HODGE-PROJECTED MULTI-AGENT LEARNING: REDUCING CYCLIC POLICY DYNAMICS VIA POTENTIAL-FLOW PROJECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

General-sum multi-agent learning is often governed by a stacked update field in which each agent's policy update changes the optimization landscape faced by the others. This coupling can entangle an integrable component of collective improvement with cyclic interaction dynamics, leading to slow or unstable multi-agent learning. Existing approaches, such as regularization, credit assignment, and consensus methods, stabilize MARL through local or algorithmic modifications; HPML complements them by projecting the joint update field onto a metric-gradient component. We introduce **HPML** (**H**odge-**P**rojected **M**ulti-agent **L**earning), which views the joint update field of a multi-agent system as an element of an L^2 space of vector fields and computes a Hodge-type projection onto the closest metric-gradient potential flow. HPML follows the projected component as the update direction, yielding the closest metric-gradient field under the chosen metric and sampling measure. The projection is defined variationally, characterized by a Poisson-type equation, and implemented through graph-based and amortized neural realizations that recover projected directions from samples. We show that the projected dynamics admit a Lyapunov potential and yield equilibrium-gap bounds with an explicit additive non-potentiality term. Controlled experiments validate circulation suppression and component recovery, and CTDE benchmarks show higher final normalized return for HPML-MAPPO than the matched MAPPO baseline on the evaluated suite.

1 INTRODUCTION

Reinforcement learning (RL) is a standard framework for sequential decision-making, and its multi-agent extension addresses networked control, robotic coordination, and strategic interaction (Puterman, 2014; Sutton et al., 1998; Shapley, 1953; Littman, 1994; Buşoniu et al., 2010). In single-agent RL, optimization is typically organized around a scalar objective, which supports Lyapunov-style analysis under standard smoothness conditions (Puterman, 2014; Bertsekas, 1997). General-sum multi-agent learning is driven by a stacked update field that may contain non-potential components and cyclic interaction dynamics even near stationary points (Balduzzi et al., 2018; Monderer & Shapley, 1996).

This issue is particularly visible in the centralized-training/decentralized-execution (CTDE) paradigm. PPO-style MARL methods produce decentralized actor updates from centralized critics or advantages (Schulman et al., 2015a; 2017; 2015b; Yu et al., 2022), and actor-critic or value-factorization baselines such as MADDPG, COMA, QMIX, and VDN induce coupled evaluations and updates across agents (Lowe et al., 2017; Foerster et al., 2018; Rashid et al., 2020; Sunehag et al., 2017). Since each agent experiences non-stationary dynamics due to the concurrent learning and policy updates of other agents, the stacked actor-update field can contain asymmetric cross-agent response terms. Even when each local update is well estimated, the stacked field can contain a non-potential component that carries closed-loop circulation. Local Jacobian antisymmetry diagnoses this effect pointwise; HPML instead targets a global projected update over the region visited by learning (Balduzzi et al., 2018; Mescheder et al., 2017; Gidel et al., 2018). Such cyclic components can slow learning, destabilize discrete-time updates, and drive recurrent policy profiles or

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

failed coordination patterns (Zinkevich et al., 2005; Balduzzi et al., 2018; Monderer & Shapley, 1996).

We propose **HPML**, a Hodge-inspired projection layer for multi-agent update fields. HPML views the supplied joint update field as an element of an $L^2(M, \mu)$ space of vector fields, where M specifies the metric and μ the sampling measure, and projects it onto the closest metric-gradient potential flow. Unlike finite-game decompositions of payoff differences (Candogan et al., 2011) or local Jacobian decompositions (Balduzzi et al., 2018), HPML operates directly on the realized learning field over the region visited by training and returns an executable projected update direction. The orthogonal residual is the non-potential component left unexplained by any metric-gradient flow and carries closed-loop circulation (Hodge, 1989; Eckmann, 1944). This gives a geometric intervention complementary to regularization, consensus, extra-gradient, Mirror-Prox, optimistic, and momentum-based methods (Korpelevich, 1977; Nemirovski, 2004; Popov, 1980; Mertikopoulos et al., 2018; Mescheder et al., 2017; Gidel et al., 2018; Daskalakis et al., 2017).

The projection admits a Poisson-type characterization, a graph-Laplacian realization from samples (Chung, 1997; Von Luxburg, 2007; Spielman, 2010; Hawrylycz, 2012; Lim, 2020), and an amortized neural realization that recovers projected directions by automatic differentiation. The projected dynamics admit a Lyapunov potential, and the residual non-potential component contributes an explicit additive term to the Variational Inequality (VI) gap (Facchinei & Pang, 2003; Nemirovski, 2004). In the CTDE paradigm, this residual measures the obstruction to a joint greedy-improvement interpretation. The main contributions are a global projection of realized MARL update fields, continuous/discrete/neural realizations from samples, original-operator VI-gap guarantees with explicit residual terms, and a CTDE integration compatible with PPO/MAPPO-style actor-update pipelines. Controlled experiments show reduced circulation and accurate potential-component recovery, while Melting Pot experiments show higher final normalized return on the evaluated suite.

2 PRELIMINARIES

This section provides the policy-space operator notation used throughout the paper. We define the feasible policy set, the VI-based equilibrium metric, and a local Jacobian-based integrability diagnostic that motivates the global projection developed later. Additional remarks on operator choice, unconstrained parameterizations, VI variants, and the canonical bilinear example are deferred to Appendix B; see in particular Remarks B.1, B.2, B.3, and Example B.5.

Markov Games and Policy Space. We consider a discounted finite Markov game with n agents $\mathcal{G} = \big(\mathcal{S}, \{\mathcal{A}_i\}_{i=1}^n, P, \{r_i\}_{i=1}^n, \gamma\big)$, $\gamma \in (0,1)$, where $\mathcal{S}$ is a finite state space, $\mathcal{A}_i$ is agent i's finite action set, $P(\cdot \mid s, a)$ is the transition kernel, and $r_i(s,a)$ is agent i's per-stage reward. A stationary policy for agent i is a mapping $\pi_i(\cdot \mid s) \in \Delta(\mathcal{A}_i)$, and we write $\pi = (\pi_1, \dots, \pi_n)$ for a joint policy. Let $\Pi := \prod_{i=1}^n \prod_{s\in\mathcal{S}} \Delta(\mathcal{A}_i)$ denote the stationary-policy set, viewed through the standard coordinate embedding as a compact convex set $\mathcal{X} \subset \mathbb{R}^d$. Each agent's discounted return under $\pi \in \Pi$ is $J_i(\pi) = \mathbb{E}_\pi\Big[\sum_{t\geq 0} \gamma^t r_i(s_t, a_t)\Big]$. We write $F : \mathcal{X} \to \mathbb{R}^d$ for a continuous joint field on policy space. A canonical game-theoretic choice is the negative pseudo-gradient operator $F(\pi) := \big(-\nabla_{\pi_1} J_1(\pi), \dots, -\nabla_{\pi_n} J_n(\pi)\big)$, which is the operator used in the VI formulation below. More generally, HPML can be applied to any continuous joint update field defined on $\mathcal{X}$. To avoid sign ambiguity, we distinguish the VI operator F_{VI} from an ascent/update field U. For the canonical game operator above, $U = -F_{\mathrm{VI}} = (\nabla_{\pi_i} J_i)_i$. The projection layer is applied to the supplied update field U; if one instead starts from a VI operator, the projected update direction is obtained by applying HPML to $-F_{\mathrm{VI}}$. In general-sum games, such a field need not be the gradient of any single scalar potential on $\mathcal{X}$; this lack of global integrability is the source of the cyclic behavior that HPML targets. See Remark B.1 and Remark B.2 for the distinction between the equilibrium operator and implementation-level parameter updates.

2.1 VARIATIONAL INEQUALITY AND DUALITY GAP

Many first-order equilibrium notions for differentiable games can be expressed as a variational inequality (VI). Let $\mathcal{X} \subset \mathbb{R}^d$ be nonempty, closed, and convex, and let $F : \mathcal{X} \to \mathbb{R}^d$. Unless otherwise stated, $\langle u, v\rangle := u^\top v$ denotes the Euclidean inner product.

Definition 2.1 (Stampacchia VI)**.** The variational inequality $\mathrm{VI}(\mathcal{X}, F)$ is to find $x^\star \in \mathcal{X}$ such that $\langle F(x^\star), x - x^\star \rangle \geq 0, \forall x \in \mathcal{X}$.

When x parametrizes a joint policy π and $F(\pi) = (-\nabla_{\pi_i} J_i(\pi))_{i=1}^n$, a solution of $\mathrm{VI}(\mathcal{X}, F)$ satisfies the standard first-order Nash condition for a differentiable game. It is generally necessary for Nash equilibrium, and it becomes sufficient under additional structure such as playerwise concavity.

Definition 2.2 (Duality gap (Stampacchia gap))**.** For $\bar{x} \in \mathcal{X}$, define $\mathrm{Gap}(\bar{x}) := \max_{x \in \mathcal{X}} \langle F(\bar{x}), \bar{x} - x \rangle$.

We use $\mathrm{Gap}(\bar{x})$ as the equilibrium-quality metric throughout. By construction $\mathrm{Gap}(\bar{x}) \geq 0$, and $\mathrm{Gap}(\bar{x}) \leq \varepsilon$ is equivalent to $\langle F(\bar{x}), x - \bar{x} \rangle \geq -\varepsilon, \forall x \in \mathcal{X}$.

Lemma 2.3 (Zero gap implies a VI solution)**.** *If* $\mathrm{Gap}(\bar{x}) = 0$*, then* $\bar{x}$ *solves* $\mathrm{VI}(\mathcal{X}, F)$ *in the sense of Definition 2.1.*

See Remark B.3 and Remark B.4 for the relation to the Minty gap and to monotonicity-based merit-function interpretations.

2.2 Jacobian Symmetry and Local Integrability

Assume F is differentiable on $\mathrm{ri}(\mathcal{X})$, and let $J(x) := \nabla F(x) \in \mathbb{R}^{d \times d}$. We decompose

$$J(x) = S(x) + A(x), \qquad S(x) = \tfrac{1}{2}\big(J(x) + J(x)^\top\big), \qquad A(x) = \tfrac{1}{2}\big(J(x) - J(x)^\top\big).$$

The symmetric part $S(x)$ is the locally potential-like component of the linearization, while the antisymmetric part $A(x)$ is a local obstruction to integrability and a diagnostic for cyclic interaction dynamics.

Definition 2.4 (Antisymmetric energy)**.** The antisymmetric energy at x is $\mathcal{R}(x) := \|A(x)\|_F^2$, where $\|\cdot\|_F$ denotes the Frobenius norm.

Proposition 2.5 (Gradient fields have symmetric Jacobians; a converse on open sets)**.** *(i) If there exists a twice-differentiable potential* ϕ *such that* $F = \nabla\phi$ *on an open set containing* $\mathcal{X}$*, then* $J(x) = \nabla^2\phi(x)$ *is symmetric for all* x *in that open set; hence* $A(x) \equiv 0$ *and* $\mathcal{R}(x) \equiv 0$ *on* $\mathcal{X}$*.*

(ii) Conversely, if $U \subset \mathbb{R}^d$ *is open and simply connected,* $F \in C^1(U)$*, and* $A(x) \equiv 0$ *for all* $x \in U$*, then there exists a scalar potential* ϕ *on* U *such that* $F = \nabla\phi$ *(unique up to an additive constant).*

Proposition 2.5 shows that antisymmetry in $J(x)$ is a local obstruction to integrability. HPML will later replace this local diagnostic by a global L^2 projection at the vector-field level. See Example B.5 for the canonical bilinear cyclic-interaction example and Remark B.6 for the local/global distinction.

3 Hodge-Projected Multi-Agent Learning

HPML follows one chain: decompose the supplied field F into its closest metric-gradient field $G^\star$ and orthogonal residual R; recover $G^\star$ from samples by graph or neural realization; and update along $G^\star$. Section 6 then relates the projected dynamics to the original VI gap. We first define the projection and orthogonality; Poisson, circulation, and metric/sampling details are deferred to Appendix C.

3.1 Metric-Gradient Projection

Let $\mathcal{X} \subset \mathbb{R}^d$ be a nonempty closed convex set. Throughout Sections 3–5, F denotes the supplied update field to be projected. All continuous projection statements are understood in coordinates on the affine hull of $\mathcal{X}$; for product-simplex policy spaces this can be implemented by dropping redundant simplex coordinates or by working in an unconstrained parameterization such as logits. Fix a symmetric positive definite matrix $M \succ 0$ and a reference probability measure μ on $\mathcal{X}$. For vectors $u, v \in \mathbb{R}^d$, define $\langle u, v \rangle_M := u^\top M v, \qquad \|u\|_M := \sqrt{u^\top M u}$. For vector fields, write $L^2(M, \mu)$ for the space of $U : \mathcal{X} \to \mathbb{R}^d$ satisfying $\mathbb{E}_{x \sim \mu}[\|U(x)\|_M^2] < \infty$, equipped with $\langle U, V \rangle_{L^2(M,\mu)} := \mathbb{E}_{x \sim \mu}\big[\langle U(x), V(x) \rangle_M\big]$.

For a scalar potential $\Phi : \mathcal{X} \to \mathbb{R}$, the metric gradient is $\nabla_M \Phi(x) := M^{-1} \nabla \Phi(x)$.

Definition 3.1 (Metric-gradient least-squares projection)**.** Let $F : \mathcal{X} \to \mathbb{R}^d$ be square-integrable under μ. Define $\mathcal{G}^M_{\mathrm{grad}} := \{\nabla_M \Phi : \Phi \in H^1(\mathcal{X})\}$, and define $\Pi^M_{\mathrm{grad}} F$ as the minimizer-induced field of $\inf_{\Phi \in H^1(\mathcal{X})} \mathbb{E}_{x\sim\mu}\big[\|F(x) - \nabla_M \Phi(x)\|^2_M\big]$. If $\Phi^\star$ is any minimizer, the projected field is $\Pi^M_{\mathrm{grad}} F := \nabla_M \Phi^\star$, which is unique μ-a.e., while $\Phi^\star$ itself is defined up to an additive constant.

Definition 3.1 is the core object of HPML: it replaces the original joint field by its closest potential-flow approximation under the geometry induced by M and the sampling distribution μ. We assume the continuous variational problem admits a minimizer; otherwise, $\Pi^M_{\mathrm{grad}} F$ can be interpreted as the projection onto the $L^2(M, \mu)$-closure of metric-gradient fields. The finite graph and neural implementations used later are finite-dimensional approximations of this same objective. A weak Poisson-type characterization of the same projection is given in Proposition C.2.

3.2 Orthogonality and Residual Energy

Let $G^\star := \Pi^M_{\mathrm{grad}} F, R := F - G^\star$. The projected field and residual satisfy the following variational orthogonality property.

Lemma 3.2 (Normal equation / orthogonality)**.** *For every test potential $\Phi \in H^1(\mathcal{X})$, $\mathbb{E}_{x\sim\mu}\big[\langle R(x), \nabla_M \Phi(x)\rangle_M\big] = 0$. Equivalently, R is $L^2(M, \mu)$-orthogonal to all metric-gradient fields.*

Corollary 3.3 (Pythagorean identity)**.** *With $G^\star$ and R defined above, $\mathbb{E}_{x\sim\mu}\big[\|F(x)\|^2_M\big] = \mathbb{E}_{x\sim\mu}\big[\|G^\star(x)\|^2_M\big] + \mathbb{E}_{x\sim\mu}\big[\|R(x)\|^2_M\big]$.*

Corollary 3.3 separates the field into a potential component and an orthogonal non-potential residual. This is the Hodge-type structure used by HPML: an $L^2(M, \mu)$-orthogonal split into a metric-gradient component and a residual. On domains or complexes with richer topology, the residual can be further refined into curl-like and harmonic components; on the sample graph, the same idea appears as a cut-space/cycle-space decomposition of edge flows. HPML uses the coarser projected-gradient/residual split because the update only requires the integrable direction. Both components are relative to the chosen metric M and sampling measure μ, so E measures non-potentiality over the region represented by the data.

3.3 Non-Potentiality

Definition 3.4 (Non-potentiality)**.** Define the residual field $R(x) := F(x) - \Pi^M_{\mathrm{grad}} F(x)$ and its $L^2(M, \mu)$ magnitude $E := \Big(\mathbb{E}_{x\sim\mu}\big[\|R(x)\|^2_M\big]\Big)^{1/2}$. We call E the *non-potentiality* of F under metric M and sampling measure μ.

The quantity E is a global, geometry-aware measure of how far the joint update field deviates from a potential flow over the region emphasized by μ. If $E = 0$, then F agrees μ-a.e. with a metric-gradient field; if $E > 0$, the residual R captures the irreducible non-potential component that can sustain cyclic interaction dynamics. Closed-loop circulation is carried entirely by this residual; see Proposition C.5 and Remark C.6.

3.4 HPML Update

Let $\mathrm{Retr} : \mathbb{R}^d \to \mathcal{X}$ be a feasibility operator. HPML updates by following only the projected component:

$$x_{t+1} = \mathrm{Retr}\Big(x_t + \eta_t \, \Pi^M_{\mathrm{grad}} F(x_t)\Big). \tag{1}$$

In practice, the projection is computed approximately from samples by solving the same variational problem on a discrete structure or by fitting an amortized potential model. The concrete procedures are given in Algorithm 1 and Algorithm 3, with additional implementation remarks in Remark C.8.

4 SAMPLE-GRAPH PROJECTION

Section 3 defines the continuous projection variationally. In practice, however, the field is observed only at sampled iterates, so we solve the same least-squares problem on a finite sample graph. The main outputs of the discrete construction are a projected potential edge flow, an orthogonal residual, and a computable proxy for non-potentiality. Detailed graph construction, lifting formulas, and the full procedure are deferred to Appendix D; see Proposition D.4, Remark D.1, Remark D.2, and Algorithm 1.

4.1 GRAPH SAMPLES AND EDGE FLOW

Let $\{x^1, \dots, x^N\} \subset \mathcal{X}$ be sampled joint-policy points. We build a weighted undirected k-NN graph $G = (\mathcal{V}, \mathcal{E})$ on these samples, with node set $\mathcal{V} = \{1, \dots, N\}$ and symmetric edge weights $w_{ij} > 0$. Choose an orientation for each edge and let $\mathcal{E}^{\rightarrow}$ denote the oriented edge set.

We represent scalar potentials as node values $\phi \in \mathbb{R}^N$. Let $B \in \mathbb{R}^{|\mathcal{E}^{\rightarrow}| \times N}$ be the oriented incidence matrix, so that the discrete gradient is the edge flow $B\phi$.

Given field estimates $F_i := F(x^i) \in \mathbb{R}^d$, define the edge 1-form $\omega_e := \left\langle M \frac{1}{2}(F_i + F_j),\, x^j - x^i \right\rangle, \quad e = (i \to j) \in \mathcal{E}^{\rightarrow}$. This is the discrete analogue of the continuous 1-form $\langle MF, dx \rangle$. Its metric consistency with potential differences is formalized in Proposition D.4.

4.2 DISCRETE PROJECTION AND NON-POTENTIALITY

Let $W = \mathrm{diag}(w_e)$ be the diagonal matrix of edge weights and define $\|u\|_W^2 := u^\top W u$.

Definition 4.1 (Discrete potential-flow projection)**.** Given an edge flow ω, define $\phi^\star \in \arg\min_{\phi \in \mathbb{R}^N} \|\omega - B\phi\|_W^2$. The projected potential edge flow is $\omega_{\mathrm{pot}} := B\phi^\star$, and the residual component is $\omega_{\mathrm{cyc}} := \omega - \omega_{\mathrm{pot}}$.

The optimality condition yields the graph Poisson system $L\phi^\star = B^\top W \omega, \quad L := B^\top W B$, with $\phi^\star$ unique up to an additive constant on each connected component.

Lemma 4.2 (Orthogonality of the discrete residual)**.** *Let $\omega_{\mathrm{cyc}} := \omega - B\phi^\star$. Then $(B\psi)^\top W\, \omega_{\mathrm{cyc}} = 0, \forall \psi \in \mathbb{R}^N$, i.e., the residual is orthogonal to all discrete gradient flows.*

Corollary 4.3 (Discrete energy decomposition)**.** *With $\omega_{\mathrm{pot}} = B\phi^\star$ and $\omega_{\mathrm{cyc}} = \omega - \omega_{\mathrm{pot}}$, $\|\omega\|_W^2 = \|\omega_{\mathrm{pot}}\|_W^2 + \|\omega_{\mathrm{cyc}}\|_W^2$.*

The graph interpretation of ω_{cyc} as a cycle-space component is recorded in Remark D.1.

4.3 DISCRETE NON-POTENTIALITY

Definition 4.4 (Discrete non-potentiality proxy)**.** Define $\mathrm{NonPot}(\omega) := \frac{\|\omega_{\mathrm{cyc}}\|_W^2}{\|\omega\|_W^2 + \varepsilon}, \qquad \varepsilon > 0$. This quantity measures the fraction of edge-flow energy unexplained by any potential flow, and serves as a computable proxy for the continuous residual energy $\mathbb{E}_\mu[\|R(x)\|_M^2]$.

Lemma 4.5 (Basic properties of NonPot)**.** *For any ω and $\varepsilon > 0$, one has $0 \le \mathrm{NonPot}(\omega) \le 1$. Moreover, $\mathrm{NonPot}(\omega) \approx 0$ iff ω is approximately potential on the graph, and it increases with the energy of circulation components supported by cycles.*

When samples cluster around recurring joint behaviors, $\mathrm{NonPot}(\omega)$ can be read as an empirical certificate of convention-level cyclic interaction dynamics; see Remark D.2.

4.4 NODE-WISE DIRECTION RECOVERY

The graph solve returns a node potential $\phi^\star$ or, equivalently, a potential edge flow $B\phi^\star$. To update a particular iterate, we recover a node-wise direction by locally fitting potential differences and mapping the result back through M^{-1}. The exact least-squares formulas are given in equation 3–equation 4, and the full graph-based procedure appears in Algorithm 1.

5 AMORTIZED AND CTDE REALIZATIONS OF HPML

This section records the amortized potential model and the CTDE parameter-space instantiation; full procedures appear in Algorithms 1–4.

5.1 AMORTIZED POTENTIAL PROJECTION

To avoid repeated large linear solves, we also use an amortized alternative that parameterizes the scalar potential and recovers the projected direction by automatic differentiation.

Let $\Phi_\theta : \mathcal{X} \to \mathbb{R}$ be a neural potential model. Its induced metric-gradient field is $g_\theta(x) := \nabla_M \Phi_\theta(x) = M^{-1}\nabla_x \Phi_\theta(x)$. Given samples $\{x_k\}_{k=1}^B$ and field estimates $\{\widehat{F}(x_k)\}$, we fit Φ_θ by minimizing

$$\min_\theta \ \widehat{\mathcal{L}}(\theta) := \frac{1}{B}\sum_{k=1}^{B} \big\|\widehat{F}(x_k) - g_\theta(x_k)\big\|_M^2 + \lambda_{\mathrm{g}}\Big(\frac{1}{B}\sum_{k=1}^{B}\Phi_\theta(x_k)\Big)^2 + \lambda_{\mathrm{wd}}\|\theta\|_2^2. \tag{2}$$

The gauge term removes the additive-constant ambiguity of the potential. Once trained, the amortized projected update is $x_{t+1} = \mathrm{Retr}\big(x_t + \eta_t\, g_\theta(x_t)\big)$. The full amortized procedure is given in Algorithm 3; implementation schedules and metric variants are summarized in Remark E.1.

5.2 CTDE PARAMETER-SPACE REALIZATION

Let $\theta := (\theta_1, \dots, \theta_n) \in \mathbb{R}^d$ denote the concatenated actor parameters, and let $\widehat{F}(\theta) := \big(\widehat{g}_1(\theta), \dots, \widehat{g}_n(\theta)\big)$ be the stacked CTDE actor-update field. A canonical example is a centralized-advantage estimator of the form $\widehat{g}_i(\theta) := \mathbb{E}\Big[\nabla_{\theta_i} \log \pi_{\theta_i}(a_i \mid o_i)\, \widehat{A}(s,a)\Big]$. HPML projects the stacked field before the actor update: $g_t \approx \Pi_{\mathrm{grad}}^M \widehat{F}(\theta_t)$, $\theta_{i,t+1} = \mathrm{Retr}\big(\theta_{i,t} + \eta_t [g_t]_i\big)$, $i = 1, \dots, n$. This realizes HPML as a plug-in projection layer on top of standard CTDE updates without changing decentralized execution. The full training loop appears in Algorithm 4; the relation between policy-space theory and parameter-space implementation is discussed in Remark E.2. In implementation, HPML changes only the actor-update direction. Rollout collection, critic training, observation preprocessing, reward normalization, and decentralized execution remain the same as in the underlying MAPPO-style pipeline. The graph variant periodically builds a local sample graph from recent stacked update-field estimates, solves the graph Poisson system, and lifts node potentials to actor directions. The neural variant amortizes the same projection objective through a scalar potential network. Thus the additional cost comes from projection refresh and graph-lifting or neural inner optimization, while the critic, reward signal, policy architecture, and execution interface remain unchanged.

