Gauge-Symmetric Dual Lagrangian Frameworks for Born-Oppenheimer Molecular Dynamics

Sungwoo Park*
Korea University
sungwoo_park@korea.ac.kr

Jongwon Lee
Korea University
disora@korea.ac.kr

Hyung-sik Yoon
Korea University
yhs915@korea.ac.kr

Jiwoong Kim
Korea University
joshuasmile@korea.ac.kr

Abstract

Born-Oppenheimer molecular dynamics delivers quantum accurate trajectories by evolving nuclei on the electronic ground-state surface while retaining electronic observables such as dipoles and polarizabilities. In Kohn-Sham density functional theory, this typically relies on an iterative self consistent field solver at each time step. Machine learning can predict electronic quantities or improve SCF initialization, yet dynamics remain sensitive to residual self consistency error, which can yield non conservative forces and energy drift, and repeated diagonalization limits differentiable and accelerator efficient implementations. We propose a gauge-symmetric dual Lagrangian framework that avoids SCF loops by propagating the physically meaningful electronic state as the occupied subspace and its density projector on a Grassmann manifold. A residual driven update stabilized by Rayleigh type dissipation with time decaying inertia yields a closed evolution of the subspace, its conjugate momentum, and the projector. Coupled with a neural gauge-symmetric Hamiltonian with analytic coordinate derivatives, the method provides closed form forces with Pulay corrections and enables stable quantum BOMD on molecular simulation benchmarks.

1 Introduction

Born-Oppenheimer molecular dynamics (BOMD) stands as a cornerstone of *ab initio quantum simulation*, providing predictive trajectories by evaluating nuclear forces on the instantaneous electronic ground-state surface. Beyond its role as a workhorse for ground-state nuclear dynamics, BOMD is widely used across chemistry and biochemistry (Marx and Hutter, 2009; Iftimie et al., 2005), physics and materials science (Marx and Hutter, 2009; Hutter et al., 2024) to uncover reaction mechanisms (Iftimie et al., 2005) condensed-phase structure (Marx and Hutter, 2009; Hutter et al., 2024) and phase transitions under realistic conditions (Marx and Hutter, 2009; Hutter et al., 2024). Unlike methods that rely solely on potential energy surfaces, the enduring value of BOMD lies in its ability to preserve access to the full electronic state, enabling the calculation of dipoles, polarizabilities, and other electronic observables essential for downstream analysis. However, this accuracy comes at the steep computational cost of solving the electronic structure problem at every time step.

Recent advances in machine learning have sought to reduce this burden by accelerating the core steps. For instance, one prominent direction involves machine-learned interatomic potentials (MLIPs) that bypass quantum calculations entirely via fast surrogate models (Behler and Parrinello, 2007; Schütt et al., 2017; Zhang et al., 2018; Batzner et al., 2022; Unke et al., 2021b). While efficient,

*Corresponding author.

these approaches often discard the underlying electronic structure, limiting their utility in quantum-mechanical workflows. A complementary strategy learns the mapping from nuclear geometry directly to the Hamiltonian matrix, thereby accelerating the evaluation of electronic properties (Zhong et al., 2023; Yu et al., 2023; Gu et al., 2024; Tang et al., 2024; Zhong et al., 2024; Zhang et al., 2024; Shakiba and Akimov, 2024; Qian et al., 2026). Nevertheless, these advances still depend on the so-called *self-consistent field* (SCF) procedure, a computationally intensive iterative algorithm, and therefore mainly serve as auxiliary components for improving initial guesses or preconditioning iterative solvers rather than removing the SCF bottleneck itself (Brockherde et al., 2017). Due to this dependence, existing approaches may exhibit unstable energy drift and can degrade severely, or even fail altogether, in regions where the SCF procedure struggles or fails to converge. Moreover, the repeated and non-differentiable nature of SCF makes it difficult to construct a fully differentiable, end-to-end trainable pipeline. These limitations suggest that the SCF bottleneck should be tackled at a more structural level, by reformulating the electronic update itself rather than merely improving solver-side acceleration. Motivated by this perspective, the goal of this work is to uncover such a fundamental structure for electronic updates, so as to simultaneously achieve computational acceleration and mitigate the instability induced by SCF iterations.

In Kohn–Sham theory, the physical electron distribution is invariant under gauge transformations among the occupied orbitals (Parr and Yang, 1989). Such gauge symmetry can be naturally encoded in the Riemannian geometry of the Grassmann manifold, which represents the physical electronic state as an occupied subspace. This geometric viewpoint opens the door to describing electronic dynamics directly on the manifold, and hence to reformulating the standard BOMD framework from a more fundamental perspective. Building on this geometric viewpoint, we propose a *gauge-symmetric dual Lagrangian* method that defines a new electronic dynamics directly on the geometry of the occupied subspace to remove the need for repeated SCF solves. Rather than enforcing self-consistency through iterative convergence, we introduce a Lagrangian within this Riemannian geometric setting and derive a stable relaxation dynamics that drives the electronic state toward the instantaneous ground state. The resulting dynamics evolves the occupied subspace together with a momentum-like auxiliary variable and the density projector, while preserving the geometric constraints and gauge symmetry of the electronic state. We further combine this dynamics with a neural Hamiltonian whose coordinate derivatives are available in closed form, yielding an SCF-free end-to-end learning pipeline. In summary, the main contributions of this paper are two-fold:

1. **Gauge-symmetric Lagrangian Dynamics for SCF-Free Kohn-Sham DFT.** We develop a theoretical formulation that recasts self-consistency as a gauge-symmetric, geometry-aware relaxation of the physically meaningful electronic state. From the dual Lagrangian perspective, conventional SCF iterations are replaced by a stable energy-decreasing propagation rule, yielding a principled SCF-free framework for tracking the adiabatic electronic ground state during molecular dynamics.

2. **Differentiable and Learning-Ready SCF-Free BOMD.** We build an end-to-end machine learning framework by coupling the SCF-free Kohn-Sham dynamics to a learnable Hamiltonian with analytic force evaluation, so both training and inference avoid differentiating through SCF solvers and repeated diagonalization. Numerically, the resulting simulator runs with zero SCF iterations per MD step, versus approximately 7 to 26 iterations in SCF-based pipelines, yielding fixed-cost electronic updates, more stable simulation, and an approximate $\mathbf{10^2}\times$ speedup in inference time while preserving physically consistent electronic data configurations.

2 Kohn-Sham DFT in Born-Oppenheimer Molecular Dynamics

To contextualize the proposed methodology, we begin by revisiting the fundamental principles of BOMD. Let $\mathbf{r} = (\mathbf{r}_1, \ldots, \mathbf{r}_{N_\mathbf{r}}) \in \mathbb{R}^{3N_\mathbf{r}}$ be nuclear coordinates with masses $\{\mathbf{m}_I\}_{I=1}^{N_\mathbf{r}}$, where $N_\mathbf{r}$ is the number of atoms. This standard framework describes the evolution of nuclei on the instantaneous electronic ground-state potential surface, with its governing dynamics formally defined as follows:

Definition 2.1 (Born-Oppenheimer Molecular Dynamics)**.** *Assume that $\mathbf{E}_{\mathrm{BO}}(\mathbf{r})$ is differentiable in $\mathbf{r}$. Then, Born-Oppenheimer molecular dynamics (BOMD) is the Newtonian dynamics of nuclei:*

$$\mathbf{m}_I \ddot{\mathbf{r}}_I(t) = -\nabla \mathbf{E}_{\mathrm{BO}}(\mathbf{r}(t)), \quad \mathbf{E}_{\mathrm{BO}}(\mathbf{r}) \coloneqq \mathbf{E}_{\mathrm{nn}}(\mathbf{r}) + \mathbf{E}^\star(\mathbf{r}), \quad I = 1, \ldots, N_\mathbf{r}. \tag{1}$$

where $\mathbf{E}_{\mathrm{nn}}$ is the nuclear repulsion and $\mathbf{E}^\star(\mathbf{r})$ is the electronic ground-state energy.

In practice, advancing the nuclei coordinates from $\mathbf{r}(t)$ to $\mathbf{r}(t+dt)$ requires evaluating the corresponding force $\nabla_{\mathbf{r}}\mathbf{E}^{\star}(\mathbf{r}(t))$ at the current geometry. In Kohn-Sham DFT, this electronic ground-state evaluation is posed as a constrained minimization over occupied orbitals and is typically carried out by an iterative *self-consistent field* (SCF) procedure at every time step. We now make this dependence explicit by introducing the standard Kohn-Sham formulation.

Kohn-Sham Density Functional Theory. We begin by recalling the standard variational Kohn-Sham problem expressed with nuclear geometry. For each $\mathbf{r}$, the *electronic ground-state energy* is defined by the constrained minimization of the energy functional $\mathcal{E}$:

$$\mathbf{E}^{\star}(\mathbf{r}) \coloneqq \min_{\mathbf{C}} \mathcal{E}(\mathbf{r},\mathbf{C}), \quad \mathbf{C}^{\dagger}\mathrm{S}(\mathbf{r})\mathbf{C} = \mathrm{I}_N, \quad \mathrm{S}_{\mu\nu}(\mathbf{r}) = \int_{\mathbb{R}^3} \chi_{\mu}(\mathbf{x};\mathbf{r})^{\dagger}\chi_{\nu}(\mathbf{x};\mathbf{r})d\mathbf{x}, \tag{2}$$

where $\{\chi_{\mu}\}_{\mu=1}^{M}$ denotes the atom-centered basis functions, M and N represent the number of *atomic* and *occupied* orbitals, respectively. To practically solve this minimization problem within a tractable framework, we adopt the *Kohn-Sham ansatz* (Cancès and Friesecke, 2023), which specifies the energy functional in (2) as:

$$\mathcal{E}(\mathbf{r},\mathbf{C}) = \mathrm{Tr}\big(\mathbf{H}(\mathbf{r})\mathbf{P}\big) + \frac{1}{2}\iint_{\mathbb{R}^3\times\mathbb{R}^3} \frac{\rho_{\mathbf{P}}(\mathbf{x})\rho_{\mathbf{P}}(\mathbf{y})}{\|\mathbf{x}-\mathbf{y}\|}d\mathbf{x}d\mathbf{y} + \mathbf{E}_{\mathrm{xc}}\big[\rho_{\mathbf{P}}\big], \quad \mathbf{P} = \mathbf{C}\mathbf{C}^{\dagger}\mathrm{S}(\mathbf{r}), \tag{3}$$

where $\mathbf{P}$ and $\mathbf{C}$ are called an *occupied-subspace matrix* and *orbital matrix*, respectively. In the expression, the second term represents the classical Coulomb repulsion between charge densities, while the last term accounts for the non-classical quantum many-body effects with the induced electron density given by $\rho_{\mathbf{P}}(\mathbf{x}) = \sum_{\mu,\nu=1}^{M} \mathbf{P}_{\mu\nu}\chi_{\mu}(\mathbf{x};\mathbf{r})\chi_{\nu}(\mathbf{x};\mathbf{r})^{\dagger}$. In the expression, the Hamiltonian matrix $\mathbf{H}(\mathbf{r}) \in \mathbb{C}^{M\times M}$ encapsulates the mean-field energetics, defined by projecting the kinetic energy and the external nuclear potential onto the basis set:

$$\mathbf{H}_{\mu\nu}(\mathbf{r}) = -\int_{\mathbb{R}^3} \chi_{\mu}(\mathbf{x};\mathbf{r})^{\dagger}\left[\frac{1}{2}\nabla^2 + \sum_{I=1}^{N_{\mathbf{r}}} \frac{\mathrm{Z}_I}{\|\mathbf{x}-\mathbf{r}_I\|}\right]\chi_{\nu}(\mathbf{x};\mathbf{r})d\mathbf{x}, \tag{4}$$

where Z_I denotes the nuclear charge of the I-th nucleus. In practice, the variational minimization of the energy functional subject to orthogonality constraints gives rise to the Roothaan-Hall equations, which must be solved numerically via the following iterative algorithm.

Self-Consistent Field. In SCF, minimizing the energy functional $\mathcal{E}$ with respect to the orbitals $\mathbf{C}$ subject to the orthonormality constraint yields the *Roothaan-Hall equations* (Helgaker et al., 2013):

$$\mathbf{F}_{\mu\nu}(\mathbf{D}) \coloneqq \frac{\partial \mathbf{E}}{\partial \mathbf{D}_{\nu\mu}} = \mathbf{H}_{\mu\nu}(\mathbf{r}) + \sum_{\kappa,\lambda}(\mu\nu|\kappa\lambda)\mathbf{D}_{\lambda\kappa} + \frac{\partial \mathbf{E}_{\mathrm{xc}}}{\partial \mathbf{D}_{\nu\mu}}, \quad \begin{cases}\mathbf{F}(\mathbf{D})\mathbf{C} = \mathrm{S}(\mathbf{r})\mathbf{C}\boldsymbol{\epsilon}, \\ \mathbf{D} \coloneqq \mathbf{P}\mathrm{S}(\mathbf{r})^{-1},\end{cases} \tag{5}$$

Here, $\boldsymbol{\epsilon}$ denotes the diagonal matrix of Lagrange orbital energies, and the effective Hamiltonian $\mathbf{F}(\mathbf{D})$ is termed the *Fock matrix*, with $(\mu\nu|\kappa\lambda)$ representing the two-electron repulsion integrals. Solving Eq. (5) yields a nonlinear generalized eigenvalue problem and is typically handled by an iterative SCF procedure that repeatedly constructs the Fock matrix and solves generalized eigenproblems, as illustrated in Fig. 1 (*left*), with a per-iteration cost that scales as $O(M^4)$ in standard implementations.

