fictitious play for diagonal payoff matrices. In *Proceedings of the Thirty-Second Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '21, page 1387–1404, USA, 2021. Society for Industrial and Applied Mathematics. ISBN 9781611976465.

Rüdiger Achilles and Andrea Bonfiglioli. The early proofs of the theorem of Campbell, Baker, Hausdorff, and Dynkin. *Archive for History of Exact Sciences*, 66(3):295–358, April 2012. ISSN 1432-0657. doi: 10.1007/s00407-012-0095-8.

Shami A. Alsallami, Jitse Niesen, and Frank W. Nijhoff. Closed-form modified Hamiltonians for integrable numerical integration schemes. *Nonlinearity*, 31(11):5110–5146, 2018. doi: 10.1088/1361-6544/aad9ac.

Vladimir I. Arnold, Valery V. Kozlov, and Anatoly I. Neishtadt. *Mathematical Aspects of Classical and Celestial Mechanics*, volume 3 of *Encyclopaedia of Mathematical Sciences*. Springer, Berlin, Germany, 3rd edition, 2006. doi: 10.1007/978-3-540-48926-9.

Kendall E. Atkinson. *An Introduction to Numerical Analysis*. Wiley, New York, 1988. ISBN 978-0471624899.

James P Bailey and Georgios Piliouras. Multi-agent learning in network zero-sum games is a Hamiltonian system. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*, pages 233–241, 2019.

James P. Bailey, Gauthier Gidel, and Georgios Piliouras. Finite regret and cycles with fixed step-size via alternating gradient descent-ascent. In *Proceedings of Thirty Third Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pages 391–407. PMLR, July 2020. URL `https://proceedings.mlr.press/v125/bailey20a.html`.

David Balduzzi, Sebastien Racaniere, James Martens, Jakob Foerster, Karl Tuyls, and Thore Graepel. The mechanics of n-player differentiable games. In *International Conference on Machine Learning*, pages 354–363. PMLR, 2018.

Abdelmalek Bedhouche and Bakir Farhi. On some products taken over prime numbers. *Journal of Integer Sequences*, 26(1), January 2023. URL `https://cs.uwaterloo.ca/journals/JIS/VOL26/Farhi/farhi32.pdf`.

Michael Betancourt. Identifying the optimal integration time in Hamiltonian Monte Carlo. *arXiv preprint arXiv:1601.00225*, 2016.

Michael Betancourt. The convergence of Markov chain Monte Carlo methods: from the Metropolis method to Hamiltonian Monte Carlo. *Annalen der Physik*, 531(3):1700214, 2019.

Michael Betancourt, Michael I Jordan, and Ashia C Wilson. On symplectic optimization. *arXiv preprint arXiv:1802.03653*, 2018.

A. J. Bosch. The factorization of a square matrix into two symmetric matrices. *The American Mathematical Monthly*, 93(6):462–464, 1986. URL `http://www.jstor.org/stable/2323471`.

Nawaf Bou-Rabee and Jesús Maria Sanz-Serna. Geometric integrators and the Hamiltonian Monte Carlo method. *Acta Numerica*, 27:113–206, 2018.

G.C. Calafiore and L. El Ghaoui. *Optimization Models*. Control systems and optimization series. Cambridge University Press, October 2014.

Bob Carpenter, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus A Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan: A probabilistic programming language. *Journal of statistical software*, 76, 2017.

Fernando Casas and Ander Murua. An efficient algorithm for computing the Baker–Campbell–Hausdorff series and some of its applications. *Journal of Mathematical Physics*, 50(3):033513, March 2009. ISSN 0022-2488. doi: 10.1063/1.3078418.

P. J. Channell and C. Scovel. Symplectic integration of Hamiltonian systems. *Nonlinearity*, 3(2):231–259, May 1990. doi: 10.1088/0951-7715/3/2/001.

Yuansi Chen, Raaz Dwivedi, Martin J. Wainwright, and Bin Yu. Fast mixing of Metropolized Hamiltonian Monte Carlo: Benefits of multi-step gradients. *JMLR*, 2020.

Zongchen Chen and Santosh S Vempala. Optimal convergence rate of Hamiltonian Monte Carlo for strongly logconcave distributions. *International Conference on Randomization and Computation (RANDOM)*, 2019.

G Bruno De Luca, Alice Gatti, and Eva Silverstein. Improving energy conserving descent for machine learning: Theory and practice. *arXiv preprint arXiv:2306.00352*, 2023.

Jelena Diakonikolas and Michael I Jordan. Generalized momentum-based methods: A Hamiltonian perspective. *SIAM Journal on Optimization*, 31(1):915–944, 2021.

Simon Duane, A. D. Kennedy, Brian J. Pendleton, and Duncan Roweth. Hybrid Monte Carlo. *Physics Letters B*, 1987.

Valentin Duruisseaux and Melvin Leok. Accelerated optimization on Riemannian manifolds via discrete constrained variational integrators. *Journal of Nonlinear Science*, 32(4):42, 2022.