5.3 CTDE INTEGRABILITY INTERPRETATION

The decomposition $F = \Pi_{\mathrm{grad}}^M F + R$ gives a direct CTDE interpretation. The projected term is the component of the joint update field compatible with ascent of a single scalar potential, and the residual R is the non-potential part outside that integrable structure.

In CTDE, let z denote the chosen joint coordinates, either policy coordinates or actor-parameter coordinates, and let $F(z) = (g_1(z), \dots, g_n(z))$ denote the stacked actor-update field induced by critic-based local improvement directions. The next proposition states that the residual is exactly the obstruction to joint integrability. The key distinction below is between *centralized evaluation* and *exact centralized optimization*: a centralized critic or a single shared reward does not by itself eliminate the residual unless the realized joint actor-update field equals the exact gradient of one common scalar objective in the full joint parameter space.

Proposition 5.1 (Residual as an obstruction to joint greedy integrability)**.** *Let $U \subset \mathrm{ri}(\mathcal{X})$ be a connected open set on which F is continuously differentiable.*

(i) If $R \equiv 0$ on U, then there exists a scalar potential Φ on U such that $F = \Pi_{\mathrm{grad}}^M F = \nabla_M \Phi$ on U. Hence the joint actor-update field is integrable on U; in the Euclidean projected-ascent setting of Section 6, Φ is a Lyapunov function. (ii) If there exists a closed piecewise-C^1

loop $C \subset U$ such that $\oint_C \langle MF(x), dx\rangle = \oint_C \langle MR(x), dx\rangle \neq 0$, then no scalar potential Φ on U can satisfy $F = \nabla_M \Phi$ on U. Consequently, the simultaneous actor-update field has no exact representation as ascent of a single common scalar objective on U.

The local source of this circulation is asymmetry in the one-form MF. In the two-agent Euclidean case, it appears as a cross-response mismatch $\nabla_{x_2} g_1(x) \neq \nabla_{x_1} g_2(x)^\top$,

6 Projected Ascent and Equilibrium-Gap Bounds

We analyze HPML in a clean Euclidean setting ($M = I$ with Euclidean projection) to isolate the main effect of the projection layer. The theory has two steps: projected ascent along a potential flow yields Lyapunov improvement, and the residual non-potential component enters the original VI gap as an explicit additive term. Technical extensions and auxiliary variants are deferred to Appendix F; see Lemma F.1, Theorem F.3, Remark F.2, and Remark F.4.

Sign convention. Throughout this section, F denotes the VI operator in Definition 2.2. If $U(x)$ is an ascent direction, then $F(x) = -U(x)$; HPML constructs a potential Φ by projecting $U = -F$ so that the projected ascent direction is $\nabla\Phi$. Equivalently, the VI operator decomposes as $F = -\nabla\Phi + R$, where R is the residual obstruction to exact potential ascent.

Assumption 6.1 (Geometry and smoothness)**.** $\mathcal{X} \subset \mathbb{R}^d$ is nonempty, closed, convex, and compact with diameter $D := \max_{x,y\in\mathcal{X}} \|x - y\|$. The potential $\Phi : \mathcal{X} \to \mathbb{R}$ is differentiable and L-smooth: $\|\nabla\Phi(x) - \nabla\Phi(y)\| \leq L\|x - y\|, \quad \forall x, y \in \mathcal{X}$. By compactness and continuity of $\nabla\Phi$, define $G := \max_{x\in\mathcal{X}} \|\nabla\Phi(x)\| < \infty, \quad \Phi_{\max} := \max_{x\in\mathcal{X}} \Phi(x), \quad \Phi_{\min} := \min_{x\in\mathcal{X}} \Phi(x)$.

Assumption 6.2 (Residual bound)**.** The VI operator admits a decomposition $F(x) = -\nabla\Phi(x) + R(x)$, where the residual satisfies $\|R(x)\| \leq \varepsilon, \quad \forall x \in \mathcal{X}$. An L^2 version is given in Theorem F.3.

Assumption 6.3 (Projection accuracy)**.** At each step, HPML uses an ascent direction g_t such that $\|g_t - \nabla\Phi(x_t)\| \leq \delta_t$, where $\nabla\Phi$ is the exact projected ascent direction.

6.1 Lyapunov Improvement Under Potential Flow

We first isolate the ideal projected dynamics.

Theorem 6.4 (Monotone Lyapunov improvement for projected gradient ascent)**.** *Under Assumption 6.1, consider the exact projected ascent step $x_{t+1} = \mathrm{Proj}_{\mathcal{X}}\big(x_t + \eta\nabla\Phi(x_t)\big), \quad \eta \leq \frac{1}{L}$. Then $\Phi(x_{t+1}) \geq \Phi(x_t) + \left(\frac{1}{\eta} - \frac{L}{2}\right)\|x_{t+1} - x_t\|^2 \geq \Phi(x_t) + \frac{1}{2\eta}\|x_{t+1} - x_t\|^2$.*

Define the projected-step mapping $\mathcal{G}_\eta(x_t) := \frac{1}{\eta}(x_{t+1} - x_t), x_{t+1} = \mathrm{Proj}_{\mathcal{X}}(x_t + \eta\nabla\Phi(x_t))$. Then $\mathcal{G}_\eta(x) = 0$ iff x is first-order stationary for the constrained maximization of Φ. Moreover, Theorem 6.4 implies $\sum_{t=0}^{T-1} \|\mathcal{G}_\eta(x_t)\|^2 \leq \frac{2}{\eta}(\Phi_{\max} - \Phi_{\min})$, so the exact projected dynamics cannot sustain a nontrivial cycle. The projection inequality used in the proof is recorded separately in Lemma F.1.

6.2 Gap Bounds with a Residual Non-Potential Component

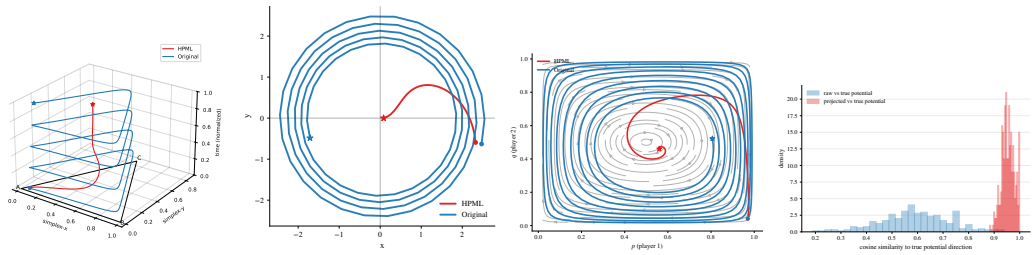
We now relate the projected dynamics to the VI gap of the original operator $F = -\nabla\Phi + R$.

Lemma 6.5 (Residual contribution to the VI gap)**.** *For any $x \in \mathcal{X}$, $\mathrm{Gap}(x) = \max_{u\in\mathcal{X}} \langle -\nabla\Phi(x) + R(x), x - u\rangle \leq \mathrm{Gap}_\Phi(x) + D\|R(x)\|$, where $\mathrm{Gap}_\Phi(x) := \max_{u\in\mathcal{X}} \langle -\nabla\Phi(x), x - u\rangle = \max_{u\in\mathcal{X}} \langle \nabla\Phi(x), u - x\rangle$.*

Lemma 6.6 (Gap bound via projected-step mapping)**.** *Under Assumption 6.1, let $x^+ = \mathrm{Proj}_{\mathcal{X}}(x + \eta\nabla\Phi(x)), \mathcal{G}_\eta(x) = \frac{1}{\eta}(x^+ - x)$. Then $\mathrm{Gap}_\Phi(x) \leq (D + \eta G)\,\|\mathcal{G}_\eta(x)\|$.*

Theorem 6.7 (Gap bound with additive non-potentiality)**.** *Under Assumptions 6.1 and 6.2, run the exact projected ascent $x_{t+1} = \mathrm{Proj}_{\mathcal{X}}(x_t + \eta\nabla\Phi(x_t)), \eta \leq \frac{1}{L}$. Let $\hat{t} \in \arg\min_{0\leq t\leq T-1} \|\mathcal{G}_\eta(x_t)\|, \hat{x}_T := x_{\hat{t}}$. Then $\mathrm{Gap}(\hat{x}_T) \leq (D + \eta G)\sqrt{\frac{2(\Phi_{\max} - \Phi_{\min})}{T\eta}} + D\varepsilon$.*

An L^2 output-law version of the same statement is given in Theorem F.3, and the circulation interpretation of the additive residual term is summarized in Remark F.2.



(a) RPS simplex (b) 2D field (c) Logistic game (d) 3D projection

Figure 1: **Mechanism tests.** HPML suppresses circulation in controlled games and recovers the known potential direction in the linear 3D test. The first three panels compare raw and projected trajectories or fields; the fourth reports cosine similarity to the potential component $g_{\mathrm{pot}}(z) = -z$.

Reading the residual term. The bound certifies the original VI operator, not only the projected potential flow. The first term is the stationarity error of the projected dynamics, controlled by Lyapunov improvement, while the additive term $D\varepsilon$ is the contribution of the non-potential residual in the original field. Thus a small residual transfers projected stationarity to the original game, whereas a large residual intentionally yields no strong equilibrium-fidelity claim. The L^2 version replaces the uniform residual bound by residual energy under the output-iterate law, matching the non-potentiality measure in Definition 3.4 and the graph proxy in Definition 4.4.

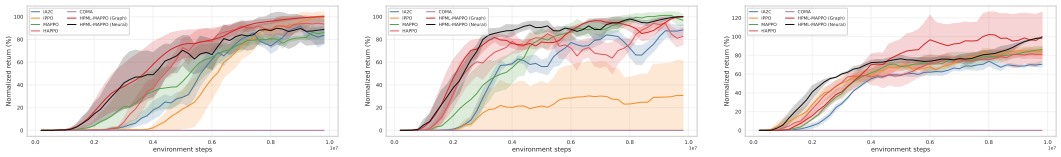
Inexact projection. Graph and neural projections are approximate in practice. Appendix F extends Theorem 6.7 to directions g_t satisfying $\|g_t - \nabla\Phi(x_t)\| \le \bar{\delta}$, adding projection-error terms while preserving the separation between structural non-potentiality and projection approximation error.

We finally quantify the effect of using an approximate projected direction g_t. Let $x_{t+1} = \mathrm{Proj}_{\mathcal{X}}(x_t + \eta g_t), \widetilde{\mathcal{G}}_\eta(x_t) := \frac{1}{\eta}(x_{t+1} - x_t)$.

Theorem 6.8 (Effect of projection error)**.** *Under Assumptions 6.1, 6.2, and 6.3, run* $x_{t+1} = \mathrm{Proj}_{\mathcal{X}}(x_t + \eta g_t), \eta \le \frac{1}{L}$, *and suppose* $\delta_t \le \bar{\delta}$ *for all* t. *Let* $\hat{t} \in \arg\min_{0\le t\le T-1} \|\widetilde{\mathcal{G}}_\eta(x_t)\|, \hat{x}_T := x_{\hat{t}}$. *Then* $\mathrm{Gap}(\hat{x}_T) \le \big(D + \eta(G + \bar{\delta})\big)\sqrt{\frac{4(\Phi_{\max} - \Phi_{\min})}{T\eta} + 4\bar{\delta}^2} + D(\varepsilon + \bar{\delta})$.

The auxiliary inexact gap-mapping estimate and a term-by-term interpretation of the bound are recorded in Lemma F.5 and Remark F.6.

7 Experiments



(a) Bach or Stravinsky (b) Cooking: Circuit (c) Clean Up

Figure 2: **Representative Melting Pot learning curves.** Normalized return versus environment steps on three scenarios covering convention selection, embodied coordination, and sequential social dilemmas. Shaded bands denote standard error over $S = 5$ seeds; remaining scenarios are reported in Appendix A.3.

We evaluate HPML in two regimes: controlled games where the potential and cyclic components are interpretable, and high-dimensional CTDE benchmarks where the projection is used as a plug-in actor-update layer.

Table 1: **Melting Pot summary.** Final normalized return (%) on three representative scenarios and the macro average over the listed scenarios. Full per-scenario results are given in Appendix A.4.

Method	BoS-Rep	CC-Circuit	Clean Up	Shown Avg
IA2C	83.7	88.3	70.6	80.9
IPPO	100.0	30.7	84.4	71.7
MAPPO	86.3	96.9	86.2	89.8
HAPPO	93.6	82.7	81.1	85.8
HPML-MAPPO (Graph)	100.0	100.0	100.0	100.0
HPML-MAPPO (Neural)	88.7	99.9	98.7	95.8

7.1 Mechanism Tests in Interpretable Games

We isolate the geometric mechanism in four controlled settings: cyclic matrix games, a two-dimensional potential-plus-skew field, a logistic 2×2 mixed-strategy game, and a linear 3D field with a known potential component. Figure 1 visualizes circulation suppression in RPS and the logistic game, while the 2D and 3D fields test recovery of analytically known potential components. We compare the raw and HPML-projected fields using circulation, path length, graph non-potentiality $\mathrm{NonPot}(\omega)$, and a gap/exploitability proxy when available. Exact settings and metrics are given in Appendix A.1. These tests verify the intended mechanism: HPML preserves the potential-flow component while reducing sample-graph circulation. The 3D test checks projection accuracy against a known potential component in addition to trajectory visualizations.

7.2 Melting Pot CTDE Benchmarks

We next evaluate HPML in high-dimensional, partially observed MARL on a 10-scenario Melting Pot suite (Leibo et al., 2021; Agapiou et al., 2022), including convention games, social dilemmas, and three fully cooperative Collaborative Cooking tasks. The main text shows Bach or Stravinsky Repeated, Collaborative Cooking: Circuit, and Clean Up; remaining curves, settings, and results are in Appendix A.2–A.5. Baselines are IA2C (Papoudakis et al., 2020), IPPO/MAPPO (Yu et al., 2022), and HAPPO (Kuba et al., 2021); COMA (Foerster et al., 2018) is retained in the full appendix table. HPML-Graph and HPML-Neural use MAPPO as the matched backbone. MAPPO and both HPML variants use identical agent-indexed, non-shared actors, critic, rollout budget, and observation interface. Curves show mean with standard-error bands over $S = 5$ seeds.

Table 1 shows Graph at the highest displayed final return on all three tasks; Neural is near ceiling on Circuit and Clean Up but lower on BoS. Circuit is already near ceiling for MAPPO, while the full table shows larger separations on harder coordination cases. Together with Figure 1, this separates mechanism verification from benchmark performance rather than equating lower residual energy with higher return.

8 Conclusion

We introduced Hodge-Projected Multi-Agent Learning (HPML), a geometric projection method that separates a joint multi-agent update field into a metric-gradient potential component and a non-potential residual. The projected component admits Lyapunov-style improvement, and the residual captures cyclic structure that appears explicitly in the VI-gap bound. Graph-based and amortized implementations make HPML usable as a CTDE plug-in; controlled tests show reduced circulation and component recovery, while the evaluated Melting Pot suite shows higher final normalized return than the matched MAPPO baseline. Key limitations are the dependence on the chosen metric, sampling distribution, field-estimation quality, and projection approximation error.

9 STATEMENTS

AI USE STATEMENT

In this work, we used generative AI tools primarily for language editing, including grammar correction, improving clarity and fluency, and refining the presentation of the manuscript. We also used generative AI tools to assist with running, debugging, and checking experimental code and computational workflows, as well as to help check mathematical derivations and improve the rigor, consistency, and clarity of proof presentation. Generative AI was not used to determine the core research ideas, theoretical results, experimental design, or scientific conclusions of this work. All AI-assisted mathematical arguments, code execution, outputs, and manuscript edits were independently inspected and verified by the authors. The authors take full responsibility for the accuracy, originality, reproducibility, and final content of the manuscript.

REFERENCES

John P Agapiou, Alexander Sasha Vezhnevets, Edgar A Duéñez-Guzmán, Jayd Matyas, Yiran Mao, Peter Sunehag, Raphael Köster, Udari Madhushani, Kavya Kopparapu, Ramona Comanescu, et al. Melting pot 2.0. *arXiv preprint arXiv:2211.13746*, 2022.

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*, pp. 354–363. PMLR, 2018.

Dimitri P Bertsekas. Nonlinear programming. *Journal of the Operational Research Society*, 48(3): 334–334, 1997.

Lucian Buşoniu, Robert Babuška, and Bart De Schutter. Multi-agent reinforcement learning: An overview. *Innovations in multi-agent systems and applications-1*, pp. 183–221, 2010.

Ozan Candogan, Ishai Menache, Asuman Ozdaglar, and Pablo A Parrilo. Flows and decompositions of games: Harmonic and potential games. *Mathematics of Operations Research*, 36(3):474–503, 2011.

Fan RK Chung. *Spectral graph theory*, volume 92. American Mathematical Soc., 1997.

Constantinos Daskalakis, Andrew Ilyas, Vasilis Syrgkanis, and Haoyang Zeng. Training gans with optimism. *arXiv preprint arXiv:1711.00141*, 2017.

Beno Eckmann. Harmonische funktionen und randwertaufgaben in einem komplex. *Commentarii Mathematici Helvetici*, 17(1):240–255, 1944.

Francisco Facchinei and Jong-Shi Pang. *Finite-dimensional variational inequalities and complementarity problems*. Springer, 2003.

Jakob Foerster, Gregory Farquhar, Triantafyllos Afouras, Nantas Nardelli, and Shimon Whiteson. Counterfactual multi-agent policy gradients. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

Gauthier Gidel, Hugo Berard, Gaëtan Vignoud, Pascal Vincent, and Simon Lacoste-Julien. A variational inequality perspective on generative adversarial networks. *arXiv preprint arXiv:1802.10551*, 2018.

Mike Hawrylycz. Discrete calculus: Applied analysis on graphs for computational science, leo grady, jonathan polimeni, springer (2010), $129.00, isbn: 978-1-84996-289-6, 2012.

William Vallance Douglas Hodge. *The theory and applications of harmonic integrals*. CUP Archive, 1989.

Galina M Korpelevich. Extragradient method for finding saddle points and other problems. *Matekon*, 13(4):35–49, 1977.

Jakub Grudzien Kuba, Ruiqing Chen, Muning Wen, Ying Wen, Fanglei Sun, Jun Wang, and Yaodong Yang. Trust region policy optimisation in multi-agent reinforcement learning. *arXiv preprint arXiv:2109.11251*, 2021.

Joel Z Leibo, Edgar A Dueñez-Guzman, Alexander Vezhnevets, John P Agapiou, Peter Sunehag, Raphael Koster, Jayd Matyas, Charlie Beattie, Igor Mordatch, and Thore Graepel. Scalable evaluation of multi-agent reinforcement learning with melting pot. In *International conference on machine learning*, pp. 6187–6199. PMLR, 2021.

Lek-Heng Lim. Hodge laplacians on graphs. *Siam Review*, 62(3):685–715, 2020.

Michael L Littman. Markov games as a framework for multi-agent reinforcement learning. In *Machine learning proceedings 1994*, pp. 157–163. Elsevier, 1994.

Ryan Lowe, Yi I Wu, Aviv Tamar, Jean Harb, OpenAI Pieter Abbeel, and Igor Mordatch. Multi-agent actor-critic for mixed cooperative-competitive environments. *Advances in neural information processing systems*, 30, 2017.

Panayotis Mertikopoulos, Bruno Lecouat, Houssam Zenati, Chuan-Sheng Foo, Vijay Chandrasekhar, and Georgios Piliouras. Optimistic mirror descent in saddle-point problems: Going the extra (gradient) mile. *arXiv preprint arXiv:1807.02629*, 2018.

Lars Mescheder, Sebastian Nowozin, and Andreas Geiger. The numerics of gans. *Advances in neural information processing systems*, 30, 2017.

Dov Monderer and Lloyd S Shapley. Potential games. *Games and economic behavior*, 14(1):124–143, 1996.

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.

Georgios Papoudakis, Filippos Christianos, Lukas Schäfer, and Stefano V Albrecht. Benchmarking multi-agent deep reinforcement learning algorithms in cooperative tasks. *arXiv preprint arXiv:2006.07869*, 2020.

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

Martin L Puterman. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.

Tabish Rashid, Mikayel Samvelyan, Christian Schroeder De Witt, Gregory Farquhar, Jakob Foerster, and Shimon Whiteson. Monotonic value function factorisation for deep multi-agent reinforcement learning. *Journal of Machine Learning Research*, 21(178):1–51, 2020.

John Schulman, Sergey Levine, Pieter Abbeel, Michael Jordan, and Philipp Moritz. Trust region policy optimization. In *International conference on machine learning*, pp. 1889–1897. PMLR, 2015a.

John Schulman, Philipp Moritz, Sergey Levine, Michael Jordan, and Pieter Abbeel. High-dimensional continuous control using generalized advantage estimation. *arXiv preprint arXiv:1506.02438*, 2015b.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

Lloyd S Shapley. Stochastic games. *Proceedings of the national academy of sciences*, 39(10): 1095–1100, 1953.

Daniel A Spielman. Algorithms, graph theory, and linear equations in laplacian matrices. In *Proceedings of the International Congress of Mathematicians 2010 (ICM 2010) (In 4 Volumes) Vol. I: Plenary Lectures and Ceremonies Vols. II–IV: Invited Lectures*, pp. 2698–2722. World Scientific, 2010.

Peter Sunehag, Guy Lever, Audrunas Gruslys, Wojciech Marian Czarnecki, Vinicius Zambaldi, Max Jaderberg, Marc Lanctot, Nicolas Sonnerat, Joel Z Leibo, Karl Tuyls, et al. Value-decomposition networks for cooperative multi-agent learning. *arXiv preprint arXiv:1706.05296*, 2017.

Richard S Sutton, Andrew G Barto, et al. *Reinforcement learning: An introduction*, volume 1. MIT press Cambridge, 1998.

Ulrike Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007.

Chao Yu, Akash Velu, Eugene Vinitsky, Jiaxuan Gao, Yu Wang, Alexandre Bayen, and Yi Wu. The surprising effectiveness of ppo in cooperative multi-agent games. *Advances in neural information processing systems*, 35:24611–24624, 2022.

Martin Zinkevich, Amy Greenwald, and Michael L. Littman. Cyclic equilibria in markov games. In *Advances in Neural Information Processing Systems 18 [Neural Information Processing Systems, NIPS 2005, December 5-8, 2005, Vancouver, British Columbia, Canada]*, pp. 1641–1648, 2005. URL https://proceedings.neurips.cc/paper/2005/hash/9752d873fa71c19dc602bf2a0696f9b5-Abstract.html.

A Additional Experimental Details and Full Results

A.1 Mechanism-test details

The controlled experiments in Section 7.1 use four diagnostic settings matching Figure 1. First, Rock–Paper–Scissors uses $K = 3$ and $\rho = 1$ to visualize circulation on the simplex. Second, the continuous field $g(z) = -\alpha z + \rho J z$ uses $\alpha = 0.05$ and $\rho = 2.5$, separating a known contractive potential component from a tunable skew component. Third, the logistic 2×2 mixed-strategy game uses $\rho = 2$ and maps logits (a, b) to $(p, q) = (\sigma(a), \sigma(b))$. Fourth, the linear 3D field $g(z) = -z + \rho S z$ uses $\rho = 2$ and a fixed skew-symmetric matrix S, so the exact potential component is $-z$.

For each setting, we compare the raw field with its sample-graph HPML projection and track orbit diameter, path length, $\mathrm{NonPot}(\omega)$, and a gap/exploitability proxy when available. The RPS panel is used only as a circulation-suppression test of the sampled learning field; it is not interpreted as a nonzero potential component of the underlying RPS payoff game. Claims of component recovery are supported by the analytically specified 2D and 3D fields.

A.2 Implementation details

This appendix provides the implementation details omitted from the main text, including Melting Pot scenario handling, observation preprocessing, CTDE training configuration, baseline-specific settings, HPML graph construction, neural-projection optimization, and evaluation protocol.

Melting Pot specifics. For Melting Pot, we follow a unified preprocessing and rollout interface across all scenarios. We keep the actor/critic backbone fixed across scenarios unless otherwise stated, and only adjust environment-dependent quantities such as action masking, episode horizon, and reward normalization.

Actor parameterization and baseline matching. The primary controlled comparison is MAPPO versus HPML-MAPPO. MAPPO and both HPML variants use identical agent-indexed, non-shared actor parameterizations, the same centralized critic, and the same rollout and observation interfaces; HPML changes only the stacked actor direction. We do not apply parameter sharing to only one side of this comparison because doing so would change both the hypothesis class and the joint coordinate space. Other baselines use comparable capacity where their algorithmic structure permits.