To alleviate the computational bottleneck, existing ML-based approaches aim to reduce SCF iterations by learning a surrogate Hamiltonian $\mathbf{H}_{\theta} \approx \mathbf{H}^{\star}$ for improved initialization of the solver. However, better initialization alone is insufficient for reliable BOMD simulations, as SCF instability allows even negligible inaccuracies in the reconstruction of $\mathbf{H}_{\theta}$ to propagate and induce chaotic behavior (Cancès et al., 2021). Consequently, the resulting self-consistency errors corrupt the force evaluation and accumulate into significant energy drift over BOMD trajectories. These limitations suggest that resolving the sensitivity of the SCF procedure requires structurally avoiding the feedback loop, rather than merely accelerating existing solvers.

In this work, we aim to overcome the aforementioned stability barriers by developing an **end-to-end SCF-free** scheme. Our approach reformulates the electronic evolution not as a series of chaotic fixed-point mappings in SCF solvers, but as a continuous-time geometric dynamics for the orbital matrix $t \mapsto \mathbf{C}_t$ and the density projector $t \mapsto \mathbf{P}_t$. This ensures that the resulting trajectory remains faithful to the Kohn–Sham variational principle in (3) by formulating a geometric Lagrangian framework that governs the joint evolution of the orbital matrix $t \mapsto \mathbf{C}_t$ and the density projector $t \mapsto \mathbf{P}_t$. In the following section, we make this approach concrete by replacing the iterative SCF loop with a single, differentiable update rule for $(\mathbf{C}_t, \mathbf{P}_t)$, so the electronic step becomes a stable module that can be trained in an end-to-end framework.

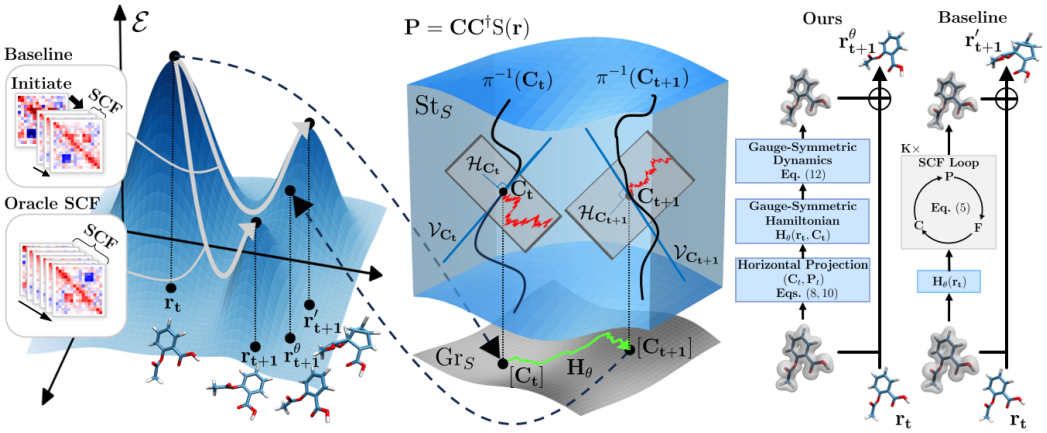


Figure 1: **Overall Pipeline.** *(left)* Conventional SCF-accelerated schemes infer an initial Hamiltonian $\mathbf{H}_\theta = \mathbf{H}_0$ to reduce the inner SCF solves, rapidly traversing the energy landscape and reaching metastable basins. *(middle)* We evolve the electronic state $\mathbf{C}_t \in \mathrm{St}_S$ on the principal bundle via a gauge-symmetric Lagrangian dynamics, providing an SCF-free detour for BOMD. *(right)* Existing methods advance nuclei by solving an expensive SCF problem at every step to obtain the ground-state electronic state at inference time. Our method bypasses this loop by propagating the electronic state via *gauge-symmetric dual Lagrangian dynamics*, updating the dual variables $(\mathbf{P}, \mathbf{C})$ through horizontal projection while evaluating $\mathbf{H}_\theta$ once per step, yielding an SCF-free scheme.

3 Gauge-Symmetric Dual Lagrangian of Molecular Dynamics

Motivated by the fragility of SCF solvers, we propose a self-consistent continuous dynamics that can avoid time-consuming generalized eigenvalue problems during BOMD simulation.

Riemannian Geometry of Electronic Configurations.[2] The electronic state in (2) is obtained by minimizing $\mathcal{E}$ under the orthonormality constraint, but the representation is known to be inherently non-unique. To be more specific, for any gauge transformation $\mathbf{U} \in \mathrm{U}(N)$, both $\mathbf{C}$ and $\mathbf{CU}$ satisfy the same constraint and span the same occupied subspace, thus yielding the same energy and physical observables. This invariance is known as **gauge symmetry** (Parr and Yang, 1989), implying that the physically meaningful electronic configuration is the *occupied subspace*, not a particular choice of orbital representatives. Accordingly, a proper formulation is required to (i) enforce orthonormality of orbital representatives and (ii) identify all gauge rotations as the same physical state.

We now establish the aforementioned geometric structure of the electronic configuration space through the interplay between the orthonormality constraint and gauge symmetry, which we formulate respectively on *Stiefel* and *Grassmann* manifolds:

$$\underbrace{\mathrm{St}_S \coloneqq \left\{\mathbf{C} \;\middle|\; \mathbf{C}^\dagger \mathrm{S}(\mathbf{r})\mathbf{C} = \mathrm{I}_N\right\}}_{\text{Subspace Orthonormality in (2)}}, \quad \underbrace{\mathrm{Gr}_S \coloneqq \mathrm{St}_S/\mathrm{U}(N) \cong \left\{[\mathbf{C}] \coloneqq \mathbf{P} \;\middle|\; \mathbf{P} = \mathbf{C}\mathbf{C}^\dagger \mathrm{S}(\mathbf{r})\right\}}_{\text{Gauge Symmetry: } \mathbf{C} \sim \mathbf{CU}}.$$

While $\mathbf{C} \in \mathrm{St_S}$ encodes orthonormality to retain gauge redundancy, $[\mathbf{C}] \coloneqq \mathbf{P} \in \mathrm{Gr_S}$ quotients out this symmetry to uniquely identify the physical state. Fig. 1 *(middle)* depicts the geometric relationship between $\mathbf{C}$ and $[\mathbf{C}]$, highlighting the structure of the configuration space. To endow this configuration space with a metric structure suitable for dynamics, we equip it with the Riemannian metric $\langle \mathbf{X}, \mathbf{Y} \rangle_{\mathrm{S}} \coloneqq \mathrm{Re}\,\mathrm{Tr}(\mathbf{X}^\dagger \mathrm{S}(\mathbf{r})\mathbf{Y})$, admitting the standard characterization of the tangent space:

$$T_{\mathbf{C}}\mathrm{St_S}(M, N) = \left\{\Delta \in \mathbb{C}^{M \times N} : \mathbf{C}^\dagger \mathrm{S}\Delta + \Delta^\dagger \mathrm{S}\mathbf{C} = 0\right\}. \tag{6}$$

Then, the right action of the unitary Lie group $\mathrm{U}(N)$, $\mathcal{R}_{\mathbf{U}}(\mathbf{C}) \coloneqq \mathbf{CU}$ endows $\mathrm{St_S}$ with the structure of a principal $\mathrm{U}(N)$-bundle over the Grassmann quotient with canonical projection $\pi : \mathrm{St_S} \to \mathrm{Gr_S}$, $\pi(\mathbf{C}) = [\mathbf{C}]$. The metric also induces a canonical principal connection specified by the $\mathfrak{u}(N)$-valued connection 1-form $\omega_{\mathbf{C}} : T_{\mathbf{C}}\mathrm{St_S} \to \mathfrak{u}(N)$, where $\omega_{\mathbf{C}}(\Delta) \coloneqq \mathbf{C}^\dagger \mathrm{S}(\mathbf{r})\Delta$ and $\mathfrak{u}(N) = \{\Omega \in \mathbb{C}^{N \times N} : \Omega^\dagger = -\Omega\}$ denotes the Lie algebra of $\mathrm{U}(N)$. With respect to established metric and connection, the tangent space splits orthogonally as follows:

$$T_{\mathbf{C}}\mathrm{St_S} = \mathcal{V}_{\mathbf{C}} \oplus \mathcal{H}_{\mathbf{C}},\ \mathcal{V}_{\mathbf{C}} \coloneqq \left\{\mathbf{C}\Omega : \Omega \in \mathfrak{u}(N)\right\},\ \mathcal{H}_{\mathbf{C}} \coloneqq \left\{\Delta \in T_{\mathbf{C}}\mathrm{St_S} : \mathbf{C}^\dagger \mathrm{S}\Delta = 0\right\}. \tag{7}$$

[2]For mathematical preliminaries and omitted notation, please refer to Appendix 10.

Physically, the vertical subspace $\mathcal{V}_{\mathbf{C}}$ corresponds to infinitesimal gauge rotations that leave the electronic state invariant, whereas the horizontal subspace $\mathcal{H}_{\mathbf{C}}$ captures the intrinsic evolution of the occupied subspace. For a dynamical trajectory $\mathbf{C}_t$, this framework identifies the *covariant derivative* as the horizontal projection of the velocity, subtracting the gauge potential $\mathcal{A}(t) \coloneqq \mathbf{C}_t^\dagger \mathrm{S} \dot{\mathbf{C}}_t$:

$$\text{(Primary Dynamics)} \qquad \nabla_t \mathbf{C}_t \coloneqq \mathrm{Hor}_{\mathbf{C}}(\dot{\mathbf{C}}_t) = \dot{\mathbf{C}}_t - \mathbf{C}_t \mathcal{A}(t) = \dot{\mathbf{C}}_t - \mathbf{C}_t(\mathbf{C}_t^\dagger \mathrm{S} \dot{\mathbf{C}}_t), \tag{8}$$

satisfying $\mathrm{Hor}_{\mathbf{C}} : T_{\mathbf{C}}\mathrm{St}_{\mathrm{S}} \to \mathcal{H}_{\mathbf{C}}$ and $\mathbf{C}_t^\dagger S \nabla_t \mathbf{C}_t = 0$. The horizontal geometry introduced above provides a gauge-covariant notion of electronic motion: the horizontal derivative $\nabla_t \mathbf{C}_t$ removes arbitrary unitary rotations within the occupied subspace, ensuring that kinetic quantities depend only on the evolving subspace $[\mathbf{C}_t] \coloneqq \{\mathbf{C}_t \mathbf{U} \mid \mathbf{U} \in \mathrm{U}(N)\} \in \mathrm{Gr}_{\mathrm{S}}$. To express the variational dynamics in phase-space form, we introduce a momentum-like dual variable $\mathbf{\Pi}$:

$$\text{(Momentum Variable)} \qquad \mathbf{\Pi}_t \coloneqq (\nabla_t \mathbf{C}_t)^\flat = \mathrm{S}(\mathbf{r}_t) \nabla_t \mathbf{C}_t, \quad \mathbf{\Pi}_t^\sharp = \mathrm{S}(\mathbf{r}_t)^{-1} \mathbf{\Pi}_t, \tag{9}$$

where $(\cdot)^\flat$ and $(\cdot)^\sharp$ denote the musical isomorphisms under predefined metric. By construction, the induced evolution of the density matrix can be written as dual dynamics mirroring (8):

$$\text{(Dual Dynamics)} \qquad \nabla_t \mathbf{P}_t = \mathbf{\Pi}_t^\sharp \mathbf{C}_t^\dagger \mathrm{S} + \mathbf{C}_t (\mathbf{\Pi}_t^\sharp)^\dagger \mathrm{S} = (\nabla_t \mathbf{C}_t) \mathbf{C}_t^\dagger \mathrm{S} + \mathbf{C}_t (\nabla_t \mathbf{C}_t)^\dagger \mathrm{S}. \tag{10}$$

Fig. 1 highlights the corresponding dual flows of the horizontal electronic motion $\nabla_t \mathbf{C}_t$ (**red**) and the induced evolution of occupied-subspace matrix $\nabla_t \mathbf{P}_t$ (**green**).