Robert D. Engle, Robert D. Skeel, and Matthew Drees. Monitoring energy drift with shadow Hamiltonians. *Journal of Computational Physics*, 206(2):432–452, July 2005. doi: 10.1016/j.jcp.2004.12.009.

Chris M. Field and Frank W. Nijhoff. A note on modified Hamiltonians for numerical integrations admitting an exact invariant. *Nonlinearity*, 16(5):1673–1683, 2003. doi: 10.1088/0951-7715/16/5/308.

Gauthier Gidel, Reyhane Askari Hemmat, Mohammad Pezeshki, Rémi Le Priol, Gabriel Huang, Simon Lacoste-Julien, and Ioannis Mitliagkas. Negative momentum for improved game dynamics. In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pages 1802–1811. PMLR, 16–18 Apr 2019. URL `https://proceedings.mlr.press/v89/gidel19a.html`.

Suriya Gunasekar, Blake Woodworth, and Nathan Srebro. Mirrorless mirror descent: A natural derivation of mirror descent. In Arindam Banerjee and Kenji Fukumizu, editors, *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pages 2305–2313. PMLR, 13–15 Apr 2021. URL `https://proceedings.mlr.press/v130/gunasekar21a.html`.

Ernst Hairer, Gerhard Wanner, and Christian Lubich. *Geometric Numerical Integration*. Springer, 2006.

Brian C. Hall. *Lie groups, Lie algebras, and representations: An elementary introduction*. Springer, 2015.

Josef Hofbauer. Evolutionary dynamics for bimatrix games: A Hamiltonian system? *Journal of mathematical biology*, 34:675–688, 1996.

Matthew D Hoffman and Andrew Gelman. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15: 1351–1381, 2014.

Yoopyo Hong and Roger A. Horn. The Jordan canonical form of a product of a Hermitian and a positive semidefinite matrix. *Linear Algebra and its Applications*, 147:373–386, 1991. ISSN 0024-3795. doi: 10.1016/0024-3795(91)90239-S.

Nathan Jacobson. *Lie Algebras*. Dover Books on Mathematics. Dover Publications, Mineola, NY, December 1979.

Jorge V. Jose and Eugene J. Saletan. *Classical dynamics*. Cambridge University Press, Cambridge, England, August 1998.

Jonas Katona and Xiuyuan Wang. AMD: Code for "A symplectic analysis of alternating mirror descent". `https://github.com/jonaskat87/AMD`, 2025. Accessed: 2025-11-12.

Jonas Katona, Xiuyuan Wang, and Andre Wibisono. A symplectic analysis of alternating mirror descent. *arXiv preprint arXiv:2405.03472*, 2024.

Yunbum Kook, Yin-Tat Lee, Ruoqi Shen, and Santosh Vempala. Sampling with Riemannian Hamiltonian Monte Carlo in a constrained space. *Advances in Neural Information Processing Systems*, 35:31684–31696, 2022.

Vladimir F. Lazutkin. *KAM Theory and Semiclassical Approximations to Eigenfunctions*. Springer Berlin Heidelberg, 1993. ISBN 9783642762475. doi: 10.1007/978-3-642-76247-5.

Yin Tat Lee and Santosh S Vempala. Convergence rate of Riemannian Hamiltonian Monte Carlo and faster polytope volume computation. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 1115–1121, 2018.

Yin Tat Lee, Ruoqi Shen, and Kevin Tian. Logsmooth gradient concentration and tighter runtimes for Metropolized Hamiltonian Monte Carlo. In *Conference on learning theory*, pages 2565–2597. PMLR, 2020.

Songtao Lu, Rahul Singh, Xiangyi Chen, Yongxin Chen, and Mingyi Hong. Alternating gradient descent ascent for nonconvex min-max problems in robust learning and GANs. In *2019 53rd Asilomar Conference on Signals, Systems, and Computers*, pages 680–684, 2019. doi: 10.1109/IEEECONF44664.2019.9048943.

Oren Mangoubi and Nisheeth Vishnoi. Dimensionally tight bounds for second-order Hamiltonian Monte Carlo. *Advances in Neural Information Processing Systems*, 31, 2018.

Aaron Meurer, Christopher P. Smith, Mateusz Paprocki, Ondřej Čertík, Sergey B. Kirpichev, Matthew Rocklin, Amit Kumar, Sergiu Ivanov, Jason K. Moore, Sartaj Singh, Thilina Rathnayake, Sean Vig, Brian E. Granger, Richard P. Muller, Francesco Bonazzi, Harsh Gupta, Shivam Vats, Fredrik Johansson, Fabian Pedregosa, Matthew J. Curry, Andy R. Terrel, Štěpán Roučka, Ashutosh Saboo, Isuru Fernando, Sumith Kulal, Robert Cimrman, and Anthony Scopatz. SymPy: Symbolic computing in Python. *PeerJ Computer Science*, 3:e103, January 2017. ISSN 2376-5992. doi: 10.7717/peerj-cs.103.

Willard Miller. *Symmetry Groups and Their Applications*. ISSN. Elsevier Science, 1973. ISBN 9780080873657.

Radford M Neal. MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, 2(11):2, 2011.