Training protocol. We train each method for a fixed environment-step budget on every Melting Pot scenario, using S seeds (default S=5). We report mean and standard error across seeds. Learning curves are smoothed only for visualization.

Baseline provenance. IA2C follows the independent actor–critic realization used in MARL benchmarking (Papoudakis et al., 2020); IPPO and MAPPO follow the PPO-based formulations of Yu et al. (2022); HAPPO follows Kuba et al. (2021); and COMA follows Foerster et al. (2018). COMA is retained as a canonical counterfactual-credit baseline, but its zero returns under the common protocol are not used as evidence for HPML; the principal matched comparison is MAPPO versus HPML-MAPPO.

HPML specifics. For HPML-Graph, we build a k-NN graph over a buffer of recent policy/parameter points $\{x^i\}$ and estimate edge flows $\omega_{ij} = \langle M \frac{F_i + F_j}{2}, x^j - x^i \rangle$ (Section 4). We solve the graph Poisson system with a fixed gauge and lift the potential to node directions via local least squares. For HPML-Neural, we fit a scalar potential network Φ_θ by minimizing the empirical projection loss (Section 5.1) and compute projected directions by autograd.

Hyperparameters. We sweep a small grid for each method and report the best validation configuration. Table 2 lists the final hyperparameters used in the main experiments.

Training-time overhead and scaling. With N buffered points and fixed k, the graph variant has $O(kN)$ edges and solves a sparse $N \times N$ Laplacian system; actor dimension affects neighbor search,

Table 2: Hyperparameters used in Melting Pot experiments.

Hyperparameter	Value	Notes
Total env steps	10,000,000	per scenario
Rollout length / batch size	128 / 256 $\times$ n_{agents}	2 vectorized envs; minibatch size 4096
Actor LR / critic LR	3×10^{-4} / 3×10^{-4}	Adam, $\epsilon = 10^{-5}$
PPO clip ϵ	0.2	for PPO-style methods
Entropy coef	0.01	
HPML graph k	4	kNN degree
HPML refresh rate	8	actor update/minibatch steps between projections
HPML ridge λ (lifting)	10^{-4}	stability
Neural potential LR / inner steps	10^{-3} / 2	if HPML-Neural

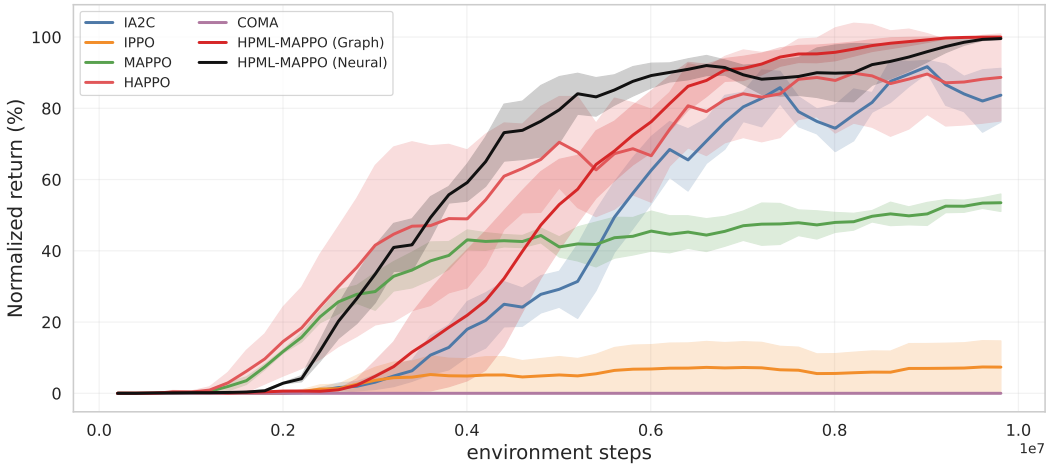


Figure 3: Normalized return versus environment steps on **Pure Coordination Repeated**.

edge-flow evaluation, and local lifting but not the Laplacian dimension. In our configuration the projection is refreshed once every eight actor-minibatch updates. HPML-Neural replaces the graph solve with two inner scalar-potential updates and one input-gradient evaluation. Both variants add training-time projection cost only; decentralized execution is unchanged.

Compute resources and software. All Melting Pot experiments were run on a single machine equipped with four NVIDIA RTX 6000 Ada Generation GPUs, each with 49,140 MiB of memory, an AMD EPYC 7713 64-Core Processor CPU, and 1.0 TiB of RAM. Software versions were Python 3.13.9, NumPy 2.4.2, and CUDA 13.0 as reported by NVIDIA-SMI.

Existing assets and licenses. The experiments use the Melting Pot benchmark (Leibo et al., 2021; Agapiou et al., 2022), which is distributed under the Apache-2.0 license. We use the benchmark only for evaluation and do not redistribute modified benchmark assets. The anonymized supplementary code package includes its own license and documentation.

A.3 Additional Melting Pot learning curves (remaining seven scenarios)

Figure 2 in the main text shows three representative scenarios (`bos_repeated`, `collaborative_cooking_circuit`, and `clean_up`). Here we report the remaining seven scenarios.

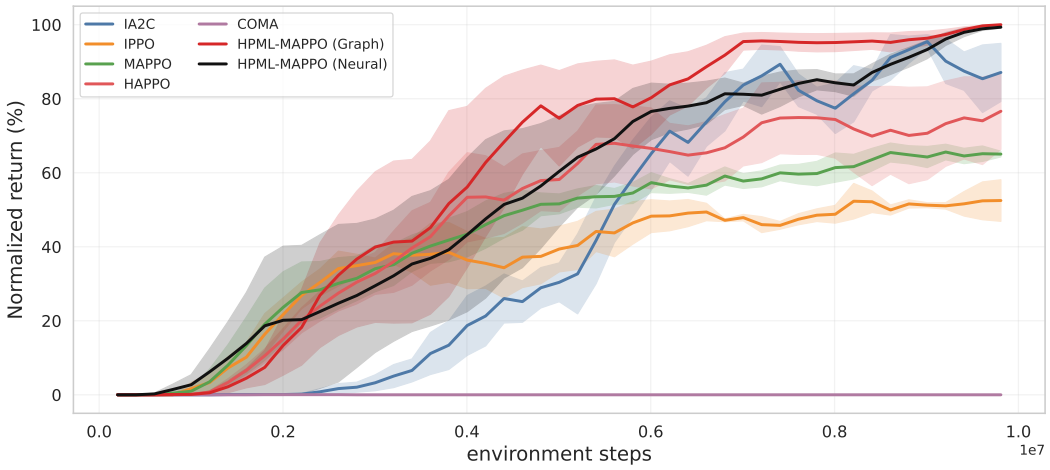


Figure 4: Normalized return versus environment steps on **Rationalizable Coordination Repeated**.

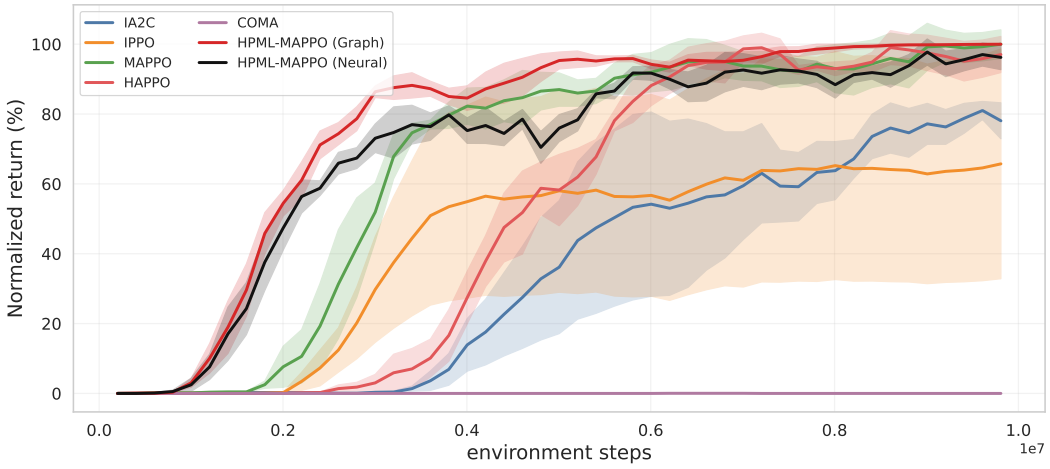


Figure 5: Normalized return versus environment steps on **Stag Hunt Repeated**.

A.4 Full Melting Pot table (per-scenario)

Interpreting scenario variation. The three Collaborative Cooking scenarios are fully cooperative but stress different coordination geometries. MAPPO is near ceiling on Circuit and Figure Eight (96.9 and 100.0) but substantially lower on Crowded (42.1), whereas HPML-Graph obtains 100.0 on Circuit and Crowded and 96.2 on Figure Eight. This pattern shows why cooperation alone does not determine the effect of projection: near-ceiling cases leave little room for change, while bottlenecks and role interference can induce stronger cross-agent coupling. We therefore use the controlled fields to establish the geometric mechanism and the benchmark suite to assess practical effect; task labels or return values alone are not treated as evidence of circulation.

A.5 Residual diagnostics and scale dependence

The dimensionless graph diagnostic $\mathrm{NonPot}(\omega) = \|\omega_{\mathrm{cyc}}\|_W^2 / (\|\omega\|_W^2 + \varepsilon)$ is meaningful for a fixed graph construction and scale. By contrast, the raw numerator $\|\omega_{\mathrm{cyc}}\|_W^2$ changes with actor-field magnitude, parameter scaling, graph weights, and projection realization, so raw Graph and Neural residual energies should not be compared as if they were a common performance metric. We therefore separate geometric mechanism tests from benchmark return and do not infer task performance from cross-realization raw residual magnitude.

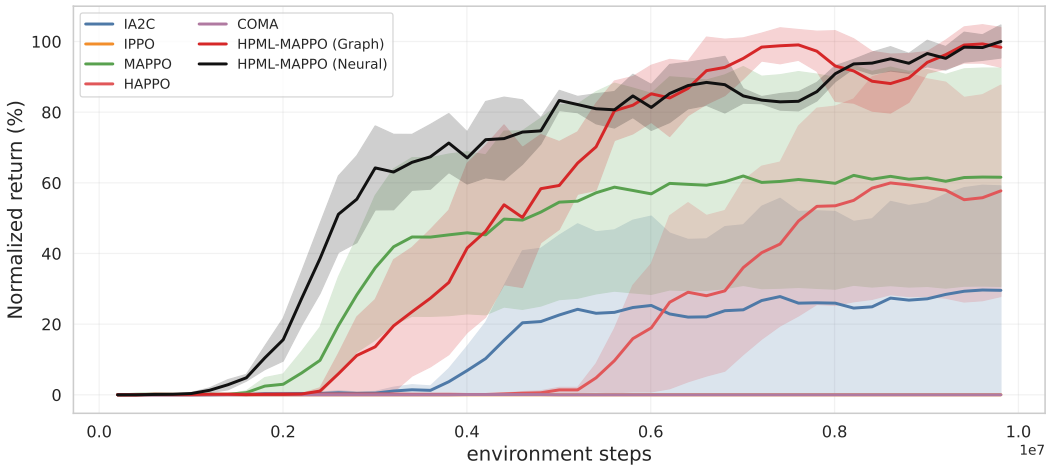


Figure 6: Normalized return versus environment steps on **Prisoners Dilemma Repeated**.

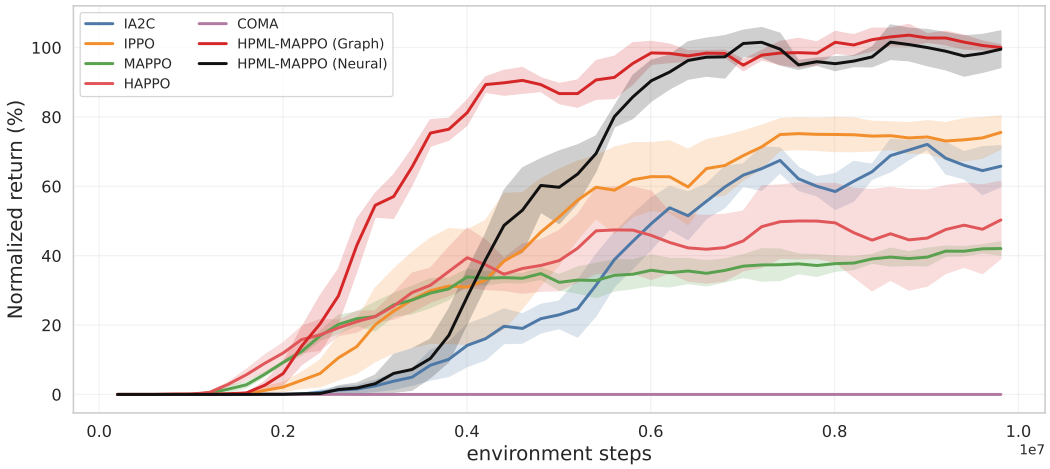


Figure 7: Normalized return versus environment steps on **Collaborative Cooking: Crowded**.

B Additional Notes for Section 2

Remark B.1 (Equilibrium operator and algorithmic update field). For equilibrium analysis, the canonical VI operator is the negative pseudo-gradient

$$F_{\mathrm{VI}}(\pi) = \big(-\nabla_{\pi_1} J_1(\pi), \ldots, -\nabla_{\pi_n} J_n(\pi) \big).$$

The corresponding simultaneous ascent/update field is $U = -F_{\mathrm{VI}}$. HPML projects the supplied update field U; if the available object is a VI operator, one first changes sign and applies the projection to $-F_{\mathrm{VI}}$. Some sections use F as a generic field to avoid excessive notation, but the intended sign is always determined by whether the object is being used for VI gap evaluation or for an actual update.

Remark B.2 (Primal policies and unconstrained parameterizations). The theory is stated on the compact convex feasible set $\mathcal{X}$ of stationary policies in primal coordinates. In implementations, one often optimizes unconstrained parameters such as logits and maps them back to $\mathcal{X}$ through a retraction or mirror map such as softmax. This changes the coordinate representation of the iterate, but not the conceptual role of F as a field defined on feasible policies.

Remark B.3 (Stampacchia and Minty gaps). Definition 2.2 uses the Stampacchia duality gap

$$\mathrm{Gap}(\bar{x}) = \max_{x \in \mathcal{X}} \langle F(\bar{x}), \bar{x} - x \rangle ,$$

which evaluates the operator at the reference point $\bar{x}$. Another common quantity is the Minty gap

$$\mathrm{Gap}_{\mathrm{Minty}}(\bar{x}) := \max_{x \in \mathcal{X}} \langle F(x), \bar{x} - x \rangle ,$$

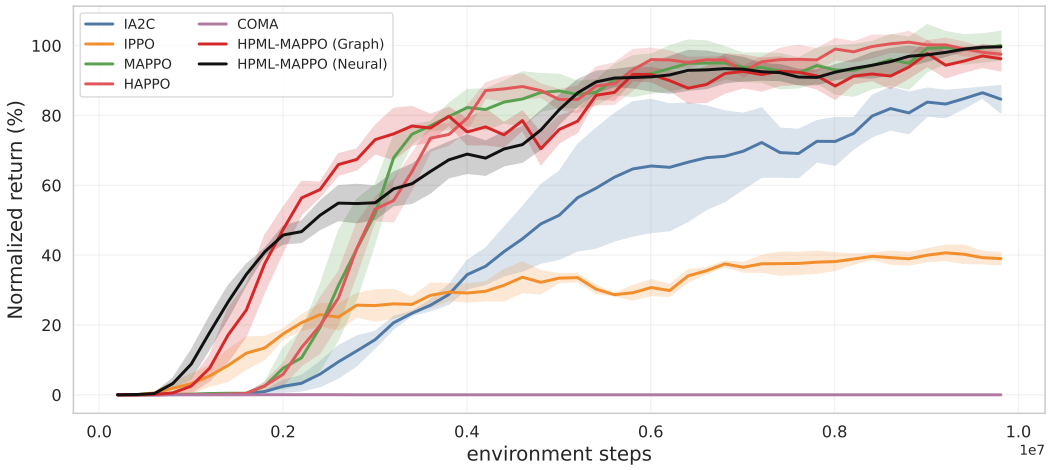


Figure 8: Normalized return versus environment steps on **Collaborative Cooking: Figure Eight**.

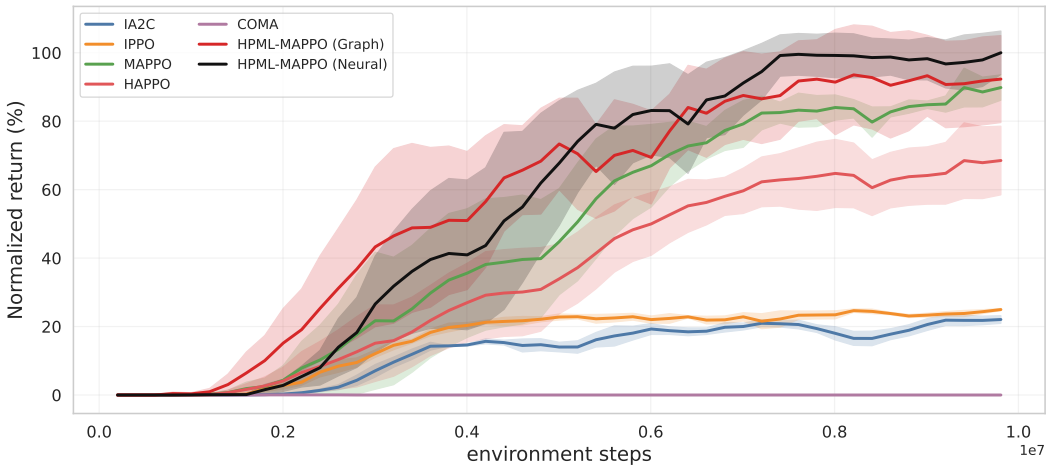


Figure 9: Normalized return versus environment steps on **Commons Harvest: Partnership**.

which corresponds to the Minty VI formulation. The present paper uses the Stampacchia gap because it aligns directly with the projected operator analysis developed later.

Remark B.4 (Gap and monotonicity). Lemma 2.3 is purely algebraic and does not require monotonicity. When F is monotone, however, the duality gap becomes a standard merit function that quantitatively certifies closeness to the VI solution set and supports classical convergence arguments. In general-sum games, the induced operator is typically non-monotone, which is precisely why cycling may persist and why a structural modification of the field is useful.

Example B.5 (Canonical bilinear cyclic-interaction field). Consider the bilinear min–max problem

$$\min_{u} \max_{v} \; u^{\top} v$$

with associated field

$$F(u, v) = (v, -u).$$

Its Jacobian is

$$J = \nabla F = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}, \qquad S = 0, \qquad A = J.$$

Hence the field is purely antisymmetric and has no potential component in the local Jacobian sense. This is the canonical example showing that simultaneous first-order updates may follow cyclic orbits around the origin rather than converge.

Table 3: Full Melting Pot results (mean normalized return % ± s.e.) on all 10 scenarios. Abbreviations: BoS-R = Bach or Stravinsky Repeated, PC-R = Pure Coordination Repeated, RC-R = Rationalizable Coordination Repeated, SH-R = Stag Hunt Repeated, PD-R = Prisoners Dilemma Repeated, CC-Cir = Collaborative Cooking Circuit, CC-Crowd = Collaborative Cooking Crowded, CC-F8 = Collaborative Cooking Figure Eight, CH-Part = Commons Harvest Partnership.

Method	BoS-R	PC-R	RC-R	SH-R	PD-R	CC-Cir	CC-Crowd	CC-F8	CH-Part	Clean Up	Macro Avg
IA2C	83.7 ± 7.5	83.7 ± 7.5	87.1 ± 7.8	78.1 ± 5.2	29.6 ± 29.6	88.3 ± 4.5	65.8 ± 5.9	84.6 ± 3.9	22.1 ± 1.1	70.6 ± 3.0	69.4 ± 3.4
IPPO	100.0 ± 4.1	7.4 ± 7.4	52.5 ± 5.6	65.7 ± 32.9	0.0 ± 0.0	30.7 ± 30.7	75.5 ± 4.8	39.0 ± 1.7	25.0 ± 0.3	84.4 ± 4.2	48.0 ± 4.7
MAPPO	86.3 ± 3.6	53.5 ± 2.5	65.0 ± 0.8	100.0 ± 4.1	61.6 ± 30.9	96.9 ± 4.9	42.1 ± 1.9	100.0 ± 4.1	89.8 ± 3.7	86.2 ± 3.6	78.1 ± 3.3
HAPPO	93.6 ± 3.9	88.7 ± 12.2	76.6 ± 10.3	97.0 ± 5.2	57.7 ± 29.9	82.7 ± 7.2	50.3 ± 11.1	97.5 ± 0.8	68.5 ± 10.0	81.1 ± 4.5	79.4 ± 3.9
COMA	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
HPML-MAPPO (Graph)	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0	98.3 ± 5.6	100.0 ± 0.0	100.0 ± 0.8	96.2 ± 3.5	92.3 ± 12.7	100.0 ± 25.9	98.7 ± 3.0
HPML-MAPPO (Neural)	88.7 ± 12.2	99.6 ± 0.2	99.3 ± 0.2	96.2 ± 3.5	100.0 ± 4.7	99.9 ± 0.0	99.5 ± 5.3	99.6 ± 0.2	100.0 ± 6.3	98.7 ± 0.2	98.2 ± 1.6

Remark B.6 (Local and global non-potentiality). Proposition 2.5 identifies antisymmetry of the Jacobian as a *local* obstruction to integrability. This is not yet a global statement: even if a field looks locally potential-like, global or sample-level obstructions may remain. The role of HPML in the main text is to move from this local Jacobian diagnostic to a global L^2 projection of the full joint field, thereby producing a residual that measures non-potentiality at the vector-field level.

C Variational and Geometric Details for HPML

This appendix records the continuous Euler–Lagrange characterization of the projection, its circulation interpretation, and several geometric remarks that are not needed in the main text but clarify the structure behind HPML.

Relation to the closest decomposition frameworks. Candogan et al. (Candogan et al., 2011) decompose a complete finite strategic-form game: nodes are pure-strategy profiles, edges are unilateral deviations, and the edge flow is formed from payoff differences; the output characterizes potential, harmonic, and nonstrategic components of the payoff game. Balduzzi et al. (Balduzzi et al., 2018) instead analyze the local symmetric–antisymmetric split of the Jacobian of a differentiable game's simultaneous-gradient field. HPML acts on a third object: the realized update field supplied by a learning algorithm over a sampled policy or parameter region. Its graph edges approximate line integrals of that field, and the output is an executable next-step direction. The accompanying VI-gap result evaluates the projected iterate against the original operator through an explicit residual term, while the inexact theorem separately accounts for projection error. Thus two algorithms on the same payoff game can share the same finite-game decomposition but have different HPML decompositions because they induce different learning fields.

Lemma C.1 (Weighted integration by parts)**.** *Let $\mathcal{X} \subset \mathbb{R}^d$ be a bounded Lipschitz domain or a flat torus. Let $d\mu(x) = \rho(x)\,dx$ with $0 < \rho_{\min} \le \rho(x) \le \rho_{\max} < \infty$ a.e. on $\mathcal{X}$. Assume boundary conditions that eliminate boundary terms (e.g., periodic boundary conditions on the torus; or $\rho U \cdot n = 0$ on $\partial\mathcal{X}$). Then for any $\psi \in H^1(\mathcal{X})$ and any vector field U for which $\div_\mu U$ is well-defined in $H^{-1}(\mathcal{X})$,*

$$\int_{\mathcal{X}} U^\top \nabla\psi \, d\mu = -\int_{\mathcal{X}} \psi \; \div_\mu U \, d\mu.$$

Equivalently, ∇ and $-\div_\mu$ are adjoint under the $L^2(\mu)$ pairing.

Proposition C.2 (Poisson-type characterization of the projection)**.** *Assume $M \succ 0$ is constant and $d\mu(x) = \rho(x)\,dx$ with $\rho > 0$. Assume boundary conditions that remove boundary terms in Lemma C.1. Let $\Phi^\star$ be any minimizer in Definition 3.1. Then the projected field is $\Pi^M_{\mathrm{grad}} F = \nabla_M \Phi^\star$ and $\Phi^\star$ satisfies the weak equation*

$$\mathcal{L}_{M,\mu}\Phi^\star = \div_\mu F, \qquad \textit{i.e.,} \qquad \div_\mu\big(M^{-1}\nabla\Phi^\star\big) = \div_\mu F.$$

Moreover, $\Phi^\star$ is unique up to an additive constant; one may fix a gauge such as $\mathbb{E}_\mu[\Phi^\star] = 0$. A natural boundary condition implied by the variational optimality is the zero-flux residual condition

$$\rho\,\big(M^{-1}\nabla\Phi^\star - F\big)\cdot n = 0 \quad \textit{on } \partial\mathcal{X},$$

whenever traces are well-defined.