Gauge-Symmetric Dual Lagrangian Dynamics. To achieve the energy minimization in (2), we next capitalize on a gauge-covariant geometric dual dynamics in (8) and (9) whose induced flow guarantees monotonic energy decrease over time while continuously defined. The key modeling choice is to place inertia only on the physically meaningful horizontal change of the occupied subspace, while evaluating the energy through the gauge-invariant projector $\mathbf{P}_t$. With this choice, we introduce a lifted extended Lagrangian over $(\mathbf{C}_t, \mathbf{P}_t)$, together with a Rayleigh dissipation term with $\gamma > 0$, so that the dynamics monotonically decreases the electronic energy, *i.e.*, $\frac{d}{dt}\mathcal{E} \leq 0$. More precisely, we take the *Lagrange-d'Alembert variation* (Marsden and Ratiu, 1999) on the Stiefel manifolds[3]

$$\text{(Lagrange-d'Alembert Variation)} \qquad \partial_{\mathbf{C}} \mathcal{L} - \nabla_t \big(\partial_{\nabla_t \mathbf{C}} \mathcal{L}\big) - \partial_{\nabla_t \mathbf{C}} \mathcal{D} = \mathbf{0}, \tag{11}$$

where $\mathcal{L}$ denotes the lifted electronic Lagrangian consisting of the horizontal kinetic term and the electronic energy, and $\mathcal{D}$ denotes the Rayleigh dissipation that damps the horizontal motion:

$$\mathcal{L}(\mathbf{C}_t, \nabla_t \mathbf{C}_t, \mathbf{P}_t; \mathbf{r}_t) \coloneqq \frac{\mu_t}{2} \|\nabla_t \mathbf{C}_t\|_{\mathrm{S}_t}^2 - \mathcal{E}(\mathbf{r}_t, \mathbf{P}_t), \quad \mathcal{D}(\mathbf{C}_t, \nabla_t \mathbf{C}_t; \mathbf{r}_t) \coloneqq \frac{\gamma}{2} \|\nabla_t \mathbf{C}_t\|_{\mathrm{S}_t}^2. \tag{12}$$

To simplify the discussion, Proposition 3.1 presents the dual dynamical system in a systematic form:

Proposition 3.1 (Gauge-Symmetric Dual Lagrangian Dynamics)**.** *For the nuclear trajectory $t \mapsto \mathbf{r}_t$, damping coefficients $\mu_t = e^{-at}$ and $a, \gamma > 0$, the following system of the physical states ensures energy minimization $\frac{d}{dt}\mathcal{E}(\mathbf{r}_t, \mathbf{C}_t) \approx -\|\mathrm{grad}_{\mathbf{C}}^{\mathrm{Hor}} \mathcal{E}(\mathbf{r}_t, \mathbf{C}_t)\|_{\mathrm{S}}^2 \leq 0$:*

$$(\mathbf{C}_t, \mathbf{\Pi}_t, \mathbf{P}_t)^\top \iff \begin{cases} \nabla_t \mathbf{C}_t = \mathbf{\Pi}_t^\sharp = \mathrm{S}_t^{-1} \mathbf{\Pi}_t, \\ \mu_t \nabla_t \mathbf{\Pi}_t + \gamma \mathbf{\Pi}_t + (\mathrm{grad}_{\mathbf{C}}^{\mathrm{Hor}} \mathcal{E})^\flat \big(\mathbf{r}_t, \mathbf{P}_t\big) = 0, \\ \nabla_t \mathbf{P}_t = \mathrm{S}_t^{-1} \mathbf{\Pi}_t \mathbf{C}_t^\dagger \mathrm{S}_t + \mathbf{C}_t \mathbf{\Pi}_t^\dagger, \end{cases} \tag{13}$$

where $\mathbf{H}^\star$ is self-consistent Hamiltonian matrix satisfying (5) *and the musical isomorphism of horizontal gradient in* (13) *can be derived as follows:*

$$(\mathrm{grad}_{\mathbf{C}}^{\mathrm{Hor}} \mathcal{E})^\flat = \mathrm{S}(\mathbf{r})(\mathrm{grad}_{\mathbf{C}}^{\mathrm{Hor}} \mathcal{E}) = \mathrm{S}(\mathbf{r})(\mathbf{I} - \mathbf{P})\mathrm{S}(\mathbf{r})^{-1} \mathbf{H}^\star \mathbf{C} = \mathbf{H}^\star \mathbf{C} - \mathrm{S}(\mathbf{r}) \mathbf{C} \big(\mathbf{C}^\dagger \mathbf{H}^\star \mathbf{C}\big).$$

Eq. (13) defines a gauge-covariant damped dynamics on the occupied subspace that serves as a continuous relaxation toward self-consistency. The Rayleigh term $\gamma \mathbf{\Pi}$ dissipates the horizontal kinetic energy and enforces monotone decay of the electronic energy so that $\frac{d}{dt}\mathcal{E}(\mathbf{r}_t, \mathbf{C}_t) \leq 0$, ensuring that this flow can be used as an SCF-free energy minimization mechanism within BOMD to track the instantaneous Born-Oppenheimer electronic ground state in (2) along the nuclear trajectory. Proposition 11.8 guarantees BOMD consistency by ensuring that the *Born-Oppenheimer ground*

[3] Proposition 11.1 provides detailed derivation process and full description.

state condition in (2) *is identical to the system in* (13), and in the fast relaxation regime ($\mu_t \to 0$ with $\gamma > 0$), the resulting forces and nuclear trajectories agree with defined BOMD up to an $\mathcal{O}(\mu_t)$ error.

In practice, we realize the continuous time dual flow in Proposition 3.1 through the discretized scheme outlined in Algorithm 1, using a first-order structure preserving discretization that ensures $\mathbf{C}_t \in \mathrm{St}_{\mathrm{S}_t}$, while strictly maintaining the horizontality of $\mathbf{\Pi}_t$ at each step:

$$\mathbf{\Pi}_{t+1} = \mathrm{Hor}_{\mathbf{C}_t}\left(\frac{\mathbf{\Pi}_t - d_t\mu_t^{-1}(\mathrm{grad}^{\mathrm{Hor}}_{\mathbf{C}}\mathcal{E})^\flat(\mathbf{r}_t,\mathbf{P}_t)}{1+d_t\gamma\mu_t^{-1}}\right), \quad \begin{cases}\mathbf{C}_{t+1} = \mathrm{Exp}_{\mathbf{C}_t}\left(d_t\mathbf{\Pi}^\sharp_{t+1}\right),\\ \mathbf{P}_t = \mathbf{C}_t\mathbf{C}_t^\dagger \mathrm{S}(\mathbf{r}_t).\end{cases} \tag{14}$$

where d_t denotes the time step of the discrete grid, and $\mathrm{Exp}_{\mathbf{C}_t} : T_{\mathbf{C}_t}\mathrm{St}_{\mathrm{S}_t} \to \mathrm{St}_{\mathrm{S}_t}$ is the Riemannian exponential map on the S_t-orthonormal Stiefel manifold.

In the discretized dynamics of (14), the only quantity not directly available during inference is the self-consistent Hamiltonian $\mathbf{H}^\star$, whose evaluation would otherwise require solving an SCF problem at every step. To remove this bottleneck, our next goal is to replace $\mathbf{H}^\star$ with a neural Hamiltonian operator $\mathbf{H}_\theta$, parameterized by θ and trained so that $\mathbf{H}_\theta \approx \mathbf{H}^\star$. This allows the same structure-preserving update scheme to be carried out in a fully SCF-free manner, producing BOMD trajectories $\{\mathbf{C}_t(\theta), \mathbf{\Pi}_t(\theta), \mathbf{P}_t(\theta)\}$ driven entirely by $\mathbf{H}_\theta$. While $\mathbf{H}^\star$ is replaced by learned Hamiltonian $\mathbf{H}_\theta$, the resulting simulator realizes the same BO-consistent, energy-relaxing electronic flow approximately in practice, with the quality of this approximation governed by how accurately $\mathbf{H}_\theta$ captures $\mathbf{H}^\star$.

Neural Hamiltonian Dynamics. To enable direct force evaluation and stable nuclear updates in (1), $\mathbf{H}_\theta$ is generally required to be differentiable with respect to the nuclear coordinates and admit an analytic expression for $\nabla_{\mathbf{r}}\mathbf{H}_\theta(\mathbf{r},\cdot)$. To satisfy this requirement by construction, we propose a *neural Hamiltonian* as a Hermitian basis expansion where coefficients depend on nuclear geometry and the gauge-symmetric subspace:

$$\mathbf{H}_\theta(\mathbf{r},\mathbf{C}) = \bar{\mathbf{H}}_\theta + \sum_{k=1}^{K}\mathfrak{C}_k^\theta(\mathbf{f})\mathbf{B}_k(\theta), \;\; \mathbf{f}(\mathbf{r},\mathbf{C}) = \left[\sum_{I<J}\mathrm{Z}_I\mathrm{Z}_J\frac{e^{-\|\mathbf{r}_I-\mathbf{r}_J\|}}{\|\mathbf{r}_I-\mathbf{r}_J\|}, \; \mathrm{Tr}\left(\mathbf{P}(\mathbf{C})\mathbf{Q}\right)\right]^\top.$$

Here, $\bar{\mathbf{H}}_\theta \in \mathbb{C}^{M\times M}$ and $\{\mathbf{B}_k\}_{k=1}^K \subset \mathbb{C}^{M\times M}$ are learnable Hermitian matrices, and the coefficient vector $\mathfrak{C}^\theta = (\mathfrak{C}_1^\theta,\ldots,\mathfrak{C}_K^\theta)^\top \in \mathbb{R}^K$ is produced by a neural network, $\{\mathbf{Q}\}$ is a Hermitian probe matrix. Then, we evaluate the network Jacobian $\mathbf{J} = \partial\mathfrak{C}^\theta/\partial\mathbf{f}$ in forward mode by propagating pairs through the neural network layers, which avoids higher-order autograd. Combined with the analytic feature derivatives $\partial\mathbf{f}/\partial\mathbf{r}$, Eq. (15) yields a closed-form coordinate derivative $\partial\mathbf{H}_\theta/\partial\mathbf{r}$ via the chain rule while preserving Hermiticity.

$$\mathbf{J} \coloneqq \frac{\partial\mathfrak{C}^\theta}{\partial\mathbf{f}} \in \mathbb{R}^{K\times N_{\mathrm{f}}}, \quad \frac{\partial\mathfrak{C}_k^\theta}{\partial\mathbf{r}} = \sum_{n=1}^{N_{\mathrm{f}}}\mathbf{J}_{k,n}\frac{\partial\mathbf{f}_n}{\partial\mathbf{r}}, \quad \frac{\partial\mathbf{H}_\theta}{\partial\mathbf{r}}[\mathbf{r},\mathbf{C}] = \sum_{k=1}^{K}\frac{\partial\mathfrak{C}_k^\theta}{\partial\mathbf{r}}\mathbf{B}_k. \tag{15}$$

For the standard energy-weighted density constructed from occupied canonicalization, the resulting $\mathbf{H}_\theta$ and $\partial\mathbf{H}_\theta/\partial\mathbf{r}$ plug directly into the *Hellmann-Feynman-Pulay operator* force (Pathak et al., 2023):

$$\overrightarrow{\mathbf{F}}_t(\theta) = -\,\mathrm{Tr}\left[\mathbf{P}_t\frac{\partial\mathbf{H}_\theta}{\partial\mathbf{r}}(\mathbf{r}_t,\mathbf{C}_t)\right] + \mathrm{Tr}\left[\mathbf{W}\frac{\partial\mathrm{S}(\mathbf{r}_t)}{\partial\mathbf{r}}\right] - \frac{\partial\mathbf{E}_{\mathrm{nn}}(\mathbf{r}_t)}{\partial\mathbf{r}}, \quad \begin{cases}\dot{\mathbf{r}}_{t+1} = \dot{\mathbf{r}}_t + \frac{d_t}{\mathbf{m}}\overrightarrow{\mathbf{F}}_t(\theta),\\ \mathbf{r}_{t+1} = \mathbf{r}_t + d_t\dot{\mathbf{r}}_{t+1}.\end{cases}$$

Here, $\mathbf{P}_t$ and $\mathbf{C}_t$ are obtained from solving the discrete dual dynamics in (14) and $\mathbf{W}_t \coloneqq \mathbf{C}_t\boldsymbol{\epsilon}_t\mathbf{C}_t^\dagger$ is the energy-weighted density matrix, with $\boldsymbol{\epsilon}_t$ the diagonal matrix of occupied orbital energies obtained by occupied canonicalization. Finally, to realize the BOMD update in Eq. (1), we use the induced forces to propagate the nuclear degrees of freedom with a standard time discretization *i.e.*, $\mathbf{r}_t \to \mathbf{r}_{t+1}$. An overview of the end-to-end BOMD loop consisting of electronic relaxation, force evaluation, and nuclear propagation, is illustrated in Fig. 1 *(right)*. Thus, our *training objective function* $\mathcal{L}(\theta)$ fits $\mathbf{H}_\theta$ by minimizing the time-integrated mismatch of the induced Hamiltonian and force along reference BOMD trajectories:

$$\min_{\theta\in\Theta}\; \mathcal{L}(\theta) \coloneqq \mathbb{E}_{\mathbf{F}^\star,\mathbf{H}^\star}\left[\int_0^T\left(\left\|\overrightarrow{\mathbf{F}}_t(\theta) - \overrightarrow{\mathbf{F}}_t^\star\right\|_2^2 + \left\|\mathbf{H}_\theta(\mathbf{r}_t,\mathbf{C}_t) - \mathbf{H}_t^\star\right\|_F^2\right)dt\right], \tag{16}$$

where $\overrightarrow{\mathbf{F}}^\star$ and $\mathbf{H}^\star$ are sampled from ground-truth SCF-based reference BOMD trajectories by running a conventional electronic-structure solver at each nuclear configuration $\mathbf{r}_t$ to obtain the self-consistent Hamiltonian and the corresponding forces.