Arkadi Nemirovski. Prox-method with rate of convergence o(1/t) for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. *SIAM Journal on Optimization*, 15(1):229–251, 2004. doi: 10.1137/S1052623403425629.

A. S. Nemirovsky and D. B. Yudin. *Problem complexity and method efficiency in optimization*. Wiley-Interscience Series in Discrete Mathematics. John Wiley & Sons, Chichester, England, April 1983.

Brendan O'Donoghue and Chris J Maddison. Hamiltonian descent for composite objectives. *Advances in Neural Information Processing Systems*, 32, 2019.

OEIS Foundation Inc. A091137. From The On-Line Encyclopedia of Integer Sequences, 2024. URL `http://oeis.org/A091137`. Last visited on 01/29/2024.

Georg Ostrovski and Sebastian van Strien. Piecewise linear Hamiltonian flows associated to zero-sum games: Transition combinatorics and questions on ergodicity. *Regular and Chaotic Dynamics*, 16:128–153, 2011.

Georgios Piliouras and Xiao Wang. Constants of motion: The antidote to chaos in optimization and game dynamics. *arXiv preprint arXiv:2109.03974*, 2021.

Garvesh Raskutti and Sayan Mukherjee. The information geometry of mirror descent. *IEEE Transactions on Information Theory*, 61(3):1451–1457, 2015. doi: 10.1109/TIT.2015.2388583.

Ralph Tyrell Rockafellar. *Convex analysis*. Princeton Landmarks in Mathematics and Physics. Princeton University Press, Princeton, NJ, December 1996.

Kenneth A. Ross. *Elementary Analysis: The Theory of Calculus*. Springer New York, New York, NY, 2013. ISBN 978-1-4614-6271-2. doi: 10.1007/978-1-4614-6271-2_4.

John Salvatier, Thomas V Wiecki, and Christopher Fonnesbeck. Probabilistic programming in Python using PyMC3. *PeerJ Computer Science*, 2:e55, 2016.

Yuzuru Sato, Eizo Akiyama, and J Doyne Farmer. Chaos in learning a simple two-person game. *Proceedings of the National Academy of Sciences*, 99(7):4748–4751, 2002.

Robert D. Skeel and David J. Hardy. Practical construction of modified Hamiltonians. *SIAM Journal on Scientific Computing*, 23(4):1172–1188, 2001. doi: 10.1137/S106482750138318X.

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

Yuri B. Suris. Integrable mappings of the standard type. *Functional Analysis and Its Applications*, 23(1):74–76, 1989. doi: 10.1007/bf01078586.

Andrew R Teel, Jorge I Poveda, and Justin Le. First-order optimization algorithms with resets and Hamiltonian flows. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 5838–5843. IEEE, 2019.

J. v. Neumann. Zur theorie der gesellschaftsspiele. *Mathematische Annalen*, 100(1):295–320, December 1928. ISSN 1432-1807. doi: 10.1007/bf01448847.

Sebastian van Strien. Hamiltonian flows with random-walk behaviour originating from zero-sum games and fictitious play. *Nonlinearity*, 24(6):1715, 2011.

V. S. Varadarajan. *Lie Groups, Lie Algebras, and Their Representations.* Springer New York, 1984. doi: 10.1007/978-1-4612-1126-6.

Jun-Kun Wang. Hamiltonian descent and coordinate Hamiltonian descent. *arXiv preprint arXiv:2402.13988*, 2024.

Andre Wibisono, Ashia C Wilson, and Michael I Jordan. A variational perspective on accelerated methods in optimization. *Proceedings of the National Academy of Sciences*, 113(47):E7351–E7358, 2016.

Andre Wibisono, Molei Tao, and Georgios Piliouras. Alternating mirror descent for constrained min-max games. In *Advances in Neural Information Processing Systems*, volume 35, pages 35201–35212. Curran Associates, Inc., 2022. URL `https://proceedings.neurips.cc/paper_files/paper/2022/file/e496e0ce207ba9cdcc7d79bd499db67e-Paper-Conference.pdf`.

Ashia C Wilson, Lester Mackey, and Andre Wibisono. Accelerating rescaled gradient descent: Fast optimization of smooth functions. *Advances in Neural Information Processing Systems*, 32, 2019.

Ashia C Wilson, Ben Recht, and Michael I Jordan. A Lyapunov analysis of accelerated methods in optimization. *The Journal of Machine Learning Research*, 22(1):5040–5073, 2021.

Haruo Yoshida. Construction of higher order symplectic integrators. *Physics Letters A*, 150 (5):262–268, 1990. ISSN 0375-9601. doi: 10.1016/0375-9601(90)90092-3.

Haruo Yoshida. Recent progress in the theory and application of symplectic integrators. *Celestial Mechanics and Dynamical Astronomy*, 56(1-2):27–43, 1993. doi: 10.1007/bf00699717.

Guodong Zhang, Yuanhao Wang, Laurent Lessard, and Roger B. Grosse. Near-optimal local convergence of alternating gradient descent-ascent for minimax optimization. In *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pages 7659–7679. PMLR, March 2022. URL `https://proceedings.mlr.press/v151/zhang22e.html`.