Remark C.3 (Local Jacobian diagnostic and global projection). The Jacobian split in Section 2 is a local diagnostic: antisymmetry in ∇F signals local cyclic interaction behavior. The HPML projection is a global $L^2(M, \mu)$ construction on the full field. For a locally linear field $F(x) \approx Bx$, Euclidean metric $M = I$, and isotropic sampling, the closest gradient field corresponds to the symmetric part of B, namely $B \mapsto \frac{1}{2}(B + B^\top)$.

Remark C.4 (Relation to Hodge decomposition). Lemma 3.2 is the structural statement used in the main text. On domains or complexes with appropriate topology and boundary conditions, the residual can often be refined into curl-like and harmonic components. The discrete construction used by HPML mirrors this picture on graphs or sample complexes, but the main text only requires the projected gradient component and its orthogonal residual.

Convention-level cyclic interaction as circulation. Cyclic interaction dynamics need not come from explicit symmetry. In CTDE training, one often observes repeated revisits to a small family of recurring joint behaviors. Geometrically, such behavior corresponds to nontrivial circulation of the joint update field along approximately closed loops in the visited region of $\mathcal{X}$.

For a piecewise C^1 closed curve $\gamma : [0, 1] \to \mathcal{X}$ with $\gamma(0) = \gamma(1)$, define

$$\circ_F(\gamma) := \int_0^1 \langle F(\gamma(t)), \dot{\gamma}(t) \rangle_M \, dt = \int_0^1 \dot{\gamma}(t)^\top M F(\gamma(t)) \, dt.$$

Proposition C.5 (Closed-loop circulation is carried by the non-potential residual)**.** *Let $G^\star := \Pi^M_{\mathrm{grad}} F$ and $R := F - G^\star$. Assume $G^\star = \nabla_M \Phi^\star$ with $\Phi^\star$ sufficiently regular along the curve under consideration. For any piecewise C^1 closed curve γ in $\mathcal{X}$,*

$$\circ_{G^\star}(\gamma) = 0, \qquad \textit{and hence} \qquad \circ_F(\gamma) = \circ_R(\gamma).$$

Moreover, if μ_γ denotes the pushforward of the uniform measure on $[0, 1]$ by γ, then

$$|\circ_F(\gamma)| \le \Big(\mathbb{E}_{t\sim \mathrm{Unif}[0,1]} \|R(\gamma(t))\|_M^2\Big)^{1/2} \Big(\mathbb{E}_{t\sim \mathrm{Unif}[0,1]} \|\dot{\gamma}(t)\|_M^2\Big)^{1/2}.$$

Remark C.6 (Operational reading of failed collaboration loops). Proposition C.5 formalizes the intuition that recurring convention cycles correspond to nonzero circulation, which cannot be generated by any potential flow. Accordingly, the non-potentiality in Definition 3.4, or its empirical discrete proxy, can be read as a quantitative certificate of such looping behavior.

Remark C.7 (Metric and sampling choices). The projection is intentionally relative to (M, μ): M determines which actor-field distortions are costly, while μ determines the region in which approximation matters. In the matched experiments we use $M = I$ because all compared MAPPO-based methods share one fixed, comparably scaled parameterization, so Euclidean projection isolates integrability rather than adding a second preconditioner. For differently scaled actor blocks, block normalization is more appropriate; when distances should reflect changes in action distributions, a Fisher-type metric is a natural alternative. Recent-iterate sampling approximates the on-policy occupation measure, placing accuracy where the update is actually applied. Too short a window can make the graph poorly connected or lifting ill conditioned; too long a window can mix stale fields from incompatible training stages. Useful diagnostics are held-out projection error near the current iterate, graph connectivity, Laplacian conditioning, and local lifting conditioning. The theory is conditional on the declared (M, μ) and does not claim invariance to these choices. Smoothly varying $M(x)$ is also admissible, leading to a variable-coefficient weak Euler–Lagrange equation.

Remark C.8 (Discrete and amortized implementations). The practical implementations of HPML do not solve the continuous Poisson problem directly. Instead, they approximate the same variational objective either on a graph or sample complex, or through an amortized potential model. The concrete procedures appear in Algorithm 1 and Algorithm 3.

D Graph Construction and Discrete Recovery Details

This appendix records the implementation-level details of the sample-graph construction, the graph interpretation of the residual, the local lifting formulas, and the graph-based projection subroutine used by HPML.

Graph construction. Let $\{x^1, \dots, x^N\} \subset \mathcal{X}$ be sampled joint-policy points. We build an undirected k-NN graph $G = (\mathcal{V}, \mathcal{E})$ on the samples, where $\mathcal{V} = \{1, \dots, N\}$ and $(i, j) \in \mathcal{E}$ indicates that x^j is among the k nearest neighbors of x^i (symmetrized). If a metric $M \succ 0$ is used in Section 3, neighbors may be computed using

$$\mathrm{Dist}_M(x^i, x^j) := \sqrt{(x^i - x^j)^\top M (x^i - x^j)};$$

otherwise Euclidean distance is used. Each undirected edge is assigned a positive weight $w_{ij} = w_{ji} > 0$.

Remark D.1 (Cycle-space interpretation on a graph). On a pure graph, the space of edge flows decomposes into the cut space (gradients) and the cycle space (circulations). Thus ω_{cyc} is precisely the cycle-space component of ω. If one augments the graph to a higher-order simplicial complex, one can further refine the decomposition into gradient, curl, and harmonic components, but the graph-only variant already removes all non-gradient circulation supported by cycles.

Remark D.2 (Convention graphs and identifying which patterns form the cycle). When samples $\{x_i\}$ cluster around a small number of recurring joint behaviors, the k-NN graph can be viewed as a convention graph whose nodes represent these patterns and whose edges capture local transitions induced by the update field. In this view, the residual ω_{cyc} in Definition 4.1 is exactly the cycle-space component on the convention graph, and $\mathrm{NonPot}(\omega)$ in Definition 4.4 quantifies how much of the observed update energy is explained only by cycling among conventions.

Moreover, one can localize which conventions participate in the loop by expanding ω_{cyc} on a cycle basis of the graph, e.g. fundamental cycles from a spanning tree. Large coefficients identify the specific subset of conventions responsible for the recurrent dynamics.

D.1 Node-wise lifting formulas

The graph projection outputs a scalar potential $\phi^\star$. To obtain a vector direction at node i, define the neighborhood

$$\mathcal{N}(i) := \{j : (i, j) \in \mathcal{E}\}.$$

We fit $h^i \in \mathbb{R}^d$ as an approximation of $\nabla\Phi$ by solving

$$h^i \in \arg\min_{h \in \mathbb{R}^d} \sum_{j \in \mathcal{N}(i)} w_{ij} \Big(\langle h, x^j - x^i \rangle - (\phi^\star_j - \phi^\star_i) \Big)^2 + \lambda \|h\|_2^2, \tag{3}$$

where $\lambda \geq 0$ is a ridge parameter. We then set

$$g^i := M^{-1} h^i. \tag{4}$$

In matrix form,

$$h^i = (X_i^\top W_i X_i + \lambda I)^{-1} X_i^\top W_i y_i,$$

where rows of X_i are $(x^j - x^i)^\top$ and entries of y_i are $(\phi^\star_j - \phi^\star_i)$.

Remark D.3 (Updating a query point not in the sample set). If the current iterate x_t is not one of the sampled nodes, one may insert x_t into the graph by connecting it to its k nearest sampled neighbors, compute local potential differences, and apply the same lifting step to obtain g_t.

Proposition D.4 (Metric-consistent discretization of potential differences)**.** *Let $M \in \mathbb{R}^{d \times d}$ be constant SPD and let $\Phi \in C^2(\mathcal{X})$. For any $x \in \mathcal{X}$ and any displacement Δx with $x + \Delta x \in \mathcal{X}$,*

$$\Phi(x + \Delta x) - \Phi(x) = \langle \nabla\Phi(x), \Delta x \rangle + O(\|\Delta x\|^2) = \langle M \nabla_M \Phi(x), \Delta x \rangle + O(\|\Delta x\|^2).$$

Consequently, on a graph with node positions $\{x^i\}$ and edges (i, j), the metric-consistent edge 1-form associated with a vector field G intended to approximate $\nabla_M \Phi$ is

$$\omega_{ij} := \Big\langle M \, \frac{G(x^i) + G(x^j)}{2}, \, x^j - x^i \Big\rangle,$$

and if $G = \nabla_M \Phi$ then

$$\omega_{ij} = \Phi(x^j) - \Phi(x^i) + O(\|x^j - x^i\|^2).$$

D.2 Graph-based projection algorithm

Algorithm 1 Discrete Hodge Projection (Graph Poisson Solve)

Require: Samples $\{x^i\}_{i=1}^N \subset \mathcal{X}$, field estimates $\{F_i = F(x^i)\}$, metric $M \succ 0$ (optional), weighted k-NN graph $G = (\mathcal{V}, \mathcal{E})$, ridge $\lambda \geq 0$.

Ensure: Node potential $\phi^\star$, potential edge flow ω_{pot}, residual ω_{cyc}, and optionally lifted node directions $\{g^i\}$.

Choose an orientation for each undirected edge and form $\mathcal{E}^{\rightarrow}$.
Construct the incidence matrix B and the diagonal weight matrix $W = \mathrm{diag}(w_e)$.
for each oriented edge $e = (i \to j) \in \mathcal{E}^{\rightarrow}$ **do**
  $\Delta x_e \leftarrow x^j - x^i$
  $\omega_e \leftarrow \left\langle M \frac{1}{2}(F_i + F_j),\, \Delta x_e \right\rangle$
end for
Form $L \leftarrow B^\top W B$ and $b \leftarrow B^\top W \omega$.
Solve $L\phi^\star = b$ with a gauge condition.
$\omega_{\mathrm{pot}} \leftarrow B\phi^\star, \quad \omega_{\mathrm{cyc}} \leftarrow \omega - \omega_{\mathrm{pot}}$.
if node directions are needed **then**
  for $i = 1$ to N **do**
    Fit h^i by equation 3.
    Set $g^i \leftarrow M^{-1} h^i$.
  end for
end if

Algorithm 2 Hodge-Projected Multi-Agent Learning (HPML)

Require: Initial $x_0 \in \mathcal{X}$, stepsizes $\{\eta_t\}$, update-field oracle/estimator $F(\cdot)$, metric $M \succ 0$ (optional), feasibility operator $\mathrm{Retr}(\cdot)$, projection routine (discrete graph or amortized neural projection).

for $t = 0, 1, 2, \ldots$ **do**
  Estimate $F(x_t)$.
  Obtain a projected direction $g_t \approx \Pi^M_{\mathrm{grad}} F(x_t)$:
    either (i) build/update a sample graph and run Algorithm 1 then lift to get g_t, or
    (ii) use an amortized potential network to compute g_t by autograd.
  $x_{t+1} \leftarrow \mathrm{Retr}(x_t + \eta_t g_t)$.
end for

E Algorithmic Procedures for HPML

This appendix records the generic outer-loop realization of HPML, the amortized neural realization, and the CTDE plug-in realization. The graph-based projection subroutine itself is given in Algorithm 1.

Remark E.1 (Practical schedules for amortized projection). The amortized projection can be updated either online (a few inner SGD steps per outer iteration) or periodically (larger refreshes every H outer steps). If M varies with x, one may use a diagonal or low-rank approximation of $M(x)$, or provide metric-related features to the potential network. These choices affect approximation quality and computational overhead, but do not change the projection objective itself.

Remark E.2 (Parameter-space realization under CTDE). The main theory is stated on a convex feasible set of policy coordinates, whereas the deep implementation projects the stacked actor field in one fixed parameterization. This construction is well defined in those coordinates but is not claimed to be invariant to neural reparameterization; Euclidean distance in parameter space need not equal policy-space distance. The metric M makes this dependence explicit and can be replaced by block-normalized or policy-aware geometry when appropriate. In the matched MAPPO comparison, the baseline and both HPML variants use exactly the same actor parameterization, so the empirical difference isolates the projection within that coordinate system. The inexact-direction bound then separates ideal projected dynamics from finite graph or neural approximation error.

Remark E.3 (Relation to Bellman residuals and the multi-agent source of circulation). The role of Proposition 5.1 is interpretive rather than evaluative. It does *not* say that single-agent Bellman or TD operators are ill-defined. The issue appears only after stacking several critic-driven local actor-improvement directions into one joint field. Each block g_i may be a valid local greedy direction, yet

Algorithm 3 Neural HPML (Amortized Potential Projection)

Require: Initial $x_0 \in \mathcal{X}$, stepsizes $\{\eta_t\}$, feasibility operator $\mathrm{Retr}(\cdot)$, field estimator $\widehat{F}(\cdot)$, buffer $\mathcal{B}$, inner steps K, batch size B, (optional) metric $M \succ 0$.

Initialize potential-network parameters θ.
for $t = 0, 1, 2, \ldots$ **do**
  Append x_t (and auxiliary rollout data, if any) to buffer $\mathcal{B}$.
  Sample $\{x_k\}_{k=1}^{B} \sim \mathcal{B}$ and compute $\{\widehat{F}(x_k)\}$.
  for $k = 1$ to K **do**
    Update θ by SGD on $\widehat{\mathcal{L}}(\theta)$ in equation 2.
  end for
  Compute $g_t \leftarrow M^{-1}\nabla_x \Phi_\theta(x_t)$ by autograd.
  $x_{t+1} \leftarrow \mathrm{Retr}(x_t + \eta_t g_t)$.
end for

Algorithm 4 CTDE-HPML (HPML as a Plug-in Projection Layer for MARL)

Require: Decentralized actors $\{\pi_{\theta_i}(a_i \mid o_i)\}_{i=1}^{n}$, centralized critic V_ψ (or Q_ψ), stepsizes $\{\eta_t\}$, projection routine $\Pi_{\mathrm{grad}}^{M}(\cdot)$ (discrete or neural), buffer $\mathcal{B}$, and an optional feasibility operator $\mathrm{Retr}(\cdot)$.

for $t = 0, 1, 2, \ldots$ **do**
  Roll out trajectories with decentralized execution and append transitions to $\mathcal{B}$.
  Sample a minibatch from $\mathcal{B}$ and update the centralized critic parameters ψ.
  Compute a centralized advantage (or TD-based signal) $\widehat{A}(s, a)$.
  For each agent i, form a stochastic ascent direction

$$\widehat{g}_i \leftarrow \mathbb{E}\big[\nabla_{\theta_i} \log \pi_{\theta_i}(a_i \mid o_i)\, \widehat{A}(s, a)\big].$$

  Stack the joint field $\widehat{F} \leftarrow (\widehat{g}_1, \ldots, \widehat{g}_n)$.
  Compute projected direction $g_t \leftarrow \Pi_{\mathrm{grad}}^{M} \widehat{F}$.
  Update each actor:

$$\theta_i \leftarrow \mathrm{Retr}(\theta_i + \eta_t [g_t]_i), \qquad i = 1, \ldots, n.$$

end for

the collection $(g_1, \ldots, g_n)$ may fail to be jointly integrable because the cross-agent response terms need not be symmetric. That is the sense in which there is a "mismatch": changing agent j can alter agent i's improvement direction without a matching reciprocity in the reverse block.

This phenomenon is natural in multi-agent CTDE because the joint control problem is often factorized into simultaneously updated local blocks for scalability. The decisive distinction is again exact centralized optimization versus centralized evaluation alone. A centralized critic and even a single shared reward do not by themselves remove the obstruction if the implemented actor update is still assembled from local or approximate blocks. The obstruction disappears only when all agents are optimized as one super-agent by the exact gradient of a single centralized scalar objective in the full joint parameter space. Sampling other agents' actions then contributes stochasticity, but not a structural non-potential cyclic component at the level of the exact expected field. Partial observability and approximation error can exacerbate the effect but are not necessary. Mathematically, non-potentiality is not exclusive to multi-agent systems, but multi-agent blockwise updates are a particularly natural mechanism generating it.

So the residual is not a Bellman-evaluation error. It is a structural inconsistency in the joint greedy-improvement dynamics induced by decentralized multi-agent updates.

E.1 CTDE INTEGRABILITY AND RESIDUAL-ONLY CYCLIC DYNAMICS

Remark E.4 (Origin of the non-potential cyclic component). The nonzero loop integral in Proposition 5.1 is generated by a local asymmetry in the Jacobian of the one-form MF. Writing

$Y(x) := MF(x), \qquad K(x) := \nabla Y(x) = M\nabla F(x)$, joint integrability requires $K(x) = K(x)^\top$. Equivalently, the antisymmetric part $B(x) := \frac{1}{2}\big(M\nabla F(x) - \nabla F(x)^\top M\big)$ must vanish. If $B(x) \neq 0$, then for any sufficiently small parallelogram $C_\varepsilon(x; u, v) \subset U$, $\oint_{C_\varepsilon(x;u,v)} \langle MF(z), dz\rangle = 2\varepsilon^2\langle B(x)u, v\rangle + o(\varepsilon^2)$, so the field has a local circulation component.

In the two-agent Euclidean case $M = I$ with $x = (x_1, x_2)$ and $F = (g_1, g_2)$, this becomes $\oint_{C_\varepsilon} \langle F(z), dz\rangle = \varepsilon^2 h^\top\big(\nabla_{x_2} g_1(x) - \nabla_{x_1} g_2(x)^\top\big)k + o(\varepsilon^2)$. Thus the relevant "mismatch" is not merely that agents prefer different actions. It is the failure of cross-agent response symmetry: changing agent 2 alters agent 1's greedy improvement direction in a way that is not matched by the reverse perturbation. Once this occurs, the stacked field cannot be the gradient of a single common potential.

This mismatch is typically created by strategic coupling in the environment together with decentralized or blockwise actor updates. Importantly, the issue is not determined by reward sharing alone. A single shared team reward and a centralized critic remove the non-potential cyclic component only when the realized joint actor-update field is exactly the gradient of the centralized team objective. If the same MARL system is trained through separate local or approximate improvement blocks, then one agent's update is still formed against the changing behavior of the others, and the induced cross-agent response terms need not be symmetric. Shared rewards, coupled transitions, role-allocation constraints, convention selection, or mixed cooperative/competitive effects can all make one agent's local improvement direction reshape another agent's landscape. Partial observability and function-approximation error can strengthen the asymmetry, but they are not required; the effect can already appear in fully observed matrix or Markov games. Hyperparameters that change the update operator itself can also change the size of the residual. By contrast, the numerical step size and the initial condition do not create the residual field; they determine whether training visits a region where the asymmetry is strong and whether the resulting cyclic motion is amplified in discrete time.

This also clarifies the multi-agent aspect. The decisive distinction is exact centralized optimization, not merely centralized evaluation. If all agents were merged into a single super-agent and updated by exact gradient ascent on one centralized scalar return, then the resulting field would be integrable and this non-potential cyclic component would disappear. By contrast, MARL with a single shared reward can still exhibit the same obstruction whenever the implemented actor update is factorized into local or approximate blocks whose cross-agent responses are not jointly consistent. Sampling the others' actions contributes stochasticity, but by itself does not imply a structural non-potential cyclic component at the level of the exact expected field; the obstruction appears when the realized update rule departs from exact joint-gradient structure. Mathematically, however, the phenomenon is not exclusive to multi-agent systems; any non-potential update rule may contain such a component. Finally, Proposition C.5 implies that any nonzero loop circulation yields nonzero residual energy, so the obstruction is not only geometric but also quantitatively detectable.

The next proposition isolates the effect of this residual cyclic component by considering the canonical residual-only skew field from Example B.5.

Proposition E.5 (Updating along the residual cyclic component alone need not approach the fixed point)**.** *Consider the residual-only field on $\mathbb{R}^2$, $R(u, v) = (v, -u)$, whose unique fixed point is $(0, 0)$. (i) Under the continuous-time dynamics $(\dot{u}, \dot{v}) = R(u, v)$, every non-stationary trajectory satisfies $u(t)^2 + v(t)^2 = u(0)^2 + v(0)^2$ for all $t \geq 0$ and therefore never reaches the fixed point. In particular, every circle $\gamma_r(s) = r(\cos 2\pi s, \sin 2\pi s), \qquad r > 0$, is an invariant orbit. (ii) Under simultaneous explicit-Euler updates $z_{t+1} = z_t + \eta R(z_t), \qquad z_t = (u_t, v_t)$, one has $\|z_{t+1}\|^2 = (1 + \eta^2)\|z_t\|^2$ for every step size $\eta > 0$. Hence every nonzero iterate moves strictly away from the fixed point, and $\|z_t\| = (1 + \eta^2)^{t/2}\|z_0\|$. Thus a pure residual cyclic component can make the fixed point unreachable in continuous time and repelling in discrete time.*

Proposition E.5 isolates the role of the residual cyclic component: in continuous time it sustains recurrent motion, while in simultaneous discrete-time updates any positive step size turns the same skew field into outward drift.

F AUXILIARY RESULTS FOR PROJECTED DYNAMICS AND GAP BOUNDS

This appendix records the projection inequality used by Theorem 6.4, the L^2 output-law gap bound, an inexact projected-step lemma used by Theorem 6.8, and several interpretive remarks omitted from the main text.

Lemma F.1 (Projection variational inequality and nonexpansiveness)**.** *Let $\mathcal{X} \subset \mathbb{R}^d$ be nonempty, closed, and convex, and let $P := \mathrm{Proj}_{\mathcal{X}}$ be the Euclidean projection. Then for any $y \in \mathbb{R}^d$ and any $z \in \mathcal{X}$,*

$$\langle y - P(y),\, z - P(y)\rangle \le 0.$$

Moreover, P is firmly nonexpansive: for any $y, y' \in \mathbb{R}^d$,

$$\|P(y) - P(y')\|^2 \le \langle P(y) - P(y'),\, y - y'\rangle, \qquad \textit{and hence} \qquad \|P(y) - P(y')\| \le \|y - y'\|.$$

Remark F.2 (Residual circulation and the additive gap term). By Proposition C.5, any closed-loop circulation of the original operator F must be carried by the residual R. Accordingly, the additive terms $D\varepsilon$ in Theorem 6.7 and DE_T in Theorem F.3 can be read as the unavoidable contribution of non-potential circulation in the original joint field.

Theorem F.3 (L^2 non-potentiality with respect to the output-iterate law)**.** *Assume Assumption 6.1 and suppose*

$$F(x) = -\nabla\Phi(x) + R(x),$$

with R square-integrable under the law of the output iterate defined below. Let $\eta \le 1/L$ and run the exact projected ascent

$$x_{t+1} = \mathrm{Proj}_{\mathcal{X}}(x_t + \eta\nabla\Phi(x_t)), \qquad t = 0, \dots, T-1.$$

Let $\tilde{t}$ be uniformly distributed over $\{0, 1, \dots, T-1\}$, independent of the iterates, and output $\tilde{x}_T := x_{\tilde{t}}$. Define

$$E_T := \Big(\mathbb{E}\big[\|R(\tilde{x}_T)\|^2\big]\Big)^{1/2}.$$

Then

$$\mathbb{E}\big[\mathrm{Gap}(\tilde{x}_T)\big] \le (D + \eta G)\sqrt{\frac{2(\Phi_{\max} - \Phi_{\min})}{T\eta}} + DE_T.$$

Remark F.4 (Rate extensions under stronger structure). The $O(1/\sqrt{T})$ rate in Theorem 6.7 is the most robust guarantee under the minimal assumptions used in the main text. Under stronger structure, such as a PL condition for Φ or stronger operator regularity, one can sharpen the stationarity rate and therefore the induced gap rate. These refinements are not needed for the main narrative.

Lemma F.5 (Gap bound via inexact projected-step mapping)**.** *Under Assumption 6.1, let*

$$x^+ = \mathrm{Proj}_{\mathcal{X}}(x + \eta g), \qquad \widetilde{\mathcal{G}}_\eta(x) := \frac{1}{\eta}(x^+ - x),$$

and let $e := g - \nabla\Phi(x)$. Then

$$\mathrm{Gap}_\Phi(x) \le \big(D + \eta\|g\|\big)\,\|\widetilde{\mathcal{G}}_\eta(x)\| + D\,\|e\|.$$

Remark F.6 (Interpretation of the inexact-projection bound). In Theorem 6.8, the residual term $D\varepsilon$ is inherited from the original non-potential component, while the additional $D\bar{\delta}$ term and the $\bar{\delta}^2$ floor arise from projection error under the stated pointwise accuracy assumption. Thus the theorem separates two effects: structural non-potentiality in the original game field and approximation error in the projection routine itself.

G BROADER IMPACTS.