Table 1: **Quantitative Result on Aspirin in MD17 Dataset.** We report physical accuracy metrics including MSE, RMSD, pairwise distance errors, planar structure preservation, and inference time, the number of SCF iterations for computational efficiency. Best and second best results are highlighted in **bold** and **blue**, respectively.

Model	Physical Metrics				Computational Efficiency	
	MSE (Å^2)	RMSD (Å)	DME (Å)	Planarity (Å)	Inference (s) (CPU / GPU)	SCF iter (CPU / GPU)
(A) Geometry-aware Hamiltonian Dynamical Models						
NHD-occ (Park, 2025) ($\mathbf{P}_t$)	6.47×10^{-4}	4.38×10^{-2}	3.11×10^{-2}	9.63×10^{-3}	4.46×10^{1} (GPU)	**0** / **0**
NHD-orb (Park, 2025) ($\mathbf{C}_t$)	6.28×10^{-4}	4.32×10^{-2}	2.96×10^{-2}	9.98×10^{-3}	4.09×10^{1} (GPU)	**0** / **0**
(B) Solver Structured and Generative Reformulations						
DEQHNet (Wang et al., 2024)	4.23×10^{-5}	1.05×10^{-2}	7.62×10^{-3}	3.10×10^{-4}	2.35×10^{3} / 3.67×10^{2}	1.75×10^{1} / 1.67×10^{1}
QHFlow (Kim et al., 2025a)	4.07×10^{-5}	1.02×10^{-2}	7.31×10^{-3}	3.99×10^{-4}	1.67×10^{3} / 1.40×10^{2}	1.85×10^{1} / 1.83×10^{1}
(C) Equivariant Operator and Wavefunction Predictions						
PhiSNet (Unke et al., 2021a)	7.33×10^{-3}	1.24×10^{-1}	5.96×10^{-2}	1.73×10^{-2}	1.55×10^{3} / 1.74×10^{2}	2.53×10^{1} / 2.61×10^{1}
QHNet (Yu et al., 2023)	3.94×10^{-5}	9.97×10^{-3}	7.27×10^{-3}	2.41×10^{-4}	1.21×10^{3} / 1.21×10^{2}	1.65×10^{1} / 1.64×10^{1}
WANet (Li et al., 2025)	4.20×10^{-5}	1.03×10^{-2}	7.45×10^{-3}	**2.37×10^{-4}**	2.55×10^{3} / 1.14×10^{2}	1.67×10^{1} / 1.60×10^{1}
SPHNet (Luo et al., 2025)	**3.67×10^{-5}**	**9.64×10^{-3}**	**6.96×10^{-3}**	3.02×10^{-4}	1.88×10^{3} / 1.03×10^{2}	1.53×10^{1} / 1.46×10^{1}
Ours	**7.38×10^{-6}**	**4.67×10^{-3}**	**3.31×10^{-3}**	**1.03×10^{-4}**	**2.97×10^{0}** (GPU)	**0** / **0**

4 Related Work

Learning Electronic Structures and Operators. Recent approaches leverage symmetry-equivariance to efficiently predict electronic quantities. (Unke et al., 2021a) and (Yu et al., 2023) utilize $\mathrm{SE}(3)$-equivariant architectures for wavefunctions and Hamiltonians, while subsequent works have improve scalability through sparsity and alignment objectives (Luo et al., 2025; Li et al., 2025). Alternative frameworks incorporate solver constraints or generative priors, such as modeling SCF fixed points (Wang et al., 2024; Kim et al., 2025a). However, these methods are generally unable to provide differentiable BOMD trajectories. In contrast, our framework derives *the electronic structure while being explicitly optimized for BOMD*, coupling a learnable Hamiltonian for accurate dynamics.

Classical SCF-free BOMD. A line of research also propagates auxiliary electronic variables alongside nuclei (Car and Parrinello, 1985; Niklasson, 2008), yet they are generally ill-suited for learning integration. These integrators often require handcrafted hyperparameters and fail to provide a differentiable operator pipeline necessary for accelerator-friendly batching. In contrast, our approach couples gauge-symmetric electronic dynamics with a learned neural Hamiltonian, enabling efficient *end-to-end training without differentiating through SCF fixed points or diagonalization heavy solvers*.

Machine Learning Interatomic Potentials. This line of work learns a surrogate potential energy surface to run large-scale MD at substantially reduced cost (Mishin, 2021). However, MLIPs explicitly discard the electronic problem: they provide no electronic state variables and lack enforcement of the Kohn–Sham variational structure, thus electronic observables are unavailable and variationally consistent quantum-level forces are not guaranteed, which can lead to physically inconsistent dynamics when electronic structure matters. Our goal is instead *electronic structure preserving ab initio* BOMD with SCF independent propagation to remove the SCF bottleneck. Since MLIPs target a different mesoscopic molecular surrogate regime, we exclude them from our experimental baselines.

5 Experiments

In this section, our primary objective is to achieve high fidelity *ab initio* molecular dynamics driven by KS-DFT simulations while eliminating the computational bottleneck associated with iterative SCF procedures. In this framework, the electronic ground state is obtained at each time step by minimizing the energy functional through iterative SCF solvers. Note that all quantum chemical calculations are conducted using the same quantum chemical setting of PySCF (Sun et al., 2018) for fair comparisons.

MD17/MD22 Molecular Benchmarks. We evaluate our method on multiple molecular systems from the MD17 (Chmiela et al., 2017) and MD22 (Chmiela et al., 2023) *ab initio* molecular dynamics benchmarks under a common simulation protocol. Specifically, we select aspirin, salicylic acid, and azobenzene from MD17 and Ac-Ala3-NHMe from MD22. For each molecular system, the dataset is constructed using 256 uniformly sampled frames and the corresponding BOMD trajectory segments. This benchmark suite covers near-planar molecules and larger flexible systems, providing a setting where small electronic-state errors can accumulate into geometric distortions and energy drift.

Tables 1 and 3 summarize the performance on aspirin and Ac-Ala3-NHMe, grouping prior methods into three categories. (A) *Manifold-valued Hamiltonian dynamical models* explicitly evolve electronic states on constrained manifolds $(\mathrm{St}, \mathrm{Gr})$ and target SCF-free propagation. (B) *Solver structured and generative reformulations* inject self-consistency structure or generative modeling

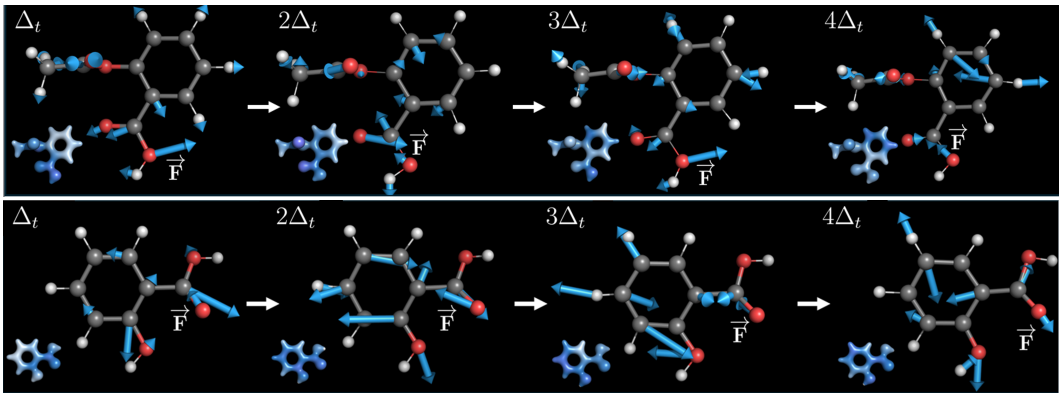


Figure 2: **Qualitative Results on Aspirin and Salicylic Acid.** We visualize molecular dynamics trajectories predicted by our model for aspirin ($C_9H_8O_4$) and salicylic acid ($C_7H_6O_3$). At intervals of Δ_t steps, blue arrows indicate the evaluated atomic forces, and the electron-density cloud is shaded by force magnitude.

into Hamiltonian learning, often retaining iterative relaxation behavior. (C) *Equivariant operator and wavefunction predictions* leverage symmetry equivariant architectures for accurate static operator or wavefunction inference. Fig. 2 visualizes representative dynamics from the proposed SCF-free simulations, illustrating the resulting trajectory and the evaluated forces.

In addition to standard trajectory level error metrics including both MSE and RMSD, we report $\mathrm{SE}(3)$-invariant geometric metric including DME, Planarity, and Rg to assess the structural integrity of the generated trajectories independent of the global frame. To evaluate computational efficiency, we report the time averaged metrics along with the inference time and the average number of SCF iterations per MD simulation step. As can be seen, the proposed framework consistently improves both inference time and trajectory level accuracy across both molecular systems. In particular, the method achieves a $4.97\times$ reduction in MSE relative to the strongest baseline, with a similar decrease in RMSD. We also observe this trend in DME and Rg, with Planarity achieving 2.30 times better performance. These results suggest that enforcing gauge-symmetric formulation provides an effective mechanism for adiabatic tracking of the occupied subspace.

Notably, SCF solvers incur a major time overhead because each MD step requires repeated electronic structure evaluations including diagonalization with $O(M^3)$ cost. Baselines using SCF require approximately **17** SCF iterations per MD step on average, with counts ranging from 8 to 26. In contrast, our method uses **0** SCF iteration throughout the simulation. This indicates that the computational gain comes from avoiding the iterative SCF loop itself rather than reducing its iteration budget. By directly propagating the occupied subspace and its corresponding density projector with a learned Hamiltonian, our method realizes an SCF-free electronic propagation scheme. It achieves an approximate $\mathbf{4.1 \times 10^2}$ speedup over CPU PySCF and an approximate $\mathbf{3.5 \times 10^1}$ speedup over GPU-accelerated PySCF. In addition, our framework achieves lower geometric errors than manifold-valued Hamiltonian dynamical baselines that also run without SCF iterations.

To further assess whether learned electronic updates can support SCF-free BOMD propagation, we remove all SCF iterations from the baselines that normally rely on them. Once this correction is disabled, errors in the predicted electronic quantities directly enter force evaluation and accumulate along the trajectory. As visualized in Section 9, these baselines show unstable molecular trajectories. This is also reflected by the pronounced deterioration in MSE and RMSD in Table 2. By comparison, the proposed SCF-free update maintains stable trajectory propagation under the same zero SCF setting while preserving the molecular geometry. Overall, we conclude that the gauge-symmetric structure of our method enables it to replace the SCF loop while improving the accuracy-efficiency trade-off. Further additional experiments (Section 9), algorithms (Section 12) and implementation details (Section 8) are provided in the Appendix.

Table 2: **Comparison under a zero SCF iterations.** We report the MSE and RMSD of SCF-based baselines when the SCF iteration budget is restricted to zero, evaluated up to the $2\Delta_t$ step on Aspirin and Salicylic Acid molecular systems.

	Aspirin		Salicylic Acid	
Models	MSE	RMSD	MSE	RMSD
WANet	1.03×10^5	5.53×10^2	4.30×10^3	1.10×10^2
SPHNet	1.78×10^3	7.05×10^1	9.56×10^2	5.08×10^1
QHFlow	1.15×10^3	5.60×10^1	1.23×10^2	1.81×10^1

Table 3: **Quantitative Result on Ac-Ala3-NHMe in MD22 Dataset.** We report physical accuracy metrics including MSE, RMSD, pairwise distance errors, radius of gyration error, and inference time, the number of SCF iterations. Best and second-best results are highlighted in **bold** and **blue**, respectively.

Model	Physical Metrics				Computational Efficiency	
	MSE (Å^2)	RMSD (Å)	DME (Å)	Rg MAE (Å)	Inference (s) (CPU / GPU)	SCF iter (CPU / GPU)
(A) Geometry-aware Hamiltonian Dynamical Models						
NHD-occ (Park, 2025) ($\mathbf{P}_t$)	3.12×10^{-4}	3.01×10^{-2}	2.06×10^{-2}	1.54×10^{-2}	9.86×10^{1} (GPU)	0 / 0
NHD-orb (Park, 2025) ($\mathbf{C}_t$)	2.91×10^{-4}	3.08×10^{-2}	1.84×10^{-2}	1.67×10^{-2}	8.46×10^{1} (GPU)	0 / 0
(B) Solver Structured and Generative Reformulations						
DEQHNet (Wang et al., 2024)	1.94×10^{-5}	7.13×10^{-3}	4.74×10^{-3}	$\mathbf{4.57 \times 10^{-4}}$	4.60×10^{3} / 3.52×10^{2}	2.16×10^{1} / 2.27×10^{1}
QHFlow (Kim et al., 2025a)	1.90×10^{-5}	7.07×10^{-3}	4.71×10^{-3}	4.63×10^{-4}	2.82×10^{3} / 2.05×10^{2}	7.91×10^{0} / 7.64×10^{0}
(C) Equivariant Operator and Wavefunction Predictions						
PhiSNet (Unke et al., 2021a)	2.89×10^{-2}	2.62×10^{-1}	1.34×10^{-1}	4.80×10^{-2}	3.14×10^{3} / 3.10×10^{2}	2.06×10^{1} / 2.04×10^{1}
QHNet (Yu et al., 2023)	1.99×10^{-5}	7.24×10^{-3}	4.79×10^{-3}	4.79×10^{-4}	3.13×10^{3} / 2.57×10^{2}	2.15×10^{1} / 2.17×10^{1}
WANet (Li et al., 2025)	$\mathbf{1.82 \times 10^{-5}}$	$\mathbf{6.92 \times 10^{-3}}$	$\mathbf{4.57 \times 10^{-3}}$	4.74×10^{-4}	5.60×10^{3} / 2.46×10^{2}	2.42×10^{1} / 2.23×10^{1}
SPHNet (Luo et al., 2025)	1.83×10^{-5}	6.95×10^{-3}	4.63×10^{-3}	5.25×10^{-4}	4.12×10^{3} / 2.22×10^{2}	1.70×10^{1} / 1.86×10^{1}
Ours	$\mathbf{3.59 \times 10^{-6}}$	$\mathbf{3.19 \times 10^{-3}}$	$\mathbf{2.16 \times 10^{-3}}$	$\mathbf{1.64 \times 10^{-4}}$	$\mathbf{6.47 \times 10^{0}}$ (GPU)	**0 / 0**

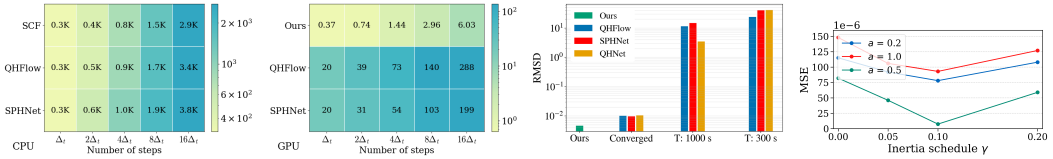


Figure 3: **Ablation Study.** *(left)* Heatmaps of total inference time in seconds across methods and step counts, measured with CPU- and GPU-accelerated PySCF. *(middle)* The plot shows the RMSD obtained with each method under full convergence and different BOMD time budgets on aspirin. *(right)* Comparison of trajectory fidelity under different inertia schedules γ and damping coefficients μ_t.

Ablation Study. We conduct a comprehensive empirical evaluation to investigate (i) inference time scaling with respect to simulation length, (ii) runtime advantage with fixed wall-clock budgets, and (iii) the specific contribution of stabilization terms in the dual Lagrangian dynamics.

(1) How does inference time scale with trajectory length? To assess computational efficiency, we measure the total time for varying the MD trajectory length in Fig. 3 *(left)* under both CPU and GPU-accelerated PySCF settings. SCF-based baselines perform a SCF solver at every step and GPU acceleration reduces the per-step inference time. Nevertheless, cumulative convergence overhead dominates and the cost increases rapidly with the trajectory length. In contrast, our method avoids SCF iterations and keeps the per-step computation essentially constant, which results in substantially lower inference time.

(2) How do benchmarks perform under runtime constraints? We examine the practical significance of our fast inference speed by imposing wall-clock budgets on the SCF-based baselines and by using the runtime regime of our method as the reference point. Specifically, we additionally test run time budgets of $T = 1000$ and $T = 300$ seconds, which is over $100\times$ larger than the runtime of our method. As shown in Fig. 3 *(middle)*, the baseline methods deteriorate rapidly with a sharp increase in RMSD under tighter time budgets. These results show that our method defines a faster regime that existing SCF-based approaches cannot reach without significant accuracy loss.

(3) Are damping and inertia necessary for stability? We conduct ablate study on the stabilization components by varying the inertia schedules γ and the decay coefficient a while keeping other modules fixed. Specifically, we compare the default time-decaying inertia schedule against variants with no damping, constant inertia, or slower relaxation. The evaluated fidelity in Fig. 3 *(right)* show that removing damping or using less aggressive relaxation amplifies force inconsistency and geometric error, whereas the default setting is the most stable. These results confirm that damping and inertia are essential for accurate and stable SCF-free dynamics.

6 Conclusion

We introduced a gauge-symmetric dual Lagrangian framework for SCF-free Born–Oppenheimer molecular dynamics by propagating the occupied electronic subspace and its density projector directly on the Grassmann manifold. Proposed dissipative dynamics stabilizes the electronic relaxation, and a gauge-symmetric neural Hamiltonian provides analytic coordinate derivatives for closed-form forces. Experiments show that the proposed approach can match self-consistent reference trajectories while eliminating SCF iterations and reducing computational overhead.

Acknowledgments and Disclosure of Funding

This work was supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP)-ICT Creative Consilience Program grant funded by the Korea government (MSIT) (IITP-2026-RS-2020-II201819). This work was also supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2026-25495269). This research was supported by the Basic Science Research Program through the NRF funded by the Korea government (MSIT) (RS-2026-25522655).

References

Simon Batzner, Albert Musaelian, Lixin Sun, Mario Geiger, Jonathan P. Mailoa, Mordechai Kornbluth, Nicola Molinari, Tess E. Smidt, and Boris Kozinsky. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature Communications*, 13(1):2453, May 2022. doi: 10.1038/s41467-022-29939-5.

Jörg Behler and Michele Parrinello. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.*, 98:146401, Apr 2007. doi: 10.1103/PhysRevLett.98.146401.

Felix Brockherde, Leslie Vogt, Li Li, Mark E. Tuckerman, Kieron Burke, and Klaus-Robert Müller. Bypassing the kohn-sham equations with machine learning. *Nature Communications*, 8(1):872, Oct 2017. doi: 10.1038/s41467-017-00839-3.

Martin Buttenschoen, Garrett M. Morris, and Charlotte M. Deane. Posebusters: Ai-based docking methods fail to generate physically valid poses or generalise to novel sequences. *Chemical Science*, 15:3130–3139, 2024. doi: 10.1039/D3SC04185A. URL `http://dx.doi.org/10.1039/D3SC04185A`.

Eric Cancès and Gero Friesecke, editors. *Density Functional Theory: Modeling, Mathematical Analysis, Computational Methods, and Applications*. Mathematics and Molecular Modeling. Springer Cham, 2023. ISBN 978-3-031-22340-2. doi: 10.1007/978-3-031-22340-2.

Éric Cancès, Gaspard Kemlin, and Antoine Levitt. Convergence analysis of direct minimization and self-consistent iterations. *SIAM Journal on Matrix Analysis and Applications*, 42(1):243–274, 2021. doi: 10.1137/20M1332864.

R. Car and M. Parrinello. Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.*, 55:2471–2474, Nov 1985. doi: 10.1103/PhysRevLett.55.2471.

Stefan Chmiela, Alexandre Tkatchenko, Huziel E. Sauceda, Igor Poltavsky, Kristof T. Schütt, and Klaus-Robert Müller. Machine learning of accurate energy-conserving molecular force fields. *Science Advances*, 3(5):e1603015, 2017. doi: 10.1126/sciadv.1603015.

Stefan Chmiela, Valentin Vassilev-Galindo, Oliver T. Unke, Adil Kabylda, Huziel E. Sauceda, Alexandre Tkatchenko, and Klaus-Robert Müller. Accurate global machine learning force fields for molecules with hundreds of atoms. *Science Advances*, 9(2):eadf0873, 2023. doi: 10.1126/sciadv.adf0873.

Qiangqiang Gu, Zhanghao Zhouyin, Shishir Kumar Pandey, Peng Zhang, Linfeng Zhang, and Weinan E. Deep learning tight-binding approach for large-scale electronic simulations at finite temperatures with ab initio accuracy. *Nature Communications*, 15(1):6772, Aug 2024. doi: 10.1038/s41467-024-51006-4.

Trygve Helgaker, Poul Jørgensen, and Jeppe Olsen. *Molecular Electronic-Structure Theory*. John Wiley & Sons, 2013. ISBN 9781118531471.

Jürg Hutter, Marcella Iannuzzi, and Thomas D. Kühne. Ab initio molecular dynamics: A guide to applications. In Manuel Yáñez and Russell J. Boyd, editors, *Comprehensive Computational Chemistry*, pages 493–517. Elsevier, 2024. doi: 10.1016/B978-0-12-821978-2.00096-9.

Radu Iftimie, Peter Minary, and Mark E. Tuckerman. Ab initio molecular dynamics: Concepts, recent developments, and future trends. *Proceedings of the National Academy of Sciences of the United States of America*, 102(19):6654–6659, 2005. doi: 10.1073/pnas.0500193102.

Seongsu Kim, Nayoung Kim, Dongwoo Kim, and Sungsoo Ahn. High-order equivariant flow matching for density functional theory hamiltonian prediction. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025a. URL https://openreview.net/forum?id=iFIjNXb0Y5.

Sunghwan Kim, Jie Chen, Tiejun Cheng, Asta Gindulyte, Jia He, Siqian He, Qingliang Li, Benjamin A Shoemaker, Paul A Thiessen, Bo Yu, et al. Pubchem 2025 update. *Nucleic Acids Research*, 53 (D1):D1516–D1525, 2025b. doi: 10.1093/nar/gkae1059.

Yunyang Li, Zaishuo Xia, Lin Huang, Xinran Wei, Samuel Harshe, Han Yang, Erpai Luo, Zun Wang, Jia Zhang, Chang Liu, Bin Shao, and Mark Gerstein. Enhancing the scalability and applicability of kohn-sham hamiltonians for molecular systems. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=twEvvkQqPS.

Xiaohui Lin, Yongquan Jiang, and Yan Yang. Molecular distance matrix prediction based on graph convolutional networks. *Journal of Molecular Structure*, 1257:132540, 2022. ISSN 0022-2860. doi: https://doi.org/10.1016/j.molstruc.2022.132540.

M Yu Lobanov, NS Bogatyreva, and OV Galzitskaya. Radius of gyration as an indicator of protein structure compactness. *Molecular Biology*, 42(4):623–628, 2008. doi: 10.1134/S0026893308040195.

Erpai Luo, Xinran Wei, Lin Huang, Yunyang Li, Han Yang, Zaishuo Xia, Zun Wang, Chang Liu, Bin Shao, and Jia Zhang. Efficient and scalable density functional theory Hamiltonian prediction through adaptive sparsity. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu, editors, *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pages 41368–41390. PMLR, 13–19 Jul 2025.

Jerrold E. Marsden and Tudor S. Ratiu. *Introduction to Mechanics and Symmetry: A Basic Exposition of Classical Mechanical Systems*, volume 17 of *Texts in Applied Mathematics*. Springer, 2 edition, 1999. doi: 10.1007/978-0-387-21792-5.

Dominik Marx and Jürg Hutter. *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*. Cambridge University Press, 2009. doi: 10.1017/CBO9780511609633.

Y. Mishin. Machine-learning interatomic potentials for materials science. *Acta Materialia*, 214: 116980, 2021. ISSN 1359-6454. doi: 10.1016/j.actamat.2021.116980.

Anders M. N. Niklasson. Extended born-oppenheimer molecular dynamics. *Phys. Rev. Lett.*, 100: 123004, Mar 2008. doi: 10.1103/PhysRevLett.100.123004.

Sungwoo Park. Neural hamiltonian diffusions for modeling structured geometric dynamics. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=VswQY0peMr.

Robert G. Parr and Weitao Yang. *Density-Functional Theory of Atoms and Molecules*. Oxford University Press, 1989. ISBN 9780195092769. URL https://doi.org/10.1093/oso/9780195092769.005.0014.

Shivesh Pathak, Ignacio Ema López, Alex J. Lee, William P. Bricker, Rafael López Fernández, Susi Lehtola, and Joshua A. Rackers. Accurate hellmann-feynman forces from density functional calculations with augmented gaussian basis sets. *The Journal of Chemical Physics*, 158(1):014104, 2023. doi: 10.1063/5.0130668.

Chen Qian, Valdas Vitartas, James R Kermode, and Reinhard J Maurer. Equivariant electronic hamiltonian prediction with many-body message passing. *npj Computational Materials*, 2026. doi: 10.1038/s41524-026-02020-1.

Kristof Schütt, Pieter-Jan Kindermans, Huziel Enoc Sauceda Felix, Stefan Chmiela, Alexandre Tkatchenko, and Klaus-Robert Müller. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

Mohammad Shakiba and Alexey V. Akimov. Machine-learned kohn–sham hamiltonian mapping for nonadiabatic molecular dynamics. *Journal of Chemical Theory and Computation*, 20(8): 2992–3007, 2024. doi: 10.1021/acs.jctc.4c00008.

Qiming Sun, Timothy C. Berkelbach, Nick S. Blunt, George H. Booth, Sheng Guo, Zhendong Li, Junzi Liu, James D. McClain, Elvira R. Sayfutyarova, Sandeep Sharma, Sebastian Wouters, and Garnet Kin-Lic Chan. Pyscf: the python-based simulations of chemistry framework. *WIREs Computational Molecular Science*, 8(1):e1340, 2018. doi: https://doi.org/10.1002/wcms.1340.

Zechen Tang, He Li, Peize Lin, Xiaoxun Gong, Gan Jin, Lixin He, Hong Jiang, Xinguo Ren, Wenhui Duan, and Yong Xu. A deep equivariant neural network approach for efficient hybrid density functional calculations. *Nature Communications*, 15(1):8815, Oct 2024. doi: 10.1038/s41467-024-53028-4.

Oliver Unke, Mihail Bogojeski, Michael Gastegger, Mario Geiger, Tess Smidt, and Klaus-Robert Müller. Se(3)-equivariant prediction of molecular wavefunctions and electronic densities. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 14434–14447. Curran Associates, Inc., 2021a.