HPML is a general-purpose MARL optimization tool. Its potential positive impact is improved stability and reproducibility in cooperative and mixed-motive multi-agent systems. A potential negative impact is that more stable multi-agent optimization could also improve systems deployed in competitive, surveillance, or resource-allocation settings; responsible use therefore depends on the application context, evaluation protocol, and deployment safeguards.

H PROOF

H.1 PROOF OF LEMMA 2.3

Proof. Fix any $\bar{x} \in \mathcal{X}$. By Definition 2.2,

$$\mathrm{Gap}(\bar{x}) = \max_{x \in \mathcal{X}} \langle F(\bar{x}), \bar{x} - x \rangle .$$

First observe that choosing the feasible point $x = \bar{x}$ yields

$$\langle F(\bar{x}), \bar{x} - \bar{x} \rangle = \langle F(\bar{x}), 0 \rangle = 0,$$

hence the maximum is at least this value and therefore

$$\mathrm{Gap}(\bar{x}) \ \geq \ 0.$$

Now assume $\mathrm{Gap}(\bar{x}) = 0$. For an arbitrary $x \in \mathcal{X}$, since $\mathrm{Gap}(\bar{x})$ is the maximum of the map $x \mapsto \langle F(\bar{x}), \bar{x} - x \rangle$ over the feasible set $\mathcal{X}$, it dominates every feasible value, i.e.,

$$\langle F(\bar{x}), \bar{x} - x \rangle \ \leq \ \max_{y \in \mathcal{X}} \langle F(\bar{x}), \bar{x} - y \rangle \ = \ \mathrm{Gap}(\bar{x}) \ = \ 0.$$

Multiplying both sides by -1 reverses the inequality and gives

$$- \langle F(\bar{x}), \bar{x} - x \rangle \ \geq \ 0.$$

Using bilinearity of the inner product and the identity $-(\bar{x} - x) = x - \bar{x}$, we rewrite the left-hand side as

$$- \langle F(\bar{x}), \bar{x} - x \rangle = \langle F(\bar{x}), -(\bar{x} - x) \rangle = \langle F(\bar{x}), x - \bar{x} \rangle .$$

Therefore, for every $x \in \mathcal{X}$ we have

$$\langle F(\bar{x}), x - \bar{x} \rangle \ \geq \ 0,$$

which is exactly the Stampacchia variational inequality condition in Definition 2.1. Hence $\bar{x}$ solves $\mathrm{VI}(\mathcal{X}, F)$. □

H.2 PROOF OF PROPOSITION 2.5

Proof. For (i), assume there exists ϕ that is twice continuously differentiable on an open set $V \supseteq \mathcal{X}$ and satisfies $F = \nabla \phi$ on V. Writing components, for each $i \in \{1, \ldots, d\}$ we have $F_i(x) = \partial_i \phi(x)$ for all $x \in V$. By definition of the Jacobian $J(x) = \nabla F(x)$, its entries are

$$J_{ij}(x) = \partial_j F_i(x) = \partial_j \big(\partial_i \phi(x)\big) = \partial_j \partial_i \phi(x).$$

Similarly,

$$J_{ji}(x) = \partial_i F_j(x) = \partial_i \big(\partial_j \phi(x)\big) = \partial_i \partial_j \phi(x).$$

Since $\phi \in C^2(V)$, Schwarz/Clairaut's theorem applies and yields $\partial_j \partial_i \phi(x) = \partial_i \partial_j \phi(x)$ for all $x \in V$, hence $J_{ij}(x) = J_{ji}(x)$ for every pair (i, j), so $J(x) = J(x)^\top$ on V. With the symmetric–antisymmetric split $A(x) = \frac{1}{2}\big(J(x) - J(x)^\top\big)$ from Section 2.2, this gives $A(x) \equiv 0$ on V, and consequently on $\mathcal{X}$ we have

$$\mathcal{R}(x) = \|A(x)\|_F^2 = 0.$$

For (ii), assume $U \subset \mathbb{R}^d$ is open and simply connected, $F \in C^1(U)$, and $A(x) \equiv 0$ on U. Then for every $x \in U$,

$$0 = A(x) = \tfrac{1}{2}\big(J(x) - J(x)^\top\big) \qquad \Longrightarrow \qquad J(x) = J(x)^\top.$$

Writing this entrywise and recalling $J_{ij}(x) = \partial_j F_i(x)$, we obtain the cross-partial equalities

$$\partial_j F_i(x) = \partial_i F_j(x), \qquad \forall\, x \in U,\ \forall\, i, j \in \{1, \ldots, d\}.$$

Lemma H.1 (Closed-loop integral vanishes under $J = J^\top$)**.** *Let $U \subset \mathbb{R}^d$ be open and simply connected and let $F \in C^1(U)$ satisfy $\partial_j F_i = \partial_i F_j$ on U. Then for every piecewise C^1 closed curve $\gamma : [0, 1] \to U$ with $\gamma(0) = \gamma(1)$,*

$$\int_0^1 \langle F(\gamma(t)), \gamma'(t) \rangle \ dt \ = \ 0.$$

Proof. Fix a piecewise C^1 closed curve γ in U. Since U is simply connected, γ is homotopic to the constant loop at $\gamma(0)$ through loops in U; concretely, there exists a (piecewise) C^1 map $H : [0,1] \times [0,1] \to U$ such that $H(0,t) = \gamma(t)$ for all t, $H(1,t) = \gamma(0)$ for all t, and $H(s,0) = H(s,1) = \gamma(0)$ for all s. Define, for each $s \in [0,1]$, the loop integral

$$I(s) := \int_0^1 \langle F(H(s,t)), \partial_t H(s,t) \rangle \, dt.$$

We compute $I'(s)$ by differentiating under the integral sign. Using the chain rule and that $\nabla F = J$, we have

$$\frac{d}{ds} \langle F(H(s,t)), \partial_t H(s,t) \rangle = \langle J(H(s,t))\, \partial_s H(s,t), \partial_t H(s,t) \rangle + \langle F(H(s,t)), \partial_{st} H(s,t) \rangle .$$

Integrating this identity over $t \in [0,1]$ gives

$$I'(s) = \int_0^1 \langle J(H)\, \partial_s H, \partial_t H \rangle \, dt + \int_0^1 \langle F(H), \partial_{st} H \rangle \, dt,$$

where for readability $H = H(s,t)$ inside the integrals. For the second term, integrate by parts in t. Since $F \in C^1$ and H is (piecewise) C^1 in both variables, the map $t \mapsto \langle F(H(s,t)), \partial_s H(s,t) \rangle$ is (piecewise) C^1 and

$$\partial_t \langle F(H), \partial_s H \rangle = \langle J(H)\, \partial_t H, \partial_s H \rangle + \langle F(H), \partial_{ts} H \rangle .$$

Rearranging yields

$$\langle F(H), \partial_{st} H \rangle = \partial_t \langle F(H), \partial_s H \rangle - \langle J(H)\, \partial_t H, \partial_s H \rangle .$$

Integrating from $t = 0$ to $t = 1$ gives

$$\int_0^1 \langle F(H), \partial_{st} H \rangle \, dt = \Big[\langle F(H(s,t)), \partial_s H(s,t) \rangle \Big]_{t=0}^{t=1} - \int_0^1 \langle J(H)\, \partial_t H, \partial_s H \rangle \, dt.$$

Because $H(s,0) = H(s,1) = \gamma(0)$ for all s, the boundary term vanishes: both endpoints coincide, and $\partial_s H(s,0) = \partial_s H(s,1) = 0$ for the fixed endpoint loop, so $\big[\langle F(H), \partial_s H \rangle \big]_0^1 = 0$. Substituting back into the formula for $I'(s)$ yields

$$I'(s) = \int_0^1 \langle J(H)\, \partial_s H, \partial_t H \rangle \, dt - \int_0^1 \langle J(H)\, \partial_t H, \partial_s H \rangle \, dt.$$

Using $\langle a, b \rangle = \langle b, a \rangle$ and the identity $\langle J\, u, v \rangle - \langle J\, v, u \rangle = \big\langle (J - J^\top) u, v \big\rangle$, we obtain

$$I'(s) = \int_0^1 \big\langle (J(H) - J(H)^\top)\, \partial_s H, \partial_t H \big\rangle \, dt.$$

Under the hypothesis $\partial_j F_i = \partial_i F_j$, we have $J = J^\top$ everywhere on U, hence $J(H) - J(H)^\top = 0$ along the homotopy and therefore $I'(s) = 0$ for all s. Thus $I(s)$ is constant in s. Evaluating at $s = 1$ gives $I(1) = 0$ because $H(1, \cdot)$ is the constant loop, so $I(0) = I(1) = 0$. Since $I(0)$ is exactly the integral along γ, the claim follows. □

Fix a base point $x_0 \in U$. For any $x \in U$, choose a piecewise C^1 curve $\gamma_{x_0 \to x} : [0,1] \to U$ with $\gamma_{x_0 \to x}(0) = x_0$ and $\gamma_{x_0 \to x}(1) = x$, and define

$$\phi(x) := \int_0^1 \big\langle F(\gamma_{x_0 \to x}(t)), \gamma'_{x_0 \to x}(t) \big\rangle \, dt.$$

We show ϕ is well-defined (independent of the chosen path). If γ_1 and γ_2 are two such paths from x_0 to x, consider the closed curve obtained by traversing γ_1 and then traversing γ_2 backwards; denote this closed loop by $\gamma_1 \circ \gamma_2^{-1}$. The integral along γ_2^{-1} is the negative of the integral along γ_2, so

$$\int_{\gamma_1 \circ \gamma_2^{-1}} \langle F, d\ell \rangle = \int_{\gamma_1} \langle F, d\ell \rangle - \int_{\gamma_2} \langle F, d\ell \rangle .$$

By Lemma H.1, the left-hand side is 0, hence $\int_{\gamma_1} \langle F, d\ell\rangle = \int_{\gamma_2} \langle F, d\ell\rangle$, proving that $\phi(x)$ does not depend on the choice of $\gamma_{x_0\to x}$.

It remains to show $\nabla\phi(x) = F(x)$. Fix $x \in U$ and a coordinate direction e_k. Because U is open, there exists $\delta > 0$ such that $x+he_k \in U$ for all $|h| < \delta$. Using path-independence, we may compute $\phi(x + he_k) - \phi(x)$ by choosing for $\phi(x + he_k)$ the path $\gamma_{x_0\to x}$ followed by the straight segment from x to $x + he_k$. Parameterizing that segment by $\sigma_h(t) = x + te_k$ for $t \in [0, h]$, we obtain

$$\phi(x + he_k) - \phi(x) = \int_0^h \langle F(\sigma_h(t)), \sigma_h'(t)\rangle \, dt = \int_0^h \langle F(x + te_k), e_k\rangle \, dt = \int_0^h F_k(x + te_k)\, dt.$$

Dividing by $h \neq 0$ gives

$$\frac{\phi(x + he_k) - \phi(x)}{h} = \frac{1}{h}\int_0^h F_k(x + te_k)\, dt.$$

Since F is continuous (indeed C^1) on U, the integrand $F_k(x+te_k)$ converges to $F_k(x)$ as $t \to 0$, and therefore the right-hand side converges to $F_k(x)$ as $h \to 0$ (it is the average value of a continuous function over an interval shrinking to a point). Hence the partial derivative exists and satisfies

$$\partial_k\phi(x) = \lim_{h\to 0} \frac{\phi(x + he_k) - \phi(x)}{h} = F_k(x).$$

Because this holds for every k, we conclude $\nabla\phi(x) = F(x)$ for all $x \in U$.

For uniqueness up to an additive constant, suppose $\tilde\phi$ is another scalar potential on U with $\nabla\tilde\phi = F$. Then $\nabla(\phi - \tilde\phi) = 0$ on U. For any $x, y \in U$, choose a piecewise C^1 path γ from x to y; applying the fundamental theorem of calculus along curves yields

$$(\phi - \tilde\phi)(y) - (\phi - \tilde\phi)(x) = \int_0^1 \Big\langle \nabla(\phi - \tilde\phi)(\gamma(t)), \gamma'(t)\Big\rangle \, dt = \int_0^1 \langle 0, \gamma'(t)\rangle \, dt = 0,$$

so $\phi - \tilde\phi$ is constant on the (path-)connected set U. This proves uniqueness up to an additive constant. □

H.3 Proof of Lemma C.1

Proof. We write $d\mu(x) = \rho(x)\, dx$ and interpret all expressions in the natural sense in which they are assumed to be well-defined; in particular, $\psi \in H^1(\mathcal{X})$ has $\nabla\psi \in L^2(\mathcal{X})$, the left-hand side $\int_{\mathcal{X}} U^\top \nabla\psi \, d\mu$ is understood as $\int_{\mathcal{X}} \rho\, U \cdot \nabla\psi \, dx$, and the right-hand side $\int_{\mathcal{X}} \psi \div_\mu U \, d\mu$ is understood as the H^{-1}–H^1 dual pairing whenever $\div_\mu U \notin L^2(\mu)$.

Recall the weighted divergence is given (in the distributional sense) by

$$\div_\mu U \;:=\; \frac{1}{\rho} \div (\rho U), \qquad \text{equivalently} \qquad \div(\rho U) = \rho \div_\mu U.$$

We first justify the identity for smooth ψ; the extension to $\psi \in H^1(\mathcal{X})$ will then follow by density. Fix $\psi \in C^\infty(\mathcal{X})$ (periodic if $\mathcal{X}$ is a torus, or smooth up to the boundary if $\mathcal{X}$ is a Lipschitz domain), and set

$$W(x) := \rho(x)\, U(x).$$

Then the left-hand side becomes

$$\int_{\mathcal{X}} U^\top \nabla\psi \, d\mu = \int_{\mathcal{X}} \rho\, U \cdot \nabla\psi \, dx = \int_{\mathcal{X}} W \cdot \nabla\psi \, dx.$$

We use the pointwise product rule for the divergence of a scalar times a vector field: for every x where the derivatives make sense,

$$\div(\psi W)(x) = \sum_{i=1}^d \partial_i\big(\psi(x)\, W_i(x)\big) = \sum_{i=1}^d \big(\partial_i\psi(x)\big)\, W_i(x) + \sum_{i=1}^d \psi(x)\, \partial_i W_i(x) = W(x)\cdot\nabla\psi(x) + \psi(x)\div W(x).$$

Rearranging this identity gives, pointwise,

$$W \cdot \nabla\psi = \div(\psi W) - \psi \div W.$$

Integrating over $\mathcal{X}$ with respect to dx yields

$$\int_{\mathcal{X}} W \cdot \nabla\psi \, dx = \int_{\mathcal{X}} \div(\psi W) \, dx - \int_{\mathcal{X}} \psi \; \div W \, dx.$$

If $\mathcal{X}$ is a bounded Lipschitz domain, the divergence theorem applies to the vector field ψW and gives

$$\int_{\mathcal{X}} \div(\psi W) \, dx = \int_{\partial\mathcal{X}} \psi \, W \cdot n \, dS = \int_{\partial\mathcal{X}} \psi \, \rho \, U \cdot n \, dS.$$

Under the stated boundary condition eliminating boundary terms (e.g. $\rho U \cdot n = 0$ on $\partial\mathcal{X}$), the boundary integral is 0, hence

$$\int_{\mathcal{X}} W \cdot \nabla\psi \, dx = -\int_{\mathcal{X}} \psi \; \div W \, dx.$$

If $\mathcal{X}$ is a flat torus, it has no boundary; equivalently, if one represents it by a fundamental periodic cell, the boundary contributions cancel pairwise due to periodicity, and one again has $\int_{\mathcal{X}} \div(\psi W) \, dx = 0$, so the same identity holds:

$$\int_{\mathcal{X}} W \cdot \nabla\psi \, dx = -\int_{\mathcal{X}} \psi \; \div W \, dx.$$

Substituting back $W = \rho U$ gives

$$\int_{\mathcal{X}} \rho \, U \cdot \nabla\psi \, dx = -\int_{\mathcal{X}} \psi \; \div (\rho U) \, dx.$$

Using $\div(\rho U) = \rho \; \div_\mu U$, we rewrite the right-hand side as

$$-\int_{\mathcal{X}} \psi \; \div (\rho U) \, dx = -\int_{\mathcal{X}} \psi \, \rho \; \div_\mu U \, dx = -\int_{\mathcal{X}} \psi \; \div_\mu U \, d\mu,$$

and therefore, for smooth ψ,

$$\int_{\mathcal{X}} U^\top \nabla\psi \, d\mu = -\int_{\mathcal{X}} \psi \; \div_\mu U \, d\mu.$$

To pass from smooth ψ to $\psi \in H^1(\mathcal{X})$, take a sequence $\{\psi_k\}$ of smooth functions (periodic on the torus, or smooth up to the boundary on a Lipschitz domain) such that $\psi_k \to \psi$ in $H^1(\mathcal{X})$. For each k the identity already proved yields

$$\int_{\mathcal{X}} \rho \, U \cdot \nabla\psi_k \, dx = -\int_{\mathcal{X}} \psi_k \, \rho \; \div_\mu U \, dx.$$

Because ρ is essentially bounded above and below, $\nabla\psi_k \to \nabla\psi$ in $L^2(\mathcal{X})$ implies $\nabla\psi_k \to \nabla\psi$ in $L^2(\mu)$, and the left-hand side converges by Cauchy–Schwarz:

$$\Big| \int_{\mathcal{X}} U \cdot (\nabla\psi_k - \nabla\psi) \, d\mu \Big| \le \|U\|_{L^2(\mu)} \, \|\nabla\psi_k - \nabla\psi\|_{L^2(\mu)} \longrightarrow 0.$$

On the right-hand side, the assumption that $\div_\mu U$ is well-defined in $H^{-1}(\mathcal{X})$ means the map $\psi \mapsto \langle \div_\mu U, \psi \rangle_{H^{-1}, H^1}$ is continuous with respect to the H^1 norm, so

$$\langle \div_\mu U, \psi_k \rangle_{H^{-1}, H^1} \longrightarrow \langle \div_\mu U, \psi \rangle_{H^{-1}, H^1}.$$

Interpreting $-\int_{\mathcal{X}} \psi_k \; \div_\mu U \, d\mu$ as this dual pairing when needed, we may pass to the limit and obtain

$$\int_{\mathcal{X}} U^\top \nabla\psi \, d\mu = -\int_{\mathcal{X}} \psi \; \div_\mu U \, d\mu \qquad \text{for all } \psi \in H^1(\mathcal{X}).$$

Reading the identity as an $L^2(\mu)$ (or H^{-1}–H^1) pairing shows precisely that ∇ and $-\div_\mu$ are adjoint under the $L^2(\mu)$ pairing. □

H.4 Proof of Proposition C.2

Proof. Write $d\mu(x) = \rho(x)\,dx$ and recall $M \succ 0$ is constant. Using $\nabla_M \Phi = M^{-1}\nabla\Phi$ and $\|v\|_M^2 = v^\top M v$, the objective in Definition 3.1 can be written as the quadratic functional

$$J(\Phi) := \int_{\mathcal{X}} \|F(x) - \nabla_M \Phi(x)\|_M^2 \, d\mu(x) = \int_{\mathcal{X}} \big(F(x) - M^{-1}\nabla\Phi(x)\big)^\top M \big(F(x) - M^{-1}\nabla\Phi(x)\big)\, d\mu(x).$$

Since M is constant, expanding the integrand gives

$$\big(F - M^{-1}\nabla\Phi\big)^\top M \big(F - M^{-1}\nabla\Phi\big) = F^\top M F - 2\,F^\top \nabla\Phi + \nabla\Phi^\top M^{-1}\nabla\Phi,$$

so $J(\Phi)$ depends on Φ only through $\nabla\Phi$ and is convex (indeed quadratic) in $\nabla\Phi$.

Lemma H.2 (First variation of J)**.** *For any $\Phi \in H^1(\mathcal{X})$ and any direction $\psi \in H^1(\mathcal{X})$, the map $\varepsilon \mapsto J(\Phi + \varepsilon\psi)$ is differentiable and*

$$\frac{d}{d\varepsilon} J(\Phi + \varepsilon\psi)\Big|_{\varepsilon=0} = 2 \int_{\mathcal{X}} \big(M^{-1}\nabla\Phi(x) - F(x)\big)^\top \nabla\psi(x)\, d\mu(x).$$

Proof. Set $\Phi_\varepsilon := \Phi + \varepsilon\psi$, so $\nabla\Phi_\varepsilon = \nabla\Phi + \varepsilon\nabla\psi$ and

$$F - M^{-1}\nabla\Phi_\varepsilon = \big(F - M^{-1}\nabla\Phi\big) - \varepsilon\, M^{-1}\nabla\psi.$$

Substituting into J and using bilinearity yields

$$\begin{aligned} J(\Phi_\varepsilon) &= \int_{\mathcal{X}} \Big(\big(F - M^{-1}\nabla\Phi\big) - \varepsilon\, M^{-1}\nabla\psi\Big)^\top M \Big(\big(F - M^{-1}\nabla\Phi\big) - \varepsilon\, M^{-1}\nabla\psi\Big)\, d\mu \\ &= \int_{\mathcal{X}} \big(F - M^{-1}\nabla\Phi\big)^\top M \big(F - M^{-1}\nabla\Phi\big)\, d\mu - 2\varepsilon \int_{\mathcal{X}} \big(F - M^{-1}\nabla\Phi\big)^\top M \big(M^{-1}\nabla\psi\big)\, d\mu \\ &\qquad + \varepsilon^2 \int_{\mathcal{X}} \big(M^{-1}\nabla\psi\big)^\top M \big(M^{-1}\nabla\psi\big)\, d\mu. \end{aligned}$$

Because $M(M^{-1}\nabla\psi) = \nabla\psi$ and $\big(M^{-1}\nabla\psi\big)^\top M \big(M^{-1}\nabla\psi\big) = \nabla\psi^\top M^{-1}\nabla\psi$, differentiating in ε gives

$$\frac{d}{d\varepsilon} J(\Phi_\varepsilon) = -2 \int_{\mathcal{X}} \big(F - M^{-1}\nabla\Phi\big)^\top \nabla\psi\, d\mu + 2\varepsilon \int_{\mathcal{X}} \nabla\psi^\top M^{-1}\nabla\psi\, d\mu,$$

and evaluating at $\varepsilon = 0$ yields

$$\frac{d}{d\varepsilon} J(\Phi + \varepsilon\psi)\Big|_{\varepsilon=0} = -2 \int_{\mathcal{X}} \big(F - M^{-1}\nabla\Phi\big)^\top \nabla\psi\, d\mu = 2 \int_{\mathcal{X}} \big(M^{-1}\nabla\Phi - F\big)^\top \nabla\psi\, d\mu,$$

as claimed. □

Let $\Phi^\star$ be any minimizer. Since J is (Gateaux) differentiable and $\Phi^\star$ is an unconstrained minimizer over the affine space $H^1(\mathcal{X})$, the first-order optimality condition is

$$\frac{d}{d\varepsilon} J(\Phi^\star + \varepsilon\psi)\Big|_{\varepsilon=0} = 0 \qquad \forall\, \psi \in H^1(\mathcal{X}).$$

Applying Lemma H.2 at $\Phi^\star$ yields the weak orthogonality condition

$$\int_{\mathcal{X}} \big(M^{-1}\nabla\Phi^\star - F\big)^\top \nabla\psi\, d\mu = 0 \qquad \forall\, \psi \in H^1(\mathcal{X}).$$

Equivalently,

$$\int_{\mathcal{X}} \big(M^{-1}\nabla\Phi^\star\big)^\top \nabla\psi\, d\mu = \int_{\mathcal{X}} F^\top \nabla\psi\, d\mu \qquad \forall\, \psi \in H^1(\mathcal{X}).$$

Now apply Lemma C.1 twice (with $U = M^{-1}\nabla\Phi^\star$ and $U = F$): under the stated boundary assumptions removing boundary terms, for every $\psi \in H^1(\mathcal{X})$ we have

$$\int_{\mathcal{X}} \big(M^{-1}\nabla\Phi^\star\big)^\top \nabla\psi\, d\mu = -\int_{\mathcal{X}} \psi \div_\mu \big(M^{-1}\nabla\Phi^\star\big)\, d\mu, \qquad \int_{\mathcal{X}} F^\top \nabla\psi\, d\mu = -\int_{\mathcal{X}} \psi \div_\mu F\, d\mu.$$

Substituting these identities into the weak orthogonality equality gives

$$-\int_{\mathcal{X}} \psi \div_\mu \big(M^{-1}\nabla\Phi^\star\big)\, d\mu = -\int_{\mathcal{X}} \psi \div_\mu F\, d\mu \qquad \forall\, \psi \in H^1(\mathcal{X}),$$

hence

$$\int_{\mathcal{X}} \psi\Big(\div_\mu \big(M^{-1}\nabla\Phi^\star\big) - \div_\mu F\Big)\, d\mu = 0 \qquad \forall\, \psi \in H^1(\mathcal{X}).$$

This is precisely the statement that

$$\div_\mu\big(M^{-1}\nabla\Phi^\star\big) = \div_\mu F \quad \text{in } H^{-1}(\mathcal{X}),$$

i.e. $\mathcal{L}_{M,\mu}\Phi^\star = \div_\mu F$ in the weak sense, and the projected field is $\Pi^M_{\mathrm{grad}} F = \nabla_M \Phi^\star = M^{-1}\nabla\Phi^\star$ by Definition 3.1.