Oliver T. Unke, Stefan Chmiela, Huziel E. Sauceda, Michael Gastegger, Igor Poltavsky, Kristof T. Schütt, Alexandre Tkatchenko, and Klaus-Robert Müller. Machine learning force fields. *Chemical Reviews*, 121(16):10142–10186, 2021b. doi: 10.1021/acs.chemrev.0c01111.

Zun Wang, Chang Liu, Nianlong Zou, He Zhang, Xinran Wei, Lin Huang, Lijun Wu, and Bin Shao. Infusing self-consistency into density functional theory hamiltonian prediction via deep equilibrium models. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang, editors, *Advances in Neural Information Processing Systems*, volume 37, pages 89652–89681. Curran Associates, Inc., 2024. doi: 10.52202/079017-2846.

Brandon M Wood, Misko Dzamba, Xiang Fu, Meng Gao, Muhammed Shuaibi, Luis Barroso-Luque, Kareem Abdelmaqsoud, Vahe Gharakhanyan, John R. Kitchin, Daniel S. Levine, Kyle Michel, Anuroop Sriram, Taco Cohen, Abhishek Das, Sushree Jagriti Sahoo, Ammar Rizvi, Zachary Ward Ulissi, and C. Lawrence Zitnick. UMA: A family of universal models for atoms. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL `https://openreview.net/forum?id=SvopaNxYWt`.

Haiyang Yu, Zhao Xu, Xiaofeng Qian, Xiaoning Qian, and Shuiwang Ji. Efficient and equivariant graph networks for predicting quantum Hamiltonian. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett, editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 40412–40424. PMLR, 23–29 Jul 2023.

He Zhang, Chang Liu, Zun Wang, Xinran Wei, Siyuan Liu, Nanning Zheng, Bin Shao, and Tie-Yan Liu. Self-consistency training for density-functional-theory Hamiltonian prediction. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp, editors, *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 59329–59357. PMLR, 21–27 Jul 2024.

Linfeng Zhang, Jiequn Han, Han Wang, Roberto Car, and Weinan E. Deep potential molecular dynamics: A scalable model with the accuracy of quantum mechanics. *Phys. Rev. Lett.*, 120: 143001, Apr 2018. doi: 10.1103/PhysRevLett.120.143001.

Yang Zhong, Hongyu Yu, Mao Su, Xingao Gong, and Hongjun Xiang. Transferable equivariant graph neural networks for the hamiltonians of molecules and solids. *npj Computational Materials*, 9(1):182, Oct 2023. doi: 10.1038/s41524-023-01130-4.

Yang Zhong, Hongyu Yu, Jihui Yang, Xingyu Guo, Hongjun Xiang, and Xingao Gong. Universal machine learning kohn–sham hamiltonian for materials. *Chinese Physics Letters*, 41(7):077103, jun 2024. doi: 10.1088/0256-307X/41/7/077103.

Appendix

7 Notation Table

Throughout this paper, we adopt standard notation where scalars are denoted by normal font, and vectors and matrices are denoted by bold variables unless otherwise specified. The specific symbols, their meanings, and their associated domains are detailed in the following table.

Symbol	Meaning	Space / shape
$(\cdot)^\dagger$	conjugate transpose	–
$\mathrm{Tr}(\cdot)$	matrix trace	–
$\mathrm{Re}(\cdot), \mathrm{Im}(\cdot)$	real / imaginary part	–
$\|\cdot\|_F$	Frobenius norm	–
$\mathrm{diag}(\cdot)$	diagonal matrix construction	–
$\mathbf{r}$	nuclear coordinates	$\mathbb{R}^{3N_r}$
$\mathbf{r}_I(t)$	position of nucleus I at time t	$\mathbb{R}^3$
$\dot{\mathbf{r}}_I(t), \ddot{\mathbf{r}}_I(t)$	nuclear velocity / acceleration	$\mathbb{R}^3$
$\mathbf{m}_I$	mass of nucleus I	$\mathbb{R}_{>0}$
Z_I	nuclear charge of nucleus I	$\mathbb{R}$
$\mathbf{E}_{\mathrm{BO}}(\mathbf{r})$	Born-Oppenheimer energy surface	
$\mathbf{E}_{\mathrm{nn}}(\mathbf{r})$	nuclear repulsion energy	$\mathbb{R}$
$\mathbf{E}^\star(\mathbf{r})$	electronic ground-state energy	$\mathbb{R}$
$\mathbf{R}_I(t)$	extra force term in BOMD	$\mathbb{R}^3$
$\nabla_{\mathbf{r}}, \nabla_{\mathbf{r}_I}$	gradient w.r.t. nuclear coordinates	–
T	trajectory horizon / final time in integrals	$\mathbb{R}_{>0}$
$\{\chi_\mu(\mathbf{x};\mathbf{r})\}_{\mu=1}^M$	atom-centered basis functions	$\chi_\mu : \mathbb{R}^3 \to \mathbb{C}$
$\mathrm{S}(\mathbf{r})$	overlap matrix, $S_{\mu\nu}(r) = \int \chi_\mu^\dagger \chi_\nu dx$	$\mathbb{C}^{M\times M}$
I_N	identity matrix	$\mathbb{R}^{N\times N}$
$\mathbf{C}$	orbital (occupied) coefficient matrix	$\mathbb{C}^{M\times N}$
$\mathbf{C}^\dagger \mathrm{S}(\mathbf{r})\mathbf{C} = \mathrm{I}_N$	S-orthonormality constraint	–
$\mathbf{P} = \mathbf{P}(\mathbf{C})$	occupied-subspace projector: $\mathbf{C}\mathbf{C}^\dagger \mathrm{S}(\mathbf{r})$	$\mathbb{C}^{M\times M}$
$\mathbf{D}$	density matrix used in SCF: $\mathbf{D} := \mathbf{P}\mathrm{S}(\mathbf{r})^{-1}$	$\mathbb{C}^{M\times M}$
$\rho_{\mathbf{P}}(\mathbf{x})$	induced electron density from $\mathbf{P}$	$\mathbb{R}$
$\mathbf{H}(\mathbf{r})$	one-electron Hamiltonian matrix (Eq. (4))	$\mathbb{C}^{M\times M}$
$\mathcal{E}(\mathbf{r}, \mathbf{C})$	Kohn–Sham energy functional (Eq. (3))	$\mathbb{R}$
$E_{\mathrm{xc}}[\rho]$	exchange–correlation functional	$\mathbb{R}$
$\mathbf{F}(\mathbf{D})$	Fock matrix / effective Hamiltonian (Eq. (5))	$\mathbb{C}^{M\times M}$
$\mathbf{H}^\star$	self-consistent Hamiltonian (SCF reference)	$\mathbb{C}^{M\times M}$
$\mathrm{U}(N)$	unitary gauge group acting as $\mathbf{C} \mapsto \mathbf{C}\mathbf{U}$	group
$\mathfrak{u}(N)$	Lie algebra of $\mathrm{U}(N)$: $\Omega^\dagger = -\Omega$	$\mathbb{C}^{N\times N}$
$\mathrm{St}_{\mathrm{S}}(M, N)$	generalized Stiefel manifold: $\{\mathbf{C} : \mathbf{C}^\dagger \mathrm{S}\mathbf{C} = \mathrm{I}_N\}$	manifold in $\mathbb{C}^{M\times N}$
$\mathrm{Gr}_{\mathrm{S}}(M, N)$	generalized Grassmann manifold: $\mathrm{St}_{\mathrm{S}}/\mathrm{U}(N)$	quotient manifold
$[\mathbf{C}]$	gauge-equivalence class / occupied subspace	element of Gr_{S}
π	canonical projection $\pi(\mathbf{C}) = [\mathbf{C}]$	map $\mathrm{St}_{\mathrm{S}} \to \mathrm{Gr}_{\mathrm{S}}$
$T_{\mathbf{C}}\mathrm{St}_{\mathrm{S}}$	tangent space at $\mathbf{C}$	$\subset \mathbb{C}^{M\times N}$
$\langle X, Y\rangle_{\mathrm{S}}$	S-weighted real inner product: $\mathrm{Re}\,\mathrm{Tr}(X^\dagger \mathrm{S} Y)$	$\mathbb{R}$
$\|X\|_{\mathrm{S}}^2$	induced norm: $\mathrm{Tr}(X^\dagger \mathrm{S} X)$	$\mathbb{R}_{\geq 0}$
$\omega_{\mathbf{C}}$	connection 1-form	$T_{\mathbf{C}}\mathrm{St}_{\mathrm{S}} \to \mathfrak{u}(N)$
$\mathcal{V}_{\mathbf{C}}$	vertical subspace (gauge directions)	$\{\mathbf{C}\Omega : \Omega \in \mathfrak{u}(N)\}$
$\mathcal{H}_{\mathbf{C}}$	horizontal subspace	
$\mathrm{Ver}_{\mathbf{C}}(\cdot)$	vertical projection	–
$\mathrm{Hor}_{\mathbf{C}}(\cdot)$	horizontal projection: $X - \mathbf{C}(\mathbf{C}^\dagger \mathrm{S} X) = (\mathrm{I} - \mathbf{P})X$	–

8 Implementation Details

Molecular Dynamics Simulation. Reference molecular dynamics simulation trajectories are generated using Kohn–Sham density functional theory as implemented in PySCF (Sun et al., 2018). All calculations employ the PBE exchange–correlation functional with the def2-SVP basis set, and numerical integration employs a level-3 grid. At each molecular-dynamics step, the electronic ground state is obtained by a fully converged self-consistent field procedure with convergence tolerance 10^{-10} and a maximum of 80 SCF iterations. Initial nuclear velocities are sampled from the Maxwell–Boltzmann distribution at $500\mathrm{K}$. The center of mass momentum is removed and velocities are rescaled to match the target temperature. For each molecular system, we generate a reference trajectory over 25,600 MD integration steps and uniformly subsample it to obtain a temporally ordered sequence of 256 frames. We use Δ_t to denote a reporting interval of 32 prediction steps in our experiment section.

Dataset Statistics. We report the dataset statistics for our experiments in Table 5, listing the molecular systems and their corresponding data shapes.

Table 5: **Matrix and orbital dimensions for each molecular system.** We report the dimensions of the overlap matrix $\mathbf{S}$, Hamiltonian $\mathbf{H}$, and orbital matrix $\mathbf{C}$, along with $N_\mathbf{r}$, M, and N, which denote the number of atoms, atomic orbitals, and occupied orbitals, respectively.

Molecule	$N_\mathbf{r}$	**Formula**	S	$\mathbf{H}$	$\mathbf{C}$	M	N
Aspirin	21	$\mathrm{C_9H_8O_4}$	222×222	222×222	222×47	222	47
Salicylic acid	16	$\mathrm{C_7H_6O_3}$	170×170	170×170	170×36	170	36
Azobenzene	24	$\mathrm{C_{12}H_{10}N_2}$	246×246	246×246	246×48	246	48
Ac-Ala3-NHMe	42	$\mathrm{C_{12}H_{22}N_4O_4}$	390×390	390×390	390×77	390	77

Training Setup. All experiments are conducted on NVIDIA RTX 5000 Ada GPUs using Python 3.10.18 with PyTorch 2.5.1 and CUDA 12.4. CUDA synchronization is enabled during evaluation to ensure consistent time measurements, with reported wall-clock times excluding initial warm-up iterations. Models are optimized using the Adam optimizer with learning rate 10^{-4} and default momentum parameters without learning-rate decay, with gradients clipped to unit ℓ_2 norm. Over repeated runs, the performance variation of all methods remained below 5%, and the qualitative trends were unchanged. Standard deviations are omitted for clarity.

Baseline Implementations. All baseline models follow the official public implementations of the respective authors, with published architectures, loss functions, and optimization procedures preserved whenever compatible with molecular coordinate data. Each model is adapted to operate on a unified trajectory format and evaluated under an identical protocol.

To enable trajectory-level evaluation, predicted electronic quantities are embedded into a unified BOMD pipeline, where forces are computed using the same PySCF-based electronic-structure implementation and nuclei are propagated under an identical time integration scheme. For methods that involve explicit SCF iterations, electronic calculations are performed using the default PySCF solver configuration with the PBE exchange-correlation functional and the def2-SVP basis set, so that all approaches are evaluated within the same electronic-structure framework. For experiments conducted on GPUs, we used the official GPU library of PySCF named GPU4PySCF with the same parameter settings as in the CPU experiments.

- **DEQHNet**: `https://github.com/Zun-Wang/DEQHNet`
- **QHFlow**: `https://github.com/seongsukim-ml/QHFlow`
- **QHNet (AIRS)**: `https://github.com/divelab/AIRS`
- **SPHNet / WANet**: `https://github.com/microsoft/SPHNet`
- **PhiSNet**: `https://github.com/AIRI-Institute/nablaDFT/tree/main`
- **UMA**: `https://github.com/facebookresearch/fairchem`

Evaluation Protocol. Let $\{\hat{\mathbf{r}}_t\}_{t=0}^{T}$ denote the predicted trajectory and $\{\mathbf{r}_t\}_{t=0}^{T}$ the self-consistent reference trajectory. All geometric metrics are computed by comparing the predicted and reference atomic coordinates at each BOMD simulation step.

For each time step t, the predicted atomic positions $\{\hat{\mathbf{r}}_{t,I}\}_{I=1}^{N}$ are compared with the corresponding reference positions $\{\mathbf{r}_{t,I}\}_{I=1}^{N}$. Per-step errors are evaluated over all atoms and subsequently averaged over the entire horizon. MSE and RMSD are computed from atomic coordinate discrepancies, with MSE measuring direct Cartesian deviations and RMSD quantifying structural differences after optimal rigid-body alignment via the Kabsch algorithm. Specifically, we compute

$$\mathsf{MSE}(t) = \frac{1}{3N}\sum_{I=1}^{N}\|\hat{\mathbf{r}}_{t,I} - \mathbf{r}_{t,I}\|_2^2, \quad \mathsf{RMSD}(t) = \sqrt{\frac{1}{N}\sum_{I=1}^{N}\|R_t^\star\hat{\mathbf{r}}_{t,I} - \mathbf{r}_{t,I}\|_2^2}, \tag{17}$$

where $R_t^\star \in SO(3)$ denotes the optimal rotation aligning the predicted configuration to the reference at step t. We additionally report the pairwise distance error (Lin et al., 2022) and planar deviation (Buttenschoen et al., 2024) defined as

$$\mathsf{DME}(t) = \sum_{I<J}\Big|\|\hat{\mathbf{r}}_{t,I} - \hat{\mathbf{r}}_{t,J}\|_2 - \|\mathbf{r}_{t,I} - \mathbf{r}_{t,J}\|_2\Big|, \quad \mathsf{Planar}(t) = \sum_{I\in\mathcal{A}}\big|\mathbf{n}_t^\top\hat{\mathbf{r}}_{t,I} + d_t\big|, \tag{18}$$

where $(\mathbf{n}_t, d_t)$ denotes the reference plane fitted to the self-consistent configuration.

We further assess global structural compactness through the radius-of-gyration error, which captures the discrepancy in mass-weighted spatial extent about the center of mass between the predicted and reference configurations (Lobanov et al., 2008). Concretely, we first compute the predicted and reference centers of mass,

$$\hat{\mathbf{r}}_{\mathrm{cm}}(t) = \frac{\sum_{I=1}^{N} m_I\hat{\mathbf{r}}_{t,I}}{\sum_{I=1}^{N} m_I}, \qquad \mathbf{r}_{\mathrm{cm}}(t) = \frac{\sum_{I=1}^{N} m_I\mathbf{r}_{t,I}}{\sum_{I=1}^{N} m_I}, \tag{19}$$

and Rg MAE follows the following computation:

$$\mathsf{Rg\ MAE}(t) = \left|\left(\frac{\sum_{I=1}^{N} m_I\|\hat{\mathbf{r}}_{t,I} - \hat{\mathbf{r}}_{\mathrm{cm}}(t)\|_2^2}{\sum_{I=1}^{N} m_I}\right)^{1/2} - \left(\frac{\sum_{I=1}^{N} m_I\|\mathbf{r}_{t,I} - \mathbf{r}_{\mathrm{cm}}(t)\|_2^2}{\sum_{I=1}^{N} m_I}\right)^{1/2}\right|, \tag{20}$$

Inference time is measured as the wall-clock time required to generate a full molecular dynamics trajectory under identical hardware and software environments. For methods involving SCF iterations, the reported time includes the full electronic update cost over the entire trajectory. For methods involving SCF iterations, inference time includes the full electronic update cost over the entire trajectory.

Table 6: **Quantitative Result on Salicylic Acid in MD17.** We report physical accuracy metrics including MSE, RMSD, pairwise distance errors, planar structure preservation, and inference time, the number of SCF iterations for computational efficiency. Best and second-best results are highlighted in **bold** and **blue**, respectively.

Model	Physical Metrics				Computational Efficiency	
	MSE (Å^2)	RMSD (Å)	DME (Å)	Planarity (Å)	Inference (s) (CPU / GPU)	SCF iter (CPU / GPU)
(A) Geometry-aware Hamiltonian Dynamical Models						
NHD-occ (Park, 2025) ($\mathbf{P}_t$)	5.73×10^{-4}	4.13×10^{-2}	2.82×10^{-2}	9.48×10^{-3}	1.26×10^{1} (GPU)	**0** / **0**
NHD-orb (Park, 2025) ($\mathbf{C}_t$)	7.74×10^{-4}	4.79×10^{-2}	3.39×10^{-2}	1.06×10^{-2}	2.21×10^{1} (GPU)	**0** / **0**
(B) Solver Structured and Generative Reformulations						
DEQHNet (Wang et al., 2024)	1.99×10^{-5}	6.96×10^{-3}	5.37×10^{-3}	4.34×10^{-4}	1.25×10^{3} / 2.88×10^{2}	1.70×10^{1} / 1.63×10^{1}
QHFlow (Kim et al., 2025a)	**1.57×10^{-5}**	**6.22×10^{-3}**	5.00×10^{-3}	4.56×10^{-4}	6.31×10^{2} / 1.03×10^{2}	1.76×10^{1} / 1.60×10^{1}
(C) Equivariant Operator and Wavefunction Predictions						
PhiSNet (Unke et al., 2021a)	6.38×10^{-3}	1.19×10^{-1}	5.11×10^{-2}	2.10×10^{-2}	3.10×10^{2} / 1.02×10^{2}	1.53×10^{1} / 1.50×10^{1}
QHNet (Yu et al., 2023)	1.81×10^{-5}	6.61×10^{-3}	5.23×10^{-3}	4.58×10^{-4}	2.96×10^{2} / 9.82×10^{1}	1.47×10^{1} / 1.94×10^{1}
WANet (Li et al., 2025)	2.02×10^{-5}	6.88×10^{-3}	5.27×10^{-3}	4.58×10^{-4}	1.81×10^{3} / 9.82×10^{1}	1.97×10^{1} / 1.94×10^{1}
SPHNet (Luo et al., 2025)	1.66×10^{-5}	6.37×10^{-3}	**4.99×10^{-3}**	**4.02×10^{-4}**	8.67×10^{2} / 6.55×10^{1}	9.46×10^{0} / 9.33×10^{0}
Ours	**8.34×10^{-6}**	**4.95×10^{-3}**	**3.40×10^{-3}**	**1.09×10^{-4}**	**5.03×10^{0}** (GPU)	**0** / **0**

Table 7: **Quantitative Result on Azobenzene in MD17.** We report physical accuracy metrics including MSE, RMSD, pairwise distance errors, planar structure preservation, and inference time, the number of SCF iterations for computational efficiency. Best and second-best results are highlighted in **bold** and **blue**, respectively.

Model	Physical Metrics				Computational Efficiency	
	MSE (Å^2)	RMSD (Å)	DME (Å)	Planarity (Å)	Inference (s) (CPU / GPU)	SCF iter (CPU / GPU)
(A) Geometry-aware Hamiltonian Dynamical Models						
NHD-occ (Park, 2025) ($\mathbf{P}_t$)	5.62×10^{-4}	4.09×10^{-2}	2.84×10^{-2}	6.29×10^{-3}	2.83×10^{1} (GPU)	**0** / **0**
NHD-orb (Park, 2025) ($\mathbf{C}_t$)	6.95×10^{-4}	4.55×10^{-2}	3.16×10^{-2}	7.15×10^{-3}	3.15×10^{1} (GPU)	**0** / **0**
(B) Solver Structured and Generative Reformulations						
DEQHNet (Wang et al., 2024)	1.28×10^{-5}	5.93×10^{-3}	5.49×10^{-3}	3.07×10^{-4}	2.63×10^{3} / 4.05×10^{2}	1.68×10^{1} / 1.66×10^{1}
QHFlow (Kim et al., 2025a)	**1.11×10^{-5}**	**5.46×10^{-3}**	**5.02×10^{-3}**	2.28×10^{-4}	1.82×10^{3} / 1.37×10^{2}	1.53×10^{1} / 1.51×10^{1}
(C) Equivariant Operator and Wavefunction Predictions						
PhiSNet (Unke et al., 2021a)	1.30×10^{-2}	1.23×10^{-1}	5.22×10^{-2}	1.64×10^{-2}	1.50×10^{3} / 1.66×10^{2}	1.67×10^{1} / 1.99×10^{1}
QHNet (Yu et al., 2023)	1.23×10^{-5}	5.71×10^{-3}	5.23×10^{-3}	**1.86×10^{-4}**	1.54×10^{3} / 1.30×10^{2}	1.52×10^{1} / 1.65×10^{1}
WANet (Li et al., 2025)	1.30×10^{-5}	5.84×10^{-3}	5.33×10^{-3}	2.63×10^{-4}	1.54×10^{3} / 8.20×10^{1}	6.52×10^{0} / 6.46×10^{0}
SPHNet (Luo et al., 2025)	1.25×10^{-5}	5.73×10^{-3}	5.27×10^{-3}	1.93×10^{-4}	1.95×10^{3} / 9.87×10^{1}	1.18×10^{1} / 1.19×10^{1}
Ours	**8.26×10^{-6}**	**4.96×10^{-3}**	**3.45×10^{-3}**	**7.06×10^{-5}**	**7.66×10^{0}** (GPU)	**0** / **0**

9 Additional Experiments

To further examine the generality of our method, we extend the evaluation beyond aspirin and Ac-Ala3-NHMe to two additional molecular systems, salicylic acid ($C_7H_6O_3$) and azobenzene ($C_{12}H_{10}N_2$). Together, the four molecular systems provide complementary evaluation cases with different molecular sizes, rigidity, and structural complexity. Aspirin is an organic molecule with multiple functional groups and serves as a representative MD17 benchmark, while Ac-Ala3-NHMe is a larger peptide-like molecule with greater conformational flexibility and is useful for assessing long-term trajectory stability. Salicylic acid provides a compact near-planar aromatic case with relatively constrained geometry, whereas azobenzene introduces a conjugated molecular structure with a flexible azo linkage. These systems allow us to examine whether the proposed SCF-free scheme maintains consistent behavior across both relatively rigid and more flexible molecular settings. For all evaluations, initial molecular structures are obtained from the PubChem database (Kim et al., 2025b). All simulations follow the unified evaluation protocol described in Section 8, using the same force evaluation procedure and SCF configuration as in the reference and comparison runs.

The results for these additional systems are reported in Tables 6 and 7. Across both systems, our method achieves the lowest errors in trajectory propagation and molecular geometry while using zero SCF iterations. On salicylic acid, it reduces MSE by a factor of 1.88 and Planarity by a factor of 3.69 relative to the strongest baseline, with lower RMSD and DME as well. A similar trend appears on azobenzene, where our method reduces MSE by a factor of 1.34 and Planarity by a factor of 2.63 relative to the strongest baseline, while also achieving the best RMSD and DME. Together with the results on aspirin and Ac-Ala3-NHMe, these additional experiments show that the proposed SCF-free propagation scheme maintains consistent physical accuracy across varied molecular settings.

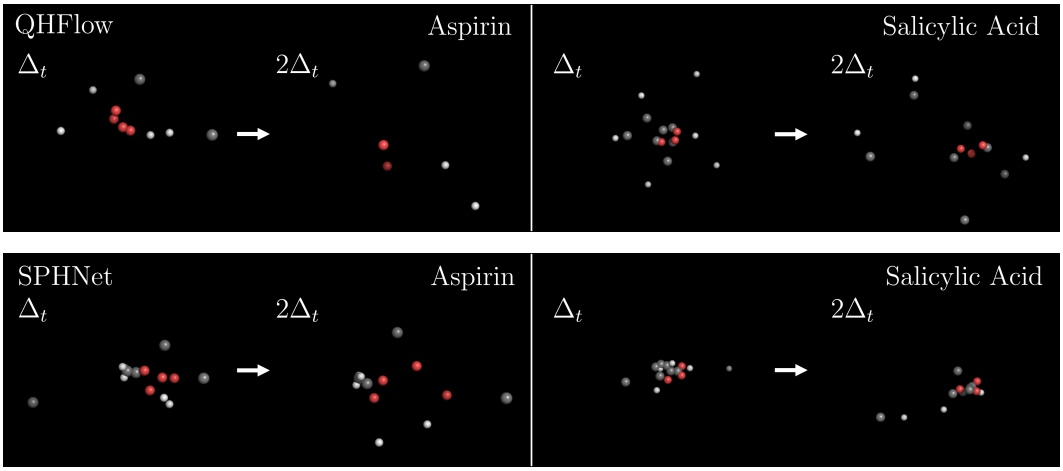


Figure 4: **Qualitative Results on Aspirin and Salicylic Acid under zero SCF iterations.** We visualize molecular dynamics trajectories of aspirin ($C_9H_8O_4$) and salicylic acid ($C_7H_6O_3$) generated by QHFlow and SPHNet without SCF refinement. Under the same propagation from Δ_t to $2\Delta_t$, both baselines exhibit visibly unstable molecular configurations.

Table 8: **Results on Trajectory Length and Temperature.** We report separate evaluations of trajectory length and temperature shifts on Ac-Ala3-NHMe using the proposed method. The reference settings are a trajectory length of 256 and a temperature of 500K, respectively.