For uniqueness up to an additive constant, let $\Phi_1^\star$ and $\Phi_2^\star$ be two minimizers and set $w := \Phi_1^\star - \Phi_2^\star \in H^1(\mathcal{X})$. Subtracting the two weak equations yields

$$\div_\mu\big(M^{-1}\nabla w\big) = 0 \quad \text{in } H^{-1}(\mathcal{X}).$$

Testing this equality with $\psi = w \in H^1(\mathcal{X})$ and applying Lemma C.1 gives

$$0 = \int_{\mathcal{X}} w \div_\mu \big(M^{-1}\nabla w\big)\, d\mu = -\int_{\mathcal{X}} \big(M^{-1}\nabla w\big)^\top \nabla w\, d\mu = -\int_{\mathcal{X}} \nabla w^\top M^{-1}\nabla w\, d\mu.$$

Therefore

$$\int_{\mathcal{X}} \nabla w^\top M^{-1}\nabla w\, d\mu = 0.$$

Since $M^{-1} \succ 0$, $\nabla w^\top M^{-1}\nabla w \geq \lambda_{\min}(M^{-1})\|\nabla w\|_2^2$ pointwise, hence the integral can vanish only if $\nabla w = 0$ μ-a.e. on $\mathcal{X}$. Consequently,

$$M^{-1}\nabla\Phi_1^\star = M^{-1}\nabla\Phi_2^\star \quad \mu\text{-a.e.},$$

so the projected field $\nabla_M\Phi^\star$ is unique μ-a.e., while $\Phi^\star$ is determined only up to an additive constant (more precisely, up to constants on each connected component). One may fix a gauge such as $\mathbb{E}_\mu[\Phi^\star] = 0$ to select a representative.

Finally, the stated boundary condition arises as the natural boundary condition of the variational problem when one does not impose boundary constraints on Φ. Writing $U := M^{-1}\nabla\Phi^\star - F$ and assuming sufficient trace regularity so that classical integration by parts applies, one has

$$\int_{\mathcal{X}} U^\top \nabla\psi\, d\mu = -\int_{\mathcal{X}} \psi \div_\mu U\, d\mu + \int_{\partial\mathcal{X}} \psi\, \rho\, U\cdot n\, dS.$$

At optimality the first variation gives $\int_{\mathcal{X}} U^\top\nabla\psi\, d\mu = 0$ for all ψ, while the interior Euler–Lagrange equation gives $\div_\mu U = 0$ in $\mathcal{X}$; therefore the boundary term must vanish for arbitrary boundary traces of ψ, forcing

$$\rho\, U\cdot n = \rho\,(M^{-1}\nabla\Phi^\star - F)\cdot n = 0 \quad \text{on } \partial\mathcal{X},$$

which is exactly the zero-flux residual condition. □

H.5 Proof of Lemma 3.2

Proof. By Definition 3.1, pick a minimizer $\Phi^\star \in H^1(\mathcal{X})$ and set

$$G^\star := \Pi^M_{\mathrm{grad}} F = \nabla_M\Phi^\star, \qquad R := F - G^\star = F - \nabla_M\Phi^\star.$$

Consider the least-squares functional on potentials

$$J(\Psi) := \mathbb{E}_{x\sim\mu}\big[\|F(x) - \nabla_M\Psi(x)\|_M^2\big], \qquad \Psi \in H^1(\mathcal{X}).$$

Fix an arbitrary test potential $\Phi \in H^1(\mathcal{X})$ and form the one-dimensional perturbation

$$\Psi_\varepsilon := \Phi^\star + \varepsilon\Phi, \qquad \varepsilon \in \mathbb{R}.$$

Using linearity of ∇_M, we have

$$\nabla_M\Psi_\varepsilon = \nabla_M(\Phi^\star + \varepsilon\Phi) = \nabla_M\Phi^\star + \varepsilon\nabla_M\Phi = G^\star + \varepsilon\nabla_M\Phi.$$

Therefore the residual under this perturbation is

$$F - \nabla_M \Psi_\varepsilon = F - (G^\star + \varepsilon \nabla_M \Phi) = (F - G^\star) - \varepsilon \nabla_M \Phi = R - \varepsilon \nabla_M \Phi.$$

Now expand the squared M-norm using $\|v\|_M^2 = \langle v, v \rangle_M$ and bilinearity of the M-inner product:

$$\begin{aligned} \|R - \varepsilon \nabla_M \Phi\|_M^2 &= \langle R - \varepsilon \nabla_M \Phi,\ R - \varepsilon \nabla_M \Phi \rangle_M \\ &= \langle R, R \rangle_M - \varepsilon \langle R, \nabla_M \Phi \rangle_M - \varepsilon \langle \nabla_M \Phi, R \rangle_M + \varepsilon^2 \langle \nabla_M \Phi, \nabla_M \Phi \rangle_M. \end{aligned}$$

Because $\langle \cdot, \cdot \rangle_M$ is symmetric (since $M \succ 0$), we have $\langle R, \nabla_M \Phi \rangle_M = \langle \nabla_M \Phi, R \rangle_M$, hence

$$\|R - \varepsilon \nabla_M \Phi\|_M^2 = \|R\|_M^2 - 2\varepsilon \langle R, \nabla_M \Phi \rangle_M + \varepsilon^2 \|\nabla_M \Phi\|_M^2.$$

Taking expectation with respect to μ gives

$$J(\Psi_\varepsilon) = \mathbb{E}_\mu[\|R\|_M^2] - 2\varepsilon \, \mathbb{E}_\mu[\langle R, \nabla_M \Phi \rangle_M] + \varepsilon^2 \, \mathbb{E}_\mu[\|\nabla_M \Phi\|_M^2].$$

Differentiate this scalar quadratic function of ε:

$$\frac{d}{d\varepsilon} J(\Psi_\varepsilon) = -2 \, \mathbb{E}_\mu[\langle R, \nabla_M \Phi \rangle_M] + 2\varepsilon \, \mathbb{E}_\mu[\|\nabla_M \Phi\|_M^2].$$

Since $\Phi^\star$ is a minimizer of J over the whole (affine) space $H^1(\mathcal{X})$, the first-order optimality condition for every direction $\Phi \in H^1(\mathcal{X})$ is

$$\frac{d}{d\varepsilon} J(\Psi_\varepsilon)\Big|_{\varepsilon=0} = 0.$$

Evaluating the derivative at $\varepsilon = 0$ yields

$$-2 \, \mathbb{E}_\mu[\langle R, \nabla_M \Phi \rangle_M] = 0,$$

and therefore

$$\mathbb{E}_{x \sim \mu}\big[\langle R(x), \nabla_M \Phi(x) \rangle_M\big] = 0 \qquad \forall \, \Phi \in H^1(\mathcal{X}).$$

This is exactly the stated $L^2(M, \mu)$-orthogonality of the residual R to every metric-gradient field $\nabla_M \Phi$. □

H.6 Proof of Corollary 3.3

Proof. Recall the decomposition $F = G^\star + R$, where $G^\star = \Pi_{\mathrm{grad}}^M F = \nabla_M \Phi^\star$ for some minimizer $\Phi^\star \in H^1(\mathcal{X})$ and $R := F - G^\star$. For each $x \in \mathcal{X}$, expand the squared M-norm using $\|v\|_M^2 = \langle v, v \rangle_M$ and bilinearity of $\langle \cdot, \cdot \rangle_M$:

$$\begin{aligned} \|F(x)\|_M^2 &= \langle F(x), F(x) \rangle_M \\ &= \langle G^\star(x) + R(x),\ G^\star(x) + R(x) \rangle_M \\ &= \langle G^\star(x), G^\star(x) \rangle_M + \langle G^\star(x), R(x) \rangle_M + \langle R(x), G^\star(x) \rangle_M + \langle R(x), R(x) \rangle_M. \end{aligned}$$

Because $\langle \cdot, \cdot \rangle_M$ is symmetric, $\langle G^\star, R \rangle_M = \langle R, G^\star \rangle_M$, so

$$\|F(x)\|_M^2 = \|G^\star(x)\|_M^2 + 2 \langle R(x), G^\star(x) \rangle_M + \|R(x)\|_M^2.$$

Taking expectation with respect to $x \sim \mu$ yields

$$\mathbb{E}_\mu[\|F(x)\|_M^2] = \mathbb{E}_\mu[\|G^\star(x)\|_M^2] + 2 \, \mathbb{E}_\mu[\langle R(x), G^\star(x) \rangle_M] + \mathbb{E}_\mu[\|R(x)\|_M^2].$$

It remains to show the cross term vanishes. By Lemma 3.2, for every test potential $\Phi \in H^1(\mathcal{X})$,

$$\mathbb{E}_\mu[\langle R(x), \nabla_M \Phi(x) \rangle_M] = 0.$$

Choose $\Phi = \Phi^\star$ (a minimizer inducing $G^\star = \nabla_M \Phi^\star$). Then

$$\mathbb{E}_\mu[\langle R(x), G^\star(x) \rangle_M] = \mathbb{E}_\mu[\langle R(x), \nabla_M \Phi^\star(x) \rangle_M] = 0.$$

Substituting this into the previous expansion gives

$$\mathbb{E}_{x \sim \mu}\big[\|F(x)\|_M^2\big] = \mathbb{E}_{x \sim \mu}\big[\|G^\star(x)\|_M^2\big] + \mathbb{E}_{x \sim \mu}\big[\|R(x)\|_M^2\big],$$

which is the claimed Pythagorean identity. □

H.7 Proof of Proposition C.5

Proof. Because $G^\star = \nabla_M \Phi^\star$, by definition of the metric-gradient we have

$$\nabla_M \Phi^\star(x) = M^{-1}\nabla\Phi^\star(x), \qquad \text{so in particular} \qquad M\, G^\star(x) = \nabla\Phi^\star(x) \ \text{ for all } x \in \mathcal{X}.$$

By the definition of circulation and the M-inner product $\langle u, v\rangle_M = u^\top M v$, for any piecewise C^1 closed curve $\gamma : [0,1] \to \mathcal{X}$ we can write

$$\circ_{G^\star}(\gamma) = \int_0^1 \langle G^\star(\gamma(t)), \dot\gamma(t)\rangle_M \, dt = \int_0^1 G^\star(\gamma(t))^\top M\, \dot\gamma(t)\, dt.$$

Substituting $G^\star(\gamma(t)) = M^{-1}\nabla\Phi^\star(\gamma(t))$ into the integrand yields

$$G^\star(\gamma(t))^\top M\, \dot\gamma(t) = \big(M^{-1}\nabla\Phi^\star(\gamma(t))\big)^\top M\, \dot\gamma(t) = \nabla\Phi^\star(\gamma(t))^\top \dot\gamma(t),$$

where we used that $(M^{-1})^\top = M^{-1}$ and $M^{-1}M = I$ because M is symmetric positive definite. Hence

$$\circ_{G^\star}(\gamma) = \int_0^1 \nabla\Phi^\star(\gamma(t))^\top \dot\gamma(t)\, dt.$$

Since γ is piecewise C^1 and $\Phi^\star$ is differentiable, the chain rule applies on each smooth sub-interval of $[0,1]$ and gives

$$\frac{d}{dt}\Phi^\star(\gamma(t)) = \nabla\Phi^\star(\gamma(t))^\top \dot\gamma(t) \quad \text{for all } t \text{ where } \gamma \text{ is differentiable.}$$

Therefore, integrating over $[0,1]$ (equivalently, summing over the finitely many smooth pieces) and applying the fundamental theorem of calculus yields

$$\circ_{G^\star}(\gamma) = \int_0^1 \frac{d}{dt}\Phi^\star(\gamma(t))\, dt = \Phi^\star(\gamma(1)) - \Phi^\star(\gamma(0)).$$

Because γ is closed, $\gamma(1) = \gamma(0)$, so $\Phi^\star(\gamma(1)) - \Phi^\star(\gamma(0)) = 0$ and thus $\circ_{G^\star}(\gamma) = 0$. Now using $F = G^\star + R$ and the linearity of both the inner product and the integral, we have

$$\circ_F(\gamma) = \int_0^1 \langle F(\gamma(t)), \dot\gamma(t)\rangle_M \, dt = \int_0^1 \langle G^\star(\gamma(t)) + R(\gamma(t)), \dot\gamma(t)\rangle_M \, dt = \circ_{G^\star}(\gamma) + \circ_R(\gamma) = \circ_R(\gamma),$$

which proves $\circ_F(\gamma) = \circ_R(\gamma)$.

For the quantitative bound, start from the identity just shown and write

$$|\circ_F(\gamma)| = |\circ_R(\gamma)| = \left| \int_0^1 \langle R(\gamma(t)), \dot\gamma(t)\rangle_M \, dt \right| = \left| \int_0^1 R(\gamma(t))^\top M\, \dot\gamma(t)\, dt \right|.$$

Let $M^{1/2}$ denote the (symmetric) principal square root of M, so that $M^{1/2}M^{1/2} = M$ and $(M^{1/2})^\top = M^{1/2}$. Then for each t where γ is differentiable,

$$R(\gamma(t))^\top M\, \dot\gamma(t) = R(\gamma(t))^\top M^{1/2} M^{1/2} \dot\gamma(t) = \big(M^{1/2}R(\gamma(t))\big)^\top \big(M^{1/2}\dot\gamma(t)\big).$$

Applying the Euclidean Cauchy–Schwarz inequality to the last inner product gives the pointwise bound

$$\big|R(\gamma(t))^\top M\, \dot\gamma(t)\big| \le \|M^{1/2}R(\gamma(t))\|_2 \, \|M^{1/2}\dot\gamma(t)\|_2 = \|R(\gamma(t))\|_M \, \|\dot\gamma(t)\|_M,$$

where we used $\|x\|_M^2 = x^\top M x = \|M^{1/2}x\|_2^2$. Using this bound inside the integral and then applying Cauchy–Schwarz in $L^2([0,1])$ yields

$$|\circ_F(\gamma)| \le \int_0^1 \|R(\gamma(t))\|_M \, \|\dot\gamma(t)\|_M \, dt \le \left(\int_0^1 \|R(\gamma(t))\|_M^2 \, dt\right)^{1/2} \left(\int_0^1 \|\dot\gamma(t)\|_M^2 \, dt\right)^{1/2}.$$

Finally, because $t \sim \mathrm{Unif}[0,1]$ satisfies $\mathbb{E}[g(t)] = \int_0^1 g(t)\, dt$ for any integrable g, we can rewrite the last display as

$$|\circ_F(\gamma)| \le \Big(\mathbb{E}_{t\sim\mathrm{Unif}[0,1]} \|R(\gamma(t))\|_M^2\Big)^{1/2} \cdot \Big(\mathbb{E}_{t\sim\mathrm{Unif}[0,1]} \|\dot\gamma(t)\|_M^2\Big)^{1/2}.$$

This is exactly the stated bound. □

H.8 Proof of Proposition 5.1

Proof. Recall the HPML decomposition

$$F \;=\; G^\star + R, \qquad G^\star := \Pi^M_{\mathrm{grad}} F, \qquad R := F - G^\star.$$

By definition of the projection, there exists a scalar potential $\Phi^\star$ such that

$$G^\star = \nabla_M \Phi^\star = M^{-1} \nabla \Phi^\star.$$

Assume $R \equiv 0$ on U. Then, by the decomposition above,

$$F = G^\star = \Pi^M_{\mathrm{grad}} F = \nabla_M \Phi^\star \qquad \text{on } U.$$

Hence the joint field is integrable on U: all agents' first-order update directions are aligned with ascent of the single scalar potential $\Phi^\star$.

In particular, when HPML is run with exact projected ascent on the feasible set $\mathcal{X}$,

$$x_{t+1} = \mathrm{Proj}_{\mathcal{X}}\big(x_t + \eta \nabla \Phi^\star(x_t)\big),$$

the Lyapunov-improvement result already established in Theorem 6.4 (or the corresponding projected-gradient theorem in the main text) applies directly. Therefore $\Phi^\star$ increases monotonically along non-stationary iterates, and the projected dynamics cannot sustain a nontrivial cycle.

Assume there exists a piecewise-C^1 closed curve $\gamma : [0,1] \to U$ with $\gamma(0) = \gamma(1)$ such that

$$\oint_\gamma \langle MF(x), dx \rangle \;=\; \circ_F(\gamma) \;\neq\; 0.$$

We first show that no scalar potential Φ on U can satisfy $F = \nabla_M \Phi$ on U.

Suppose, for contradiction, that such a Φ exists. Then along the loop γ,

$$\circ_F(\gamma) = \int_0^1 \langle \nabla_M \Phi(\gamma(t)), \dot\gamma(t) \rangle_M \, dt = \int_0^1 \langle M \nabla_M \Phi(\gamma(t)), \dot\gamma(t) \rangle \, dt.$$

Since $M \nabla_M \Phi = \nabla \Phi$, this becomes

$$\circ_F(\gamma) = \int_0^1 \langle \nabla \Phi(\gamma(t)), \dot\gamma(t) \rangle \, dt = \int_0^1 \frac{d}{dt} \Phi(\gamma(t)) \, dt = \Phi(\gamma(1)) - \Phi(\gamma(0)) = 0,$$

because γ is closed. This contradicts the assumption $\circ_F(\gamma) \neq 0$. Hence no scalar potential on U can generate F as a metric-gradient field.

It remains to identify which component carries this circulation. By Proposition C.5, for every closed curve,

$$\circ_{G^\star}(\gamma) = 0, \qquad \text{and hence} \qquad \circ_F(\gamma) = \circ_R(\gamma).$$

Therefore,

$$\oint_\gamma \langle MF(x), dx \rangle = \oint_\gamma \langle MR(x), dx \rangle \neq 0.$$

So any nonzero closed-loop circulation of the original joint field is carried entirely by the residual R.

Consequently, R is the obstruction to joint greedy integrability: when such a loop exists, the simultaneous critic-driven actor updates encoded by F cannot be interpreted as ascent of a single Bellman-compatible scalar objective on the joint policy space. Equivalently, the original joint improvement field retains an irreducible circulation component, which rules out a global potential representation and permits recurrent convention/role-allocation loops. □

H.9 Proof of Proposition E.5

Proof. The equation $R(u, v) = (0, 0)$ implies $v = 0$ and $u = 0$, so $(0, 0)$ is the unique fixed point.

For the continuous-time dynamics $(\dot{u}, \dot{v}) = (v, -u)$,

$$\frac{d}{dt}\big(u(t)^2 + v(t)^2\big) = 2u(t)\dot{u}(t) + 2v(t)\dot{v}(t) = 2u(t)v(t) - 2v(t)u(t) = 0.$$

Hence $u(t)^2 + v(t)^2$ is constant along trajectories. Therefore every non-stationary solution stays on the circle determined by its initial radius and never reaches the origin. In particular, each circle

$$\gamma_r(s) = r(\cos 2\pi s, \sin 2\pi s)$$

is an invariant orbit.

For the explicit-Euler update,

$$u_{t+1} = u_t + \eta v_t, \qquad v_{t+1} = v_t - \eta u_t.$$

Thus

$$u_{t+1}^2 + v_{t+1}^2 = (u_t + \eta v_t)^2 + (v_t - \eta u_t)^2 = (1 + \eta^2)(u_t^2 + v_t^2).$$

Equivalently,

$$\|z_{t+1}\|^2 = (1 + \eta^2)\|z_t\|^2.$$

Since $1 + \eta^2 > 1$ for every $\eta > 0$, every nonzero iterate moves strictly away from the origin, and iteration yields

$$\|z_t\| = (1 + \eta^2)^{t/2}\|z_0\|.$$

This proves both claims. □

H.10 Proof of Lemma 4.2

Proof. Endow $\mathbb{R}^m$ (edge-flow space) with the weighted inner product $\langle a, b\rangle_W := a^\top W b$ and weighted norm $\|a\|_W^2 := a^\top W a$ (here W is symmetric positive definite, typically diagonal with positive weights). Consider the weighted least-squares projection of the flow ω onto the discrete gradient subspace $\mathrm{Range}(B)$, i.e. define the objective

$$J(\phi) := \frac{1}{2}\|\omega - B\phi\|_W^2 = \frac{1}{2}(\omega - B\phi)^\top W(\omega - B\phi), \qquad \phi \in \mathbb{R}^N.$$

Let $\phi^\star$ be any minimizer of J, and define the residual (cyclic component)

$$\omega_{\mathrm{cyc}} := \omega - B\phi^\star.$$

Fix an arbitrary direction $\psi \in \mathbb{R}^N$ and consider the one-dimensional perturbation $\phi_\varepsilon := \phi^\star + \varepsilon\psi$. Then

$$\omega - B\phi_\varepsilon = \omega - B(\phi^\star + \varepsilon\psi) = (\omega - B\phi^\star) - \varepsilon B\psi = \omega_{\mathrm{cyc}} - \varepsilon B\psi.$$

Substituting into J and expanding the quadratic form gives

$$\begin{aligned} J(\phi_\varepsilon) &= \frac{1}{2}(\omega_{\mathrm{cyc}} - \varepsilon B\psi)^\top W(\omega_{\mathrm{cyc}} - \varepsilon B\psi) \\ &= \frac{1}{2}\Big(\omega_{\mathrm{cyc}}^\top W\omega_{\mathrm{cyc}} - \varepsilon\,\omega_{\mathrm{cyc}}^\top WB\psi - \varepsilon\,(B\psi)^\top W\omega_{\mathrm{cyc}} + \varepsilon^2 (B\psi)^\top W(B\psi)\Big). \end{aligned}$$

Because W is symmetric, $\omega_{\mathrm{cyc}}^\top WB\psi = (B\psi)^\top W\omega_{\mathrm{cyc}}$, hence

$$J(\phi_\varepsilon) = \frac{1}{2}\,\omega_{\mathrm{cyc}}^\top W\omega_{\mathrm{cyc}} - \varepsilon\,(B\psi)^\top W\omega_{\mathrm{cyc}} + \frac{\varepsilon^2}{2}\,(B\psi)^\top W(B\psi).$$

Differentiate with respect to ε:

$$\frac{d}{d\varepsilon}J(\phi_\varepsilon) = -(B\psi)^\top W\omega_{\mathrm{cyc}} + \varepsilon\,(B\psi)^\top W(B\psi).$$

Since $\phi^\star$ is a minimizer of J over the whole space $\mathbb{R}^N$, the first-order optimality condition along every direction ψ is

$$\frac{d}{d\varepsilon}J(\phi^\star + \varepsilon\psi)\Big|_{\varepsilon=0} = 0.$$

Evaluating the derivative at $\varepsilon = 0$ yields

$$-(B\psi)^\top W\omega_{\mathrm{cyc}} = 0,$$

i.e.

$$(B\psi)^\top W\,\omega_{\mathrm{cyc}} = 0, \qquad \forall\,\psi \in \mathbb{R}^N.$$

This is exactly the stated orthogonality: the residual $\omega_{\mathrm{cyc}} = \omega - B\phi^\star$ is orthogonal (in the weighted inner product $\langle\cdot,\cdot\rangle_W$) to every discrete gradient flow $B\psi$. Equivalently, since $(B\psi)^\top W\omega_{\mathrm{cyc}} = \psi^\top(B^\top W\omega_{\mathrm{cyc}})$ for all ψ, one may write the normal equation $B^\top W\omega_{\mathrm{cyc}} = 0$. □

H.11 Proof of Corollary 4.3

Proof. Recall that the weighted inner product and norm are $\langle a, b\rangle_W := a^\top W b$ and $\|a\|_W^2 := a^\top W a$. By definition,

$$\omega_{\mathrm{pot}} = B\phi^\star, \qquad \omega_{\mathrm{cyc}} = \omega - \omega_{\mathrm{pot}}, \qquad \text{hence} \qquad \omega = \omega_{\mathrm{pot}} + \omega_{\mathrm{cyc}}.$$

Expand the weighted energy using bilinearity and symmetry of $\langle\cdot,\cdot\rangle_W$ (since $W = W^\top$):

$$\begin{aligned}\|\omega\|_W^2 &= \langle \omega, \omega\rangle_W \\ &= \langle \omega_{\mathrm{pot}} + \omega_{\mathrm{cyc}},\, \omega_{\mathrm{pot}} + \omega_{\mathrm{cyc}}\rangle_W \\ &= \langle \omega_{\mathrm{pot}}, \omega_{\mathrm{pot}}\rangle_W + \langle \omega_{\mathrm{pot}}, \omega_{\mathrm{cyc}}\rangle_W + \langle \omega_{\mathrm{cyc}}, \omega_{\mathrm{pot}}\rangle_W + \langle \omega_{\mathrm{cyc}}, \omega_{\mathrm{cyc}}\rangle_W \\ &= \|\omega_{\mathrm{pot}}\|_W^2 + 2\,\langle \omega_{\mathrm{pot}}, \omega_{\mathrm{cyc}}\rangle_W + \|\omega_{\mathrm{cyc}}\|_W^2.\end{aligned}$$

It remains to show the cross term vanishes. By Lemma 4.2, for every $\psi \in \mathbb{R}^N$,

$$(B\psi)^\top W\, \omega_{\mathrm{cyc}} = 0.$$

Choose $\psi = \phi^\star$. Then

$$(B\phi^\star)^\top W\, \omega_{\mathrm{cyc}} = 0.$$

Since $B\phi^\star = \omega_{\mathrm{pot}}$, this is exactly

$$\langle \omega_{\mathrm{pot}}, \omega_{\mathrm{cyc}}\rangle_W = \omega_{\mathrm{pot}}^\top W\, \omega_{\mathrm{cyc}} = 0.$$

Substituting $\langle \omega_{\mathrm{pot}}, \omega_{\mathrm{cyc}}\rangle_W = 0$ into the expansion above yields

$$\|\omega\|_W^2 = \|\omega_{\mathrm{pot}}\|_W^2 + \|\omega_{\mathrm{cyc}}\|_W^2,$$

which is the claimed discrete energy decomposition. □

H.12 Proof of Lemma 4.5

Proof. Recall the discrete non-potentiality proxy (Definition 4.4): for $\varepsilon > 0$,

$$\mathrm{NonPot}(\omega) := \frac{\|\omega_{\mathrm{cyc}}\|_W^2}{\|\omega\|_W^2 + \varepsilon}, \qquad \|\eta\|_W^2 := \eta^\top W \eta,$$

where $\omega_{\mathrm{pot}} = B\phi^\star$ is the weighted least-squares potential flow and $\omega_{\mathrm{cyc}} := \omega - \omega_{\mathrm{pot}}$ is the residual.

Since $W \succeq 0$ (in fact $W \succ 0$ in typical constructions), we have $\|\omega_{\mathrm{cyc}}\|_W^2 \geq 0$. Also $\|\omega\|_W^2 \geq 0$ and $\varepsilon > 0$, hence $\|\omega\|_W^2 + \varepsilon > 0$. Therefore

$$\mathrm{NonPot}(\omega) = \frac{\|\omega_{\mathrm{cyc}}\|_W^2}{\|\omega\|_W^2 + \varepsilon} \geq 0.$$

To show $\mathrm{NonPot}(\omega) \leq 1$, we use the discrete energy decomposition (Corollary 4.3):

$$\|\omega\|_W^2 = \|\omega_{\mathrm{pot}}\|_W^2 + \|\omega_{\mathrm{cyc}}\|_W^2.$$

In particular, $\|\omega\|_W^2 \geq \|\omega_{\mathrm{cyc}}\|_W^2$, and hence

$$\|\omega\|_W^2 + \varepsilon \;\geq\; \|\omega_{\mathrm{cyc}}\|_W^2.$$

Dividing both sides by the positive quantity $\|\omega\|_W^2 + \varepsilon$ yields

$$\frac{\|\omega_{\mathrm{cyc}}\|_W^2}{\|\omega\|_W^2 + \varepsilon} \leq 1,$$

i.e. $\mathrm{NonPot}(\omega) \leq 1$.

Next, $\mathrm{NonPot}(\omega) = 0$ if and only if $\|\omega_{\mathrm{cyc}}\|_W^2 = 0$ (because the denominator is strictly positive). If $W \succ 0$, then $\|\omega_{\mathrm{cyc}}\|_W^2 = 0$ implies $\omega_{\mathrm{cyc}} = 0$, hence

$$\omega = \omega_{\mathrm{pot}} = B\phi^\star \in \mathrm{Range}(B),$$

so ω is exactly a potential (gradient/cut-space) flow on the graph. Conversely, if ω is a potential flow, i.e. $\omega = B\phi$ for some ϕ, then the best-fit potential flow is $\omega_{\mathrm{pot}} = \omega$ and $\omega_{\mathrm{cyc}} = 0$, hence $\mathrm{NonPot}(\omega) = 0$.

The phrase "$\mathrm{NonPot}(\omega) \approx 0$" can be made precise as a relative-distance statement. Since ω_{pot} is the weighted least-squares projection onto $\mathrm{Range}(B)$, one has

$$\mathrm{dist}_W(\omega, \mathrm{Range}(B)) := \inf_{\phi \in \mathbb{R}^N} \|\omega - B\phi\|_W = \|\omega - \omega_{\mathrm{pot}}\|_W = \|\omega_{\mathrm{cyc}}\|_W.$$

Thus

$$\mathrm{NonPot}(\omega) = \frac{\mathrm{dist}_W(\omega, \mathrm{Range}(B))^2}{\|\omega\|_W^2 + \varepsilon},$$

so $\mathrm{NonPot}(\omega)$ is small exactly when the residual distance to the potential-flow subspace is small relative to the total energy scale $\|\omega\|_W^2 + \varepsilon$.

Finally, to formalize "$\mathrm{NonPot}(\omega)$ increases with the energy of circulation components supported by cycles", note that on a graph (a 1-dimensional complex) the residual ω_{cyc} is precisely the cycle-space (circulation) component of ω, and its circulation energy is $\|\omega_{\mathrm{cyc}}\|_W^2$. Using the decomposition $\|\omega\|_W^2 = \|\omega_{\mathrm{pot}}\|_W^2 + \|\omega_{\mathrm{cyc}}\|_W^2$, define

$$p := \|\omega_{\mathrm{pot}}\|_W^2 \geq 0, \qquad s := \|\omega_{\mathrm{cyc}}\|_W^2 \geq 0.$$

Then

$$\mathrm{NonPot}(\omega) = \frac{s}{p + s + \varepsilon}.$$

For fixed p and $\varepsilon > 0$, the function $f(s) := \frac{s}{p+s+\varepsilon}$ is strictly increasing in s: for $s \geq 0$,

$$f'(s) = \frac{p + \varepsilon}{(p + s + \varepsilon)^2} > 0.$$

Hence, when the potential component is fixed, increasing the circulation (cycle-space) energy $s = \|\omega_{\mathrm{cyc}}\|_W^2$ strictly increases $\mathrm{NonPot}(\omega)$, matching the intended interpretation. □

H.13 Proof of Proposition D.4

Proof. Fix $x \in \mathcal{X}$ and a displacement Δx such that the segment $x + t\Delta x \in \mathcal{X}$ for all $t \in [0, 1]$. Define the one-dimensional function

$$g(t) := \Phi(x + t\Delta x), \qquad t \in [0, 1].$$

Since $\Phi \in C^2(\mathcal{X})$, we have $g \in C^2([0, 1])$, and by the chain rule,

$$g'(t) = \langle \nabla\Phi(x + t\Delta x),\, \Delta x \rangle, \qquad g''(t) = \Delta x^\top \nabla^2\Phi(x + t\Delta x)\, \Delta x.$$

Using the second-order Taylor formula with integral remainder for g around $t = 0$ evaluated at $t = 1$,

$$g(1) = g(0) + g'(0) \cdot 1 + \int_0^1 (1 - t)\, g''(t)\, dt,$$

and substituting the expressions for $g(0), g(1), g'(0), g''(t)$ gives

$$\Phi(x + \Delta x) - \Phi(x) = \langle \nabla\Phi(x),\, \Delta x \rangle + \int_0^1 (1 - t)\, \Delta x^\top \nabla^2\Phi(x + t\Delta x)\, \Delta x\, dt.$$

The remainder term is quadratic in Δx: indeed, letting

$$H_* := \sup_{t \in [0,1]} \left\| \nabla^2\Phi(x + t\Delta x) \right\|_{\mathrm{op}} < \infty$$

(which is finite by continuity of $\nabla^2\Phi$ on the compact segment), we have for every $t \in [0, 1]$,

$$\left| \Delta x^\top \nabla^2\Phi(x + t\Delta x)\, \Delta x \right| \leq \left\| \nabla^2\Phi(x + t\Delta x) \right\|_{\mathrm{op}} \|\Delta x\|^2 \leq H_* \, \|\Delta x\|^2,$$

hence

$$\left| \int_0^1 (1 - t)\, \Delta x^\top \nabla^2\Phi(x + t\Delta x)\, \Delta x\, dt \right| \leq \int_0^1 (1 - t)\, H_* \, \|\Delta x\|^2\, dt = \frac{H_*}{2} \, \|\Delta x\|^2 = O(\|\Delta x\|^2).$$

Therefore

$$\Phi(x+\Delta x)-\Phi(x)=\langle\nabla\Phi(x),\,\Delta x\rangle+O(\|\Delta x\|^2).$$

Since $M\succ 0$ is constant and $\nabla_M\Phi:=M^{-1}\nabla\Phi$, we have the identity

$$M\nabla_M\Phi(x)=M(M^{-1}\nabla\Phi(x))=\nabla\Phi(x),$$

and thus

$$\langle\nabla\Phi(x),\,\Delta x\rangle=\langle M\nabla_M\Phi(x),\,\Delta x\rangle,$$

which gives the second displayed equality in the proposition.

For the graph statement, let nodes have positions $\{x^i\}$ and write $\Delta_{ij}:=x^j-x^i$. Define the edge 1-form

$$\omega_{ij}:=\Big\langle M\,\frac{G(x^i)+G(x^j)}{2},\,\Delta_{ij}\Big\rangle.$$

If $G=\nabla_M\Phi$, then $MG=\nabla\Phi$, so

$$\omega_{ij}=\Big\langle\frac{\nabla\Phi(x^i)+\nabla\Phi(x^j)}{2},\,\Delta_{ij}\Big\rangle.$$

On the other hand, by the fundamental theorem of calculus along the straight segment $\gamma(t):=x^i+t\Delta_{ij}$, we have

$$\Phi(x^j)-\Phi(x^i)=\int_0^1\frac{d}{dt}\Phi(\gamma(t))\,dt=\int_0^1\langle\nabla\Phi(\gamma(t)),\,\gamma'(t)\rangle\,dt=\int_0^1\langle\nabla\Phi(x^i+t\Delta_{ij}),\,\Delta_{ij}\rangle\,dt.$$

To compare this integral to the endpoint average, use the integral representation of the gradient. For each $t\in[0,1]$, the map $s\mapsto\nabla\Phi(x^i+s\Delta_{ij})$ is C^1, and

$$\nabla\Phi(x^i+t\Delta_{ij})=\nabla\Phi(x^i)+\int_0^t\nabla^2\Phi(x^i+s\Delta_{ij})\,\Delta_{ij}\,ds.$$

Substituting into the line integral yields

$$\begin{aligned}\Phi(x^j)-\Phi(x^i)&=\int_0^1\Big\langle\nabla\Phi(x^i)+\int_0^t\nabla^2\Phi(x^i+s\Delta_{ij})\,\Delta_{ij}\,ds,\;\Delta_{ij}\Big\rangle\,dt\\&=\int_0^1\langle\nabla\Phi(x^i),\Delta_{ij}\rangle\,dt+\int_0^1\int_0^t\Delta_{ij}^\top\nabla^2\Phi(x^i+s\Delta_{ij})\,\Delta_{ij}\,ds\,dt\\&=\langle\nabla\Phi(x^i),\Delta_{ij}\rangle+\int_0^1(1-s)\,\Delta_{ij}^\top\nabla^2\Phi(x^i+s\Delta_{ij})\,\Delta_{ij}\,ds,\end{aligned}$$

where the last line uses Fubini's theorem to swap the order of integration: $\int_0^1\int_0^t(\cdot)\,ds\,dt=\int_0^1\int_s^1(\cdot)\,dt\,ds=\int_0^1(1-s)(\cdot)\,ds$. Similarly, using the same identity at $t=1$ gives

$$\nabla\Phi(x^j)=\nabla\Phi(x^i+\Delta_{ij})=\nabla\Phi(x^i)+\int_0^1\nabla^2\Phi(x^i+s\Delta_{ij})\,\Delta_{ij}\,ds,$$

hence

$$\Big\langle\frac{\nabla\Phi(x^i)+\nabla\Phi(x^j)}{2},\,\Delta_{ij}\Big\rangle=\langle\nabla\Phi(x^i),\Delta_{ij}\rangle+\frac12\int_0^1\Delta_{ij}^\top\nabla^2\Phi(x^i+s\Delta_{ij})\,\Delta_{ij}\,ds.$$

Subtracting the two expressions yields

$$\begin{aligned}\omega_{ij}-\big(\Phi(x^j)-\Phi(x^i)\big)&=\int_0^1\Big(\frac12-(1-s)\Big)\,\Delta_{ij}^\top\nabla^2\Phi(x^i+s\Delta_{ij})\,\Delta_{ij}\,ds\\&=\int_0^1\Big(s-\frac12\Big)\,\Delta_{ij}^\top\nabla^2\Phi(x^i+s\Delta_{ij})\,\Delta_{ij}\,ds.\end{aligned}$$

Let $H_{ij}:=\sup_{s\in[0,1]}\|\nabla^2\Phi(x^i+s\Delta_{ij})\|_{\mathrm{op}}<\infty$. Then

$$\big|\omega_{ij}-\big(\Phi(x^j)-\Phi(x^i)\big)\big|\le\int_0^1\Big|s-\frac12\Big|\,H_{ij}\,\|\Delta_{ij}\|^2\,ds=H_{ij}\,\|\Delta_{ij}\|^2\int_0^1\Big|s-\frac12\Big|\,ds=\frac{H_{ij}}{4}\,\|\Delta_{ij}\|^2.$$

Thus

$$\omega_{ij}=\Phi(x^j)-\Phi(x^i)+O(\|x^j-x^i\|^2),$$

which is the claimed second-order (metric-consistent) approximation of potential differences. □

H.14 Proof of Lemma F.1

Proof. Fix $y \in \mathbb{R}^d$ and write $p := P(y) \in \mathcal{X}$. By definition of the Euclidean projection onto a nonempty closed convex set, p is the unique minimizer of the strongly convex problem

$$\min_{x \in \mathcal{X}} f(x) := \frac{1}{2}\|x - y\|^2.$$

Let $z \in \mathcal{X}$ be arbitrary and consider the line segment

$$x(t) := p + t(z - p), \qquad t \in [0, 1].$$

Because $\mathcal{X}$ is convex, $x(t) \in \mathcal{X}$ for all $t \in [0, 1]$. Since p minimizes f over $\mathcal{X}$, the one-variable function $t \mapsto f(x(t))$ attains its minimum at $t = 0$ on $[0, 1]$, hence its right derivative at $t = 0$ is nonnegative:

$$\frac{d}{dt} f(x(t))\Big|_{t=0^+} \geq 0.$$

We compute this derivative explicitly. First,

$$f(x(t)) = \frac{1}{2}\|x(t) - y\|^2 = \frac{1}{2}\|p + t(z - p) - y\|^2.$$

Differentiate using $\frac{d}{dt}\frac{1}{2}\|u(t)\|^2 = \langle u(t), u'(t)\rangle$ with $u(t) = p + t(z - p) - y$ and $u'(t) = z - p$:

$$\frac{d}{dt} f(x(t)) = \big\langle p + t(z - p) - y,\ z - p \big\rangle.$$

Evaluating at $t = 0$ gives

$$\frac{d}{dt} f(x(t))\Big|_{t=0} = \langle p - y,\ z - p\rangle.$$

Combining with the necessary optimality condition yields

$$\langle p - y,\ z - p\rangle \geq 0.$$

Multiplying by -1 and recalling $p = P(y)$ gives the projection variational inequality

$$\langle y - P(y),\ z - P(y)\rangle \leq 0, \qquad \forall z \in \mathcal{X}.$$

For firm nonexpansiveness, take arbitrary $y, y' \in \mathbb{R}^d$ and denote

$$p := P(y), \qquad q := P(y').$$

Applying the projection variational inequality with $(y, z) = (y, q)$ gives

$$\langle y - p, q - p\rangle \leq 0,$$

or equivalently,

$$\langle y - p, p - q\rangle \geq 0.$$

Applying it again with $(y', z) = (y', p)$ gives

$$\langle y' - q, p - q\rangle \leq 0.$$

Subtracting the second inequality from the first yields

$$\langle (y - p) - (y' - q), p - q\rangle \geq 0.$$

Since

$$(y - p) - (y' - q) = (y - y') - (p - q),$$

we obtain

$$\langle y - y', p - q\rangle - \|p - q\|^2 \geq 0,$$

and therefore

$$\|P(y) - P(y')\|^2 \leq \langle P(y) - P(y'), y - y'\rangle.$$

Finally, Cauchy–Schwarz gives

$$\|P(y) - P(y')\|^2 \leq \|P(y) - P(y')\|\,\|y - y'\|,$$

so

$$\|P(y) - P(y')\| \leq \|y - y'\|.$$

□

H.15 Proof of Theorem 6.4

Proof. Let $g_t := \nabla\Phi(x_t)$ and denote the step $\Delta_t := x_{t+1} - x_t$. We will combine (a) the L-smooth lower quadratic bound for Φ and (b) the projection variational inequality.

Lemma H.3 (Two-sided smoothness (quadratic lower bound))**.** *If Φ has L-Lipschitz gradient on $\mathcal{X}$, then for all $x, y \in \mathcal{X}$,*

$$\Phi(y) \geq \Phi(x) + \langle \nabla\Phi(x), y - x\rangle - \frac{L}{2}\|y - x\|^2.$$

Proof. Fix x, y and define $h(t) := \Phi(x + t(y - x))$ for $t \in [0, 1]$. Then h is differentiable and

$$h'(t) = \langle \nabla\Phi(x + t(y - x)),\, y - x\rangle.$$

By the fundamental theorem of calculus,

$$\Phi(y) - \Phi(x) = h(1) - h(0) = \int_0^1 h'(t)\, dt = \int_0^1 \langle \nabla\Phi(x + t(y - x)),\, y - x\rangle\, dt.$$

Add and subtract $\nabla\Phi(x)$ inside the integrand:

$$\Phi(y) - \Phi(x) = \int_0^1 \langle \nabla\Phi(x),\, y - x\rangle\, dt + \int_0^1 \langle \nabla\Phi(x + t(y - x)) - \nabla\Phi(x),\, y - x\rangle\, dt.$$

The first integral equals $\langle \nabla\Phi(x), y - x\rangle$. For the second, apply Cauchy–Schwarz and L-Lipschitzness of the gradient:

$$\begin{aligned} |\langle \nabla\Phi(x + t(y - x)) - \nabla\Phi(x),\, y - x\rangle| &\leq \|\nabla\Phi(x + t(y - x)) - \nabla\Phi(x)\|\, \|y - x\| \\ &\leq L\, \|t(y - x)\|\, \|y - x\| \\ &= L\, t\, \|y - x\|^2. \end{aligned}$$

Therefore the second integral is bounded below by the negative of its absolute value:

$$\int_0^1 \langle \nabla\Phi(x + t(y - x)) - \nabla\Phi(x),\, y - x\rangle\, dt \geq -\int_0^1 L\, t\, \|y - x\|^2\, dt = -\frac{L}{2}\|y - x\|^2.$$

Combining the identities yields

$$\Phi(y) - \Phi(x) \geq \langle \nabla\Phi(x), y - x\rangle - \frac{L}{2}\|y - x\|^2,$$

which is the claimed inequality. □

Apply Lemma H.3 with $x = x_t$ and $y = x_{t+1}$:

$$\Phi(x_{t+1}) \geq \Phi(x_t) + \langle g_t, \Delta_t\rangle - \frac{L}{2}\|\Delta_t\|^2.$$

It remains to lower bound $\langle g_t, \Delta_t\rangle$ in terms of $\|\Delta_t\|^2$ using the projection step. Let

$$y_t := x_t + \eta g_t, \qquad x_{t+1} = \mathrm{Proj}_{\mathcal{X}}(y_t).$$

By Lemma F.1 (projection variational inequality), for any $z \in \mathcal{X}$,

$$\langle y_t - x_{t+1},\, z - x_{t+1}\rangle \leq 0.$$

Choose $z = x_t \in \mathcal{X}$. Then

$$\langle y_t - x_{t+1},\, x_t - x_{t+1}\rangle \leq 0.$$

Substitute $y_t = x_t + \eta g_t$ and $\Delta_t = x_{t+1} - x_t$:

$$y_t - x_{t+1} = (x_t + \eta g_t) - x_{t+1} = -(x_{t+1} - x_t) + \eta g_t = -\Delta_t + \eta g_t, \qquad x_t - x_{t+1} = -(x_{t+1} - x_t) = -\Delta_t.$$

Hence

$$\langle -\Delta_t + \eta g_t,\, -\Delta_t\rangle \leq 0.$$

Expand the inner product using bilinearity:

$$\langle -\Delta_t,\, -\Delta_t\rangle + \eta\langle g_t,\, -\Delta_t\rangle \le 0 \quad \Longleftrightarrow \quad \|\Delta_t\|^2 - \eta\langle g_t, \Delta_t\rangle \le 0.$$

Rearranging yields the key inequality

$$\langle g_t, \Delta_t\rangle \ge \frac{1}{\eta}\|\Delta_t\|^2.$$

Substitute this into the smoothness lower bound:

$$\Phi(x_{t+1}) \ge \Phi(x_t) + \frac{1}{\eta}\|\Delta_t\|^2 - \frac{L}{2}\|\Delta_t\|^2 = \Phi(x_t) + \Big(\frac{1}{\eta} - \frac{L}{2}\Big)\|\Delta_t\|^2,$$

which proves the first inequality in the theorem.