Longer trajectories			**Temperature shifts**		
Trajectory length	Inference (s)	MSE (Å^2)	Temperature (K)	MSE (Å^2)	RMSD (Å)
256	6.47	3.59×10^{-6}	480	4.97×10^{-6}	5.63×10^{-3}
512	9.13	8.67×10^{-6}	500	3.59×10^{-6}	3.19×10^{-3}
1,024	12.64	2.20×10^{-4}	520	6.08×10^{-6}	6.46×10^{-3}

Consistent with the main experiments, the additional systems show the same computational advantage from removing the SCF bottleneck. SCF baselines require repeated PySCF refinement, increasing inference cost even when the learned electronic predictor is accurate. By replacing this loop with a fixed electronic update, our method yields $59\times$ and $200\times$ speedups over CPU PySCF on salicylic acid and azobenzene, respectively, and $13\times$ and $11\times$ speedups over PySCF with GPU acceleration. These results suggest that the efficiency gain arises primarily from eliminating the iterative SCF loop at each MD step, rather than from the choice of molecular system.

Furthermore, Figure 4 illustrates why removing SCF refinement is challenging for existing baselines. For QHFlow and SPHNet, the molecular configurations on aspirin and salicylic acid become visibly distorted after the same propagation from Δ_t to $2\Delta_t$. This qualitative behavior is consistent with the zero SCF comparison in Table 2, where the baseline methods show pronounced deterioration in MSE and RMSD. In contrast, the proposed method directly propagates the occupied subspace and density projector, maintaining stable trajectory and geometric accuracy without SCF iterations. Taken together, the additional experiments support gauge-symmetric occupied-subspace propagation as a stable and efficient method for SCF-free molecular dynamics across the evaluated molecular systems.

Effect of Trajectory Length and Temperature. We examine how trajectory length and temperature affect the prediction accuracy of the proposed method on Ac-Ala3-NHMe. Longer trajectories test whether accuracy is maintained during extended simulations, while temperature variations probe sensitivity to changes in the configurational distribution. As shown in Table 8, prediction errors increase with trajectory length, with a more pronounced degradation in the longest setting, consistent with error accumulation during extended simulations. Both temperature-shifted conditions also yield higher MSE and RMSD than the reference condition.

Basis-set Comparison. We further evaluate our method on aspirin using STO-3G and def2-SVP to examine its trajectory prediction accuracy across different AO basis sizes. As a minimal basis, STO-

Table 9: **Basis-set comparison on Aspirin.** We compare trajectory MSE, RMSD, and Planarity error using the compact STO-3G basis and the larger def2-SVP basis.

Basis set	MSE (Å^2)	RMSD (Å)	Planarity (Å)
STO-3G	5.45×10^{-6}	2.68×10^{-3}	6.9×10^{-5}
def2-SVP	7.38×10^{-6}	4.67×10^{-3}	1.03×10^{-4}

Table 10: **Additional Quantitative Results on Aspirin.** We report physical accuracy metrics and electronic-state metrics for UMA, SPHNet, and the proposed method. Best and second-best results are highlighted in **bold** and **blue**, respectively.

Model	**Physical Metrics**			**Electronic-State Metrics**		
	MSE (Å^2)	RMSD (Å)	Planarity (Å)	Coefficient Similarity (%)	Occupied-orbital MAE	HOMO–LUMO gap MAE
UMA (Scratch)	1.30×10^{-5}	5.50×10^{-3}	$\mathbf{1.13 \times 10^{-4}}$	—	—	—
UMA (Fine-tuning)	$\mathbf{9.93 \times 10^{-6}}$	$\mathbf{4.73 \times 10^{-3}}$	3.65×10^{-4}	—	—	—
SPHNet	3.67×10^{-5}	9.64×10^{-3}	3.02×10^{-4}	**97.63**	$\mathbf{1.88 \times 10^{0}}$	$\mathbf{3.54 \times 10^{0}}$
Ours	$\mathbf{7.38 \times 10^{-6}}$	$\mathbf{4.67 \times 10^{-3}}$	$\mathbf{1.03 \times 10^{-4}}$	**99.13**	$\mathbf{9.70 \times 10^{-2}}$	$\mathbf{1.14 \times 10^{-1}}$

3G uses fewer AO basis functions than def2-SVP, leading to a smaller M and reduced dimensions of the overlap, Hamiltonian, and orbital-coefficient matrices. This lowers the computational demands of basis-dependent linear algebra. As shown in Table 9, the STO-3G configuration yields lower errors across the reported trajectory metrics in this comparison.

Trajectory and electronic-state comparison. We compare our method with UMA (Wood et al., 2025), a representative universal MLIP that predicts total energies and atomic forces across diverse chemical systems. This comparison on aspirin assesses whether our explicit electronic-state formulation also achieves competitive trajectory accuracy relative to an MLIP baseline. As shown in Table 10, our method achieves lower MSE and Planarity error than both UMA baselines.

Beyond nuclear trajectories, our method provides explicit electronic-state information that MLIPs such as UMA do not predict. We assess this capability against SPHNet (Luo et al., 2025) using three electronic-state metrics. Coefficient Similarity measures the cosine similarity between predicted and reference occupied-orbital coefficient matrices, while Occupied-orbital MAE is the mean absolute difference between corresponding predicted and reference occupied-orbital energies. HOMO–LUMO gap MAE is the mean absolute difference between predicted and reference gaps, where each gap is the energy separation between the highest occupied and lowest unoccupied orbitals. Our method achieves higher Coefficient Similarity and lower values for both MAE metrics than SPHNet. Taken together, these results support the effectiveness of our framework in achieving accurate molecular dynamics while preserving explicit electronic-state information.

10 Mathematical Preliminary

This section establishes the basic definitions and mathematical framework necessary for our formulation. Let us fix a nuclear configuration $\mathbf{r} \in \Omega \subset \mathbb{R}^{3N_\mathbf{r}}$, and $\mathrm{S}(\mathbf{r}) \in \mathbb{C}^{M\times M}$ denote the overlap matrix induced by the atom-centered basis. We view $\mathbb{C}^{M\times N}$ equipped with the $\mathrm{S}(\mathbf{r})$-weighted real inner product:

$$\langle X, Y\rangle_{\mathrm{S}(\mathbf{r})} \coloneqq \mathrm{Re}\,\mathrm{Tr}\big(X^\dagger \mathrm{S}(\mathbf{r}) Y\big), \quad X, Y \in \mathbf{C}^{M\times N}. \tag{21}$$

Definition 10.1 (Generalized Stiefel manifold)**.** *For fixed $S \succ 0$, the generalized (complex) Stiefel manifold is defined as the set of all S-orthonormal frames:*

$$\mathrm{St}_S(M,N) \coloneqq \big\{\mathbf{C} \in \mathbb{C}^{M\times N} : \mathbf{C}^\dagger S\mathbf{C} = \mathrm{I}_N\big\}. \tag{22}$$

$\mathbf{C} \in \mathrm{St}_S(M,N)$ span an N-dimensional occupied subspace of $\mathbb{C}^M$, representing the physical electronic state variables subject to the orthonormality constraint.

Lemma 10.2 (Tangent space)**.** *The tangent space $T_\mathbf{C}\mathrm{St}_S(M,N)$ is obtained by linearizing the defining constraint $\mathbf{C}^\dagger S\mathbf{C} = \mathrm{I}_N$. A variation Δ belongs to the tangent space if and only if it satisfies:*

$$T_\mathbf{C}\mathrm{St}_S(M,N) = \big\{\Delta \in \mathbb{C}^{M\times N} : \mathbf{C}^\dagger S\Delta + \Delta^\dagger S\mathbf{C} = 0\big\}. \tag{23}$$

Geometrically, this condition implies that the matrix $\mathbf{C}^\dagger S\Delta$ is skew-Hermitian, reflecting the infinitesimal unitary rotation of the frame.

In this paper, we endow the Stiefel manifold $\mathrm{St}_S(M,N)$ with a Riemannian structure by restricting the ambient inner product to the tangent space. Since $\mathbb{C}^{M\times N}$ is equipped with the S-weighted metric, the induced metric on the submanifold is given by:

$$g^{\mathrm{St}}_\mathbf{C}(\Delta_1, \Delta_2) \coloneqq \langle \Delta_1, \Delta_2\rangle_S = \mathrm{Re}\,\mathrm{Tr}(\Delta_1^\dagger S\Delta_2), \quad \Delta_1, \Delta_2 \in T_\mathbf{C}\mathrm{St}_S(M,N). \tag{24}$$

This metric is invariant under the unitary group action discussed below, which is crucial for defining a consistent geometry on the quotient space.

Gauge Symmetry and Principal Bundle Structure. The physical electronic state is determined strictly by the subspace spanned by the orbitals, independent of the specific choice of basis within that subspace. This rotational redundancy is mathematically formalized by the right action of the unitary group $G = \mathrm{U}(N)$ on $\mathrm{St}_S(M,N)$:

$$\mathcal{R}_\mathbf{U}(\mathbf{C}) \coloneqq \mathbf{C}\mathbf{U}, \quad \mathbf{U} \in \mathrm{U}(N). \tag{25}$$

This action is *isometric* with respect to the metric g^{St} and preserves the orthogonality constraint, as $(\mathbf{C}\mathbf{U})^\dagger S(\mathbf{C}\mathbf{U}) = \mathbf{U}^\dagger(\mathbf{C}^\dagger S\mathbf{C})\mathbf{U} = \mathrm{I}_N$. Since the action is free and proper, the Stiefel manifold forms a principal $\mathrm{U}(N)$-bundle over the quotient space.

Definition 10.3 (Generalized Grassmann manifold)**.** *The generalized Grassmann manifold is defined as the quotient space of the Stiefel manifold under the gauge action,*

$$\mathrm{Gr}_S(M,N) \coloneqq \mathrm{St}_S(M,N)/\mathrm{U}(N) = \big\{[\mathbf{C}] : \mathbf{C} \in \mathrm{St}_S(M,N)\big\}, \tag{26}$$

where $[\mathbf{C}] \coloneqq \{\mathbf{C}\mathbf{U} : \mathbf{U} \in \mathrm{U}(N)\}$ denotes the equivalence class of $\mathbf{C}$. Any point $[\mathbf{C}]$ physically represent the occupied subspace.

To provide a concrete realization of this abstract quotient, we introduce the density matrix (projector) representation. Define the map $P : \mathrm{St}_S(M,N) \to \mathbf{C}^{M\times M}$ by $P(\mathbf{C}) \coloneqq \mathbf{C}\mathbf{C}^\dagger S$. This map is invariant under the gauge action, i.e., $P(\mathbf{C}\mathbf{U}) = P(\mathbf{C})$, allowing it to descend to the quotient.

Proposition 10.4 (Projector realization of Gr_S)**.** *The quotient $\mathrm{Gr}_S(M,N)$ is diffeomorphic to the set of rank-N S-orthogonal projectors. Specifically, the map $[\mathbf{C}] \mapsto P(\mathbf{C})$ identifies the Grassmannian with the submanifold*

$$\mathcal{P}_S(M,N) \coloneqq \Big\{P \in \mathbb{C}^{M\times M} : P^2 = P, \quad \mathrm{rank}(P) = N, \quad P^\dagger S = SP\Big\}. \tag{27}$$

Here, the condition $P^\dagger S = SP$ ensures that P is self-adjoint with respect to the S-inner product, and its image corresponds to the subspace $\mathrm{span}(\mathbf{C})$.

Finally, we characterize the tangent structure compatible with this bundle. Let $\pi : \mathrm{St}_S(M,N) \to \mathrm{Gr}_S(M,N)$ be the canonical projection $\pi(\mathbf{C}) = [\mathbf{C}]$. The *vertical space* $\mathcal{V}_\mathbf{C}$ at a point $\mathbf{C}$ consists of tangent vectors purely along the gauge orbit, which corresponds to the kernel of the projection differential:

$$\mathcal{V}_\mathbf{C} \coloneqq \ker(d\pi_\mathbf{C}) = \big\{\mathbf{C}\Omega : \Omega \in \mathfrak{u}(N)\big\}, \tag{28}$$

where $\mathfrak{u}(N) = \{\Omega \in \mathbb{C}^{N\times N} : \Omega^\dagger = -\Omega\}$ is the Lie algebra of the gauge group.

Connection Form and Horizontal Distribution. To disentangle the physical evolution of the subspace from redundant gauge rotations, we invoke the *mechanical connection* induced by the S-metric. This geometric structure defines the vertical component of any tangent vector via the connection 1-form, which extracts the infinitesimal generator of the gauge rotation:

$$\omega_\mathbf{C} : T_\mathbf{C}\mathrm{St}_S(M,N) \to \mathfrak{u}(N), \quad \omega_\mathbf{C}(\Delta) \coloneqq \mathbf{C}^\dagger S\Delta. \tag{29}$$

By Lemma 10.2, the value $\omega_\mathbf{C}(\Delta)$ is always skew-Hermitian, ensuring it lies in the Lie algebra $\mathfrak{u}(N)$. The *horizontal subspace* is then identified as the orthogonal complement to the vertical space. It consists of variations that are purely physical in the sense that they induce no mixing of the basis vectors within the subspace. Formally, it is the kernel of the connection form:

$$\mathcal{H}_\mathbf{C} \coloneqq \ker(\omega_\mathbf{C}) = \big\{\Delta \in T_\mathbf{C}\mathrm{St}_S(M,N) : \mathbf{