Finally, if $\eta \le 1/L$, then $L \le 1/\eta$, hence

$$\frac{1}{\eta} - \frac{L}{2} \ge \frac{1}{\eta} - \frac{1}{2\eta} = \frac{1}{2\eta}.$$

Therefore

$$\Phi(x_{t+1}) \ge \Phi(x_t) + \frac{1}{2\eta}\,\|x_{t+1} - x_t\|^2,$$

which is the second inequality. □

H.16 Proof of Lemma 6.5

Proof. Fix any $x \in \mathcal{X}$. By definition,

$$\mathrm{Gap}(x) = \max_{u\in\mathcal{X}} \langle -\nabla\Phi(x) + R(x),\, x - u\rangle.$$

For an arbitrary $u \in \mathcal{X}$, expand the inner product by bilinearity:

$$\langle -\nabla\Phi(x) + R(x),\, x - u\rangle = \langle -\nabla\Phi(x),\, x - u\rangle + \langle R(x),\, x - u\rangle.$$

We upper bound the residual term using Cauchy–Schwarz:

$$\langle R(x),\, x - u\rangle \le \|R(x)\|\,\|x - u\|.$$

Since $\mathcal{X}$ has diameter $D := \max_{a,b\in\mathcal{X}} \|a - b\|$, in particular for this fixed $x \in \mathcal{X}$ and any $u \in \mathcal{X}$,

$$\|x - u\| \le D.$$

Therefore, for every $u \in \mathcal{X}$,

$$\langle R(x),\, x - u\rangle \le \|R(x)\|\, D.$$

Substitute this bound back into the expansion:

$$\langle -\nabla\Phi(x) + R(x),\, x - u\rangle \le \langle -\nabla\Phi(x),\, x - u\rangle + D\,\|R(x)\|.$$

Now take the maximum over $u \in \mathcal{X}$ on both sides. The right-hand side is a maximum of a function plus a constant (independent of u), so

$$\begin{aligned}\mathrm{Gap}(x) = \max_{u\in\mathcal{X}}\langle -\nabla\Phi(x) + R(x),\, x - u\rangle &\le \max_{u\in\mathcal{X}} \Big(\langle -\nabla\Phi(x),\, x - u\rangle + D\,\|R(x)\|\Big) \\ &= \max_{u\in\mathcal{X}}\langle -\nabla\Phi(x),\, x - u\rangle + D\,\|R(x)\| \\ &= \mathrm{Gap}_\Phi(x) + D\,\|R(x)\|.\end{aligned}$$

Finally, the alternative expression for $\mathrm{Gap}_\Phi(x)$ follows from pulling out the minus sign: for every $u \in \mathcal{X}$,

$$\langle -\nabla\Phi(x),\, x - u\rangle = \langle \nabla\Phi(x),\, u - x\rangle,$$

hence

$$\mathrm{Gap}_\Phi(x) = \max_{u\in\mathcal{X}}\langle -\nabla\Phi(x),\, x - u\rangle = \max_{u\in\mathcal{X}}\langle \nabla\Phi(x),\, u - x\rangle.$$

□

H.17 Proof of Lemma 6.6

Proof. Fix $x \in \mathcal{X}$ and set

$$x^+ := \mathrm{Proj}_{\mathcal{X}}\big(x + \eta\nabla\Phi(x)\big), \qquad \Delta := x^+ - x, \qquad \mathcal{G}_\eta(x) := \frac{\Delta}{\eta}.$$

For any $u \in \mathcal{X}$, the projection variational inequality in Lemma F.1 gives

$$\big\langle x + \eta\nabla\Phi(x) - x^+,\, u - x^+ \big\rangle \le 0,$$

and hence

$$\eta\langle\nabla\Phi(x), u - x^+\rangle \le \langle\Delta, u - x^+\rangle.$$

Adding $\eta\langle\nabla\Phi(x), \Delta\rangle$ to both sides yields

$$\eta\langle\nabla\Phi(x), u - x\rangle \le \langle\Delta, u - x^+\rangle + \eta\langle\nabla\Phi(x), \Delta\rangle.$$

By Cauchy–Schwarz, the diameter bound, and Assumption 6.1,

$$\langle\Delta, u - x^+\rangle \le D\|\Delta\|, \qquad \eta\langle\nabla\Phi(x), \Delta\rangle \le \eta G\|\Delta\|.$$

Therefore, for every $u \in \mathcal{X}$,

$$\langle\nabla\Phi(x), u - x\rangle \le (D + \eta G)\frac{\|\Delta\|}{\eta} = (D + \eta G)\|\mathcal{G}_\eta(x)\|.$$

Taking the maximum over $u \in \mathcal{X}$ proves the claim. □

H.18 Proof of Theorem 6.7

Proof. Fix any $x \in \mathcal{X}$. By Assumption 6.2, the VI operator decomposes as $F(x) = -\nabla\Phi(x) + R(x)$ with $\|R(x)\| \le \varepsilon$. Applying Lemma 6.5 gives

$$\mathrm{Gap}(x) = \max_{u\in\mathcal{X}}\langle -\nabla\Phi(x) + R(x),\, x - u\rangle \le \mathrm{Gap}_\Phi(x) + D\|R(x)\| \le \mathrm{Gap}_\Phi(x) + D\varepsilon.$$

Now apply Lemma 6.6 to bound Gap_Φ in terms of the projected-step mapping $\mathcal{G}_\eta(x) = \frac{1}{\eta}(x^+ - x)$ with $x^+ = \mathrm{Proj}_{\mathcal{X}}(x + \eta\nabla\Phi(x))$:

$$\mathrm{Gap}_\Phi(x) \le (D + \eta G)\,\|\mathcal{G}_\eta(x)\|.$$

Combining the two inequalities yields, for every $x \in \mathcal{X}$,

$$\mathrm{Gap}(x) \le (D + \eta G)\,\|\mathcal{G}_\eta(x)\| + D\varepsilon.$$

We now specialize to the iterates $x_{t+1} = \mathrm{Proj}_{\mathcal{X}}(x_t + \eta\nabla\Phi(x_t))$ with $\eta \le 1/L$. For each t, define $x_t^+ := x_{t+1}$ and $\Delta_t := x_{t+1} - x_t$, so that

$$\mathcal{G}_\eta(x_t) = \frac{1}{\eta}(x_t^+ - x_t) = \frac{1}{\eta}\Delta_t, \qquad \text{and hence} \qquad \|\Delta_t\|^2 = \eta^2\|\mathcal{G}_\eta(x_t)\|^2.$$

By Theorem 6.4, for every t,

$$\Phi(x_{t+1}) \ge \Phi(x_t) + \frac{1}{2\eta}\,\|x_{t+1} - x_t\|^2 = \Phi(x_t) + \frac{1}{2\eta}\,\|\Delta_t\|^2 = \Phi(x_t) + \frac{\eta}{2}\,\|\mathcal{G}_\eta(x_t)\|^2.$$

Rearranging gives

$$\frac{\eta}{2}\,\|\mathcal{G}_\eta(x_t)\|^2 \le \Phi(x_{t+1}) - \Phi(x_t).$$

Sum this inequality from $t = 0$ to $T - 1$ and use telescoping:

$$\frac{\eta}{2}\sum_{t=0}^{T-1}\|\mathcal{G}_\eta(x_t)\|^2 \le \sum_{t=0}^{T-1}\big(\Phi(x_{t+1}) - \Phi(x_t)\big) = \Phi(x_T) - \Phi(x_0).$$

By definition of $\Phi_{\max}$ and $\Phi_{\min}$ and the fact that all iterates lie in $\mathcal{X}$ (since $\mathrm{Proj}_{\mathcal{X}}(\cdot) \in \mathcal{X}$),

$$\Phi(x_T) \le \Phi_{\max}, \qquad \Phi(x_0) \ge \Phi_{\min}, \qquad \text{so} \qquad \Phi(x_T) - \Phi(x_0) \le \Phi_{\max} - \Phi_{\min}.$$

Therefore

$$\frac{\eta}{2}\sum_{t=0}^{T-1}\|\mathcal{G}_\eta(x_t)\|^2 \le \Phi_{\max}-\Phi_{\min}, \qquad \text{equivalently} \qquad \sum_{t=0}^{T-1}\|\mathcal{G}_\eta(x_t)\|^2 \le \frac{2(\Phi_{\max}-\Phi_{\min})}{\eta}.$$

Divide both sides by T:

$$\frac{1}{T}\sum_{t=0}^{T-1}\|\mathcal{G}_\eta(x_t)\|^2 \le \frac{2(\Phi_{\max}-\Phi_{\min})}{T\,\eta}.$$

Since $\min_{0\le t\le T-1}\|\mathcal{G}_\eta(x_t)\|^2 \le \frac{1}{T}\sum_{t=0}^{T-1}\|\mathcal{G}_\eta(x_t)\|^2$, we obtain

$$\min_{0\le t\le T-1}\|\mathcal{G}_\eta(x_t)\|^2 \le \frac{2(\Phi_{\max}-\Phi_{\min})}{T\,\eta},$$

and taking square roots (all terms are nonnegative) yields

$$\min_{0\le t\le T-1}\|\mathcal{G}_\eta(x_t)\| \le \sqrt{\frac{2(\Phi_{\max}-\Phi_{\min})}{T\,\eta}}.$$

Let $\hat{t}\in\arg\min_{0\le t\le T-1}\|\mathcal{G}_\eta(x_t)\|$ and $\hat{x}_T := x_{\hat{t}}$. Then

$$\|\mathcal{G}_\eta(\hat{x}_T)\| = \|\mathcal{G}_\eta(x_{\hat{t}})\| = \min_{0\le t\le T-1}\|\mathcal{G}_\eta(x_t)\| \le \sqrt{\frac{2(\Phi_{\max}-\Phi_{\min})}{T\,\eta}}.$$

Apply the pointwise gap bound $\mathrm{Gap}(x)\le (D+\eta G)\|\mathcal{G}_\eta(x)\| + D\varepsilon$ at $x=\hat{x}_T$ to get

$$\mathrm{Gap}(\hat{x}_T) \le (D+\eta G)\,\|\mathcal{G}_\eta(\hat{x}_T)\| + D\varepsilon \le (D+\eta G)\sqrt{\frac{2(\Phi_{\max}-\Phi_{\min})}{T\,\eta}} + D\varepsilon,$$

which is the claimed best-iterate guarantee. □

H.19 Proof of Theorem F.3

Proof. Fix any deterministic $x\in\mathcal{X}$. Using the decomposition $F(x) = -\nabla\Phi(x) + R(x)$ and applying Lemma 6.5, we have

$$\mathrm{Gap}(x) = \max_{u\in\mathcal{X}}\langle -\nabla\Phi(x)+R(x),\, x-u\rangle \le \mathrm{Gap}_\Phi(x) + D\|R(x)\|.$$

By Lemma 6.6, with $x^+ = \mathrm{Proj}_{\mathcal{X}}(x+\eta\nabla\Phi(x))$ and $\mathcal{G}_\eta(x) = \frac{1}{\eta}(x^+-x)$,

$$\mathrm{Gap}_\Phi(x) \le (D+\eta G)\,\|\mathcal{G}_\eta(x)\|.$$

Combining the two bounds yields the pointwise inequality valid for every $x\in\mathcal{X}$:

$$\mathrm{Gap}(x) \le (D+\eta G)\,\|\mathcal{G}_\eta(x)\| + D\,\|R(x)\|.$$

Now specialize to the iterates generated by the algorithm and the random output $\tilde{x}_T := x_{\tilde{t}}$, where $\tilde{t}$ is uniform on $\{0,1,\dots,T-1\}$ and independent of the iterates. Taking expectations gives

$$\mathbb{E}[\mathrm{Gap}(\tilde{x}_T)] \le (D+\eta G)\,\mathbb{E}\big[\|\mathcal{G}_\eta(\tilde{x}_T)\|\big] + D\,\mathbb{E}\big[\|R(\tilde{x}_T)\|\big].$$

We bound the residual term by Cauchy–Schwarz:

$$\mathbb{E}\big[\|R(\tilde{x}_T)\|\big] \le \Big(\mathbb{E}\big[\|R(\tilde{x}_T)\|^2\big]\Big)^{1/2} = E_T.$$

Next bound $\mathbb{E}\big[\|\mathcal{G}_\eta(\tilde{x}_T)\|\big]$ using Jensen/Cauchy–Schwarz:

$$\mathbb{E}\big[\|\mathcal{G}_\eta(\tilde{x}_T)\|\big] \le \Big(\mathbb{E}\big[\|\mathcal{G}_\eta(\tilde{x}_T)\|^2\big]\Big)^{1/2}.$$

Because $\tilde{t}$ is uniform and independent, conditioning on the iterates yields

$$\mathbb{E}\big[\|\mathcal{G}_\eta(\tilde{x}_T)\|^2 \,\big|\, x_0,\dots,x_{T-1}\big] = \frac{1}{T}\sum_{t=0}^{T-1}\|\mathcal{G}_\eta(x_t)\|^2,$$

and taking expectation over the iterates gives

$$\mathbb{E}\big[\|\mathcal{G}_\eta(\tilde{x}_T)\|^2\big] = \mathbb{E}\left[\frac{1}{T}\sum_{t=0}^{T-1}\|\mathcal{G}_\eta(x_t)\|^2\right] = \frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\big[\|\mathcal{G}_\eta(x_t)\|^2\big].$$

For the present (exact, deterministic) projected ascent dynamics, the iterates are deterministic given x_0, so the expectations may be dropped; the same bound below holds verbatim even if randomness is present.

By Theorem 6.4, for each t,

$$\Phi(x_{t+1}) \geq \Phi(x_t) + \frac{\eta}{2}\|\mathcal{G}_\eta(x_t)\|^2,$$

equivalently,

$$\frac{\eta}{2}\|\mathcal{G}_\eta(x_t)\|^2 \leq \Phi(x_{t+1}) - \Phi(x_t).$$

Summing from $t = 0$ to $T - 1$ telescopes:

$$\frac{\eta}{2}\sum_{t=0}^{T-1}\|\mathcal{G}_\eta(x_t)\|^2 \leq \sum_{t=0}^{T-1}\big(\Phi(x_{t+1}) - \Phi(x_t)\big) = \Phi(x_T) - \Phi(x_0).$$

Since all iterates lie in $\mathcal{X}$, we have $\Phi(x_T) \leq \Phi_{\max}$ and $\Phi(x_0) \geq \Phi_{\min}$, hence

$$\Phi(x_T) - \Phi(x_0) \leq \Phi_{\max} - \Phi_{\min}.$$

Therefore

$$\sum_{t=0}^{T-1}\|\mathcal{G}_\eta(x_t)\|^2 \leq \frac{2(\Phi_{\max} - \Phi_{\min})}{\eta}, \qquad \text{and thus} \qquad \frac{1}{T}\sum_{t=0}^{T-1}\|\mathcal{G}_\eta(x_t)\|^2 \leq \frac{2(\Phi_{\max} - \Phi_{\min})}{T\,\eta}.$$

Using $\mathbb{E}\big[\|\mathcal{G}_\eta(\tilde{x}_T)\|^2\big] = \frac{1}{T}\sum_{t=0}^{T-1}\|\mathcal{G}_\eta(x_t)\|^2$, we obtain

$$\mathbb{E}\big[\|\mathcal{G}_\eta(\tilde{x}_T)\|^2\big] \leq \frac{2(\Phi_{\max} - \Phi_{\min})}{T\,\eta},$$

and hence

$$\mathbb{E}\big[\|\mathcal{G}_\eta(\tilde{x}_T)\|\big] \leq \Big(\mathbb{E}\big[\|\mathcal{G}_\eta(\tilde{x}_T)\|^2\big]\Big)^{1/2} \leq \sqrt{\frac{2(\Phi_{\max} - \Phi_{\min})}{T\,\eta}}.$$

Substituting the two expectation bounds back into the gap inequality yields

$$\begin{aligned}\mathbb{E}[\mathrm{Gap}(\tilde{x}_T)] &\leq (D + \eta G)\,\mathbb{E}\big[\|\mathcal{G}_\eta(\tilde{x}_T)\|\big] + D\,\mathbb{E}\big[\|R(\tilde{x}_T)\|\big] \\ &\leq (D + \eta G)\sqrt{\frac{2(\Phi_{\max} - \Phi_{\min})}{T\,\eta}} + D\,E_T,\end{aligned}$$

which is the desired bound. □

H.20 Proof of Theorem 6.8

Lemma H.4 (Potential-gap control by an inexact projected-step mapping)**.** *Fix any $x \in \mathcal{X}$ and any direction $g \in \mathbb{R}^d$. Define*

$$x^+ := \mathrm{Proj}_{\mathcal{X}}(x + \eta g), \qquad \widetilde{\mathcal{G}}_\eta(x) := \frac{1}{\eta}(x^+ - x).$$

Then, for every $u \in \mathcal{X}$,

$$\langle g,\, u - x\rangle \;\leq\; (D + \eta\|g\|)\,\|\widetilde{\mathcal{G}}_\eta(x)\|,$$

and consequently

$$\mathrm{Gap}_\Phi(x) = \max_{u\in\mathcal{X}}\langle\nabla\Phi(x),\, u - x\rangle \;\leq\; (D + \eta\|g\|)\,\|\widetilde{\mathcal{G}}_\eta(x)\| + D\,\|\nabla\Phi(x) - g\|.$$

Proof. Fix $u \in \mathcal{X}$ and write

$$\langle g,\, u - x\rangle = \langle g,\, u - x^+\rangle + \langle g,\, x^+ - x\rangle.$$

We bound the two terms separately.

Because $x^+ = \mathrm{Proj}_{\mathcal{X}}(x + \eta g)$, Lemma F.1 (the VI inequality) with $y = x + \eta g$ and $z = u \in \mathcal{X}$ gives

$$\langle (x + \eta g) - x^+,\, u - x^+\rangle \le 0.$$

Expanding and rearranging,

$$\eta\,\langle g,\, u - x^+\rangle \le \langle x^+ - x,\, u - x^+\rangle.$$

Apply Cauchy–Schwarz and the diameter bound $\|u - x^+\| \le D$:

$$\langle g,\, u - x^+\rangle \le \frac{1}{\eta}\|x^+ - x\|\,\|u - x^+\| \le \frac{D}{\eta}\|x^+ - x\| = D\,\|\widetilde{\mathcal{G}}_\eta(x)\|.$$

For the second term,

$$\langle g,\, x^+ - x\rangle \le \|g\|\,\|x^+ - x\| = \eta\|g\|\,\|\widetilde{\mathcal{G}}_\eta(x)\|.$$

Adding the two bounds yields

$$\langle g,\, u - x\rangle \le \big(D + \eta\|g\|\big)\,\|\widetilde{\mathcal{G}}_\eta(x)\|.$$

Taking $\max_{u\in\mathcal{X}}$ on the left gives

$$\max_{u\in\mathcal{X}}\langle g,\, u - x\rangle \le (D + \eta\|g\|)\,\|\widetilde{\mathcal{G}}_\eta(x)\|.$$

Now decompose $\nabla\Phi(x) = g + \big(\nabla\Phi(x) - g\big)$ and use $\|u - x\| \le D$:

$$\langle \nabla\Phi(x),\, u{-}x\rangle = \langle g,\, u{-}x\rangle + \langle \nabla\Phi(x){-}g,\, u{-}x\rangle \le \langle g,\, u{-}x\rangle + \|\nabla\Phi(x){-}g\|\,\|u{-}x\| \le \langle g,\, u{-}x\rangle + D\|\nabla\Phi(x){-}g\|.$$

Taking $\max_{u\in\mathcal{X}}$ and substituting the previous bound yields the claimed inequality for $\mathrm{Gap}_\Phi(x)$. □

Lemma H.5 (One-step Lyapunov lower bound under direction error)**.** *Under Assumption 6.1, let $x^+ = \mathrm{Proj}_{\mathcal{X}}(x + \eta g)$ with $\eta \le 1/L$. Assume $\|g - \nabla\Phi(x)\| \le \delta$. Then*

$$\Phi(x^+) \;\ge\; \Phi(x) + \frac{1}{4\eta}\,\|x^+ - x\|^2 - \eta\,\delta^2.$$

Equivalently, with $\widetilde{\mathcal{G}}_\eta(x) = \frac{1}{\eta}(x^+ - x)$,

$$\Phi(x^+) \;\ge\; \Phi(x) + \frac{\eta}{4}\,\|\widetilde{\mathcal{G}}_\eta(x)\|^2 - \eta\,\delta^2.$$

Proof. Let $\Delta := x^+ - x$. L-smoothness of Φ implies the standard lower bound

$$\Phi(x^+) \ge \Phi(x) + \langle \nabla\Phi(x),\, \Delta\rangle - \frac{L}{2}\|\Delta\|^2.$$

Write $\nabla\Phi(x) = g - (g - \nabla\Phi(x))$ and denote the direction error $e := g - \nabla\Phi(x)$, so $\|e\| \le \delta$ and

$$\langle \nabla\Phi(x),\, \Delta\rangle = \langle g,\, \Delta\rangle - \langle e,\, \Delta\rangle.$$

Because $x^+ = \mathrm{Proj}_{\mathcal{X}}(x + \eta g)$, Lemma F.1 with $y = x + \eta g$ and $z = x \in \mathcal{X}$ gives

$$\langle (x + \eta g) - x^+,\, x - x^+\rangle \le 0.$$

Since $x - x^+ = -\Delta$, this becomes

$$\langle \eta g - \Delta,\, -\Delta\rangle \le 0 \qquad\Longleftrightarrow\qquad \eta\,\langle g,\, \Delta\rangle \ge \|\Delta\|^2 \qquad\Longleftrightarrow\qquad \langle g,\, \Delta\rangle \ge \frac{1}{\eta}\|\Delta\|^2.$$

For the error term, Cauchy–Schwarz gives $|\langle e, \Delta\rangle| \le \|e\|\,\|\Delta\| \le \delta\|\Delta\|$. Apply Young's inequality in the form

$$\delta\|\Delta\| \le \frac{1}{4\eta}\|\Delta\|^2 + \eta\,\delta^2,$$

which holds because $2ab \le \alpha a^2 + \alpha^{-1} b^2$ with $a = \|\Delta\|$, $b = \delta$, and $\alpha = \frac{1}{2\eta}$. Therefore,

$$-\langle e,\, \Delta\rangle \ge -\delta\|\Delta\| \ge -\frac{1}{4\eta}\|\Delta\|^2 - \eta\,\delta^2.$$

Combine the bounds:

$$\langle \nabla\Phi(x),\, \Delta\rangle = \langle g,\, \Delta\rangle - \langle e,\, \Delta\rangle \ge \frac{1}{\eta}\|\Delta\|^2 - \frac{1}{4\eta}\|\Delta\|^2 - \eta\,\delta^2 = \frac{3}{4\eta}\|\Delta\|^2 - \eta\,\delta^2.$$

Substitute into the smoothness lower bound:

$$\Phi(x^+) \ge \Phi(x) + \Big(\frac{3}{4\eta} - \frac{L}{2}\Big)\|\Delta\|^2 - \eta\,\delta^2.$$

Using $\eta \le 1/L$ implies $\frac{3}{4\eta} - \frac{L}{2} \ge \frac{3}{4\eta} - \frac{1}{2\eta} = \frac{1}{4\eta}$, hence

$$\Phi(x^+) \ge \Phi(x) + \frac{1}{4\eta}\|\Delta\|^2 - \eta\,\delta^2,$$

which is the claim. Replacing $\Delta = \eta\,\widetilde{\mathcal{G}}_\eta(x)$ yields the equivalent form. □

Proof of Theorem 6.8. Let the iterates satisfy $x_{t+1} = \mathrm{Proj}_{\mathcal{X}}(x_t + \eta g_t)$ and define

$$\widetilde{\mathcal{G}}_\eta(x_t) := \frac{1}{\eta}(x_{t+1} - x_t).$$

Fix any t and apply Lemma 6.5:

$$\mathrm{Gap}(x_t) \le \mathrm{Gap}_\Phi(x_t) + D\|R(x_t)\|.$$

By Assumption 6.2, $\|R(x_t)\| \le \varepsilon$, hence

$$\mathrm{Gap}(x_t) \le \mathrm{Gap}_\Phi(x_t) + D\varepsilon.$$

Now apply Lemma F.5 with $x = x_t$ and $g = g_t$:

$$\mathrm{Gap}_\Phi(x_t) \le \big(D + \eta\|g_t\|\big)\,\|\widetilde{\mathcal{G}}_\eta(x_t)\| + D\,\|\nabla\Phi(x_t) - g_t\|.$$

By Assumption 6.3, $\|\nabla\Phi(x_t) - g_t\| \le \delta_t \le \bar{\delta}$. Also $\|g_t\| \le \|\nabla\Phi(x_t)\| + \|g_t - \nabla\Phi(x_t)\| \le G + \bar{\delta}$ by Assumption 6.1. Therefore,

$$\mathrm{Gap}_\Phi(x_t) \le \big(D + \eta(G + \bar{\delta})\big)\,\|\widetilde{\mathcal{G}}_\eta(x_t)\| + D\,\bar{\delta},$$

and hence

$$\mathrm{Gap}(x_t) \le \big(D + \eta(G + \bar{\delta})\big)\,\|\widetilde{\mathcal{G}}_\eta(x_t)\| + D(\varepsilon + \bar{\delta}).$$

Now choose $\hat{t} \in \arg\min_{0\le t\le T-1}\|\widetilde{\mathcal{G}}_\eta(x_t)\|$ and set $\hat{x}_T = x_{\hat{t}}$. Then

$$\mathrm{Gap}(\hat{x}_T) \le \big(D + \eta(G + \bar{\delta})\big)\min_{0\le t\le T-1}\|\widetilde{\mathcal{G}}_\eta(x_t)\| + D(\varepsilon + \bar{\delta}).$$

It remains to bound the best-iterate stationarity measure. Apply Lemma H.5 to each step with $x = x_t$, $x^+ = x_{t+1}$, and $\delta = \delta_t$:

$$\Phi(x_{t+1}) \ge \Phi(x_t) + \frac{\eta}{4}\,\|\widetilde{\mathcal{G}}_\eta(x_t)\|^2 - \eta\,\delta_t^2.$$

Rearrange and sum from $t = 0$ to $T - 1$:

$$\frac{\eta}{4}\sum_{t=0}^{T-1}\|\widetilde{\mathcal{G}}_\eta(x_t)\|^2 \le \Phi(x_T) - \Phi(x_0) + \eta\sum_{t=0}^{T-1}\delta_t^2.$$

Use $\Phi(x_T) - \Phi(x_0) \le \Phi_{\max} - \Phi_{\min}$ and $\delta_t \le \bar{\delta}$:

$$\frac{\eta}{4}\sum_{t=0}^{T-1}\|\widetilde{\mathcal{G}}_\eta(x_t)\|^2 \le (\Phi_{\max} - \Phi_{\min}) + \eta\,T\,\bar{\delta}^2.$$

Divide by T and use $\min_t a_t^2 \le \frac{1}{T}\sum_t a_t^2$:

$$\min_{0\le t\le T-1} \|\widetilde{\mathcal{G}}_\eta(x_t)\|^2 \le \frac{1}{T}\sum_{t=0}^{T-1} \|\widetilde{\mathcal{G}}_\eta(x_t)\|^2 \le \frac{4(\Phi_{\max}-\Phi_{\min})}{T\,\eta} + 4\,\bar{\delta}^2.$$

Taking square-roots and substituting into the earlier gap bound yields the explicit estimate

$$\mathrm{Gap}(\hat{x}_T) \le \big(D + \eta(G+\bar{\delta})\big)\sqrt{\frac{4(\Phi_{\max}-\Phi_{\min})}{T\,\eta} + 4\bar{\delta}^2} + D(\varepsilon+\bar{\delta}).$$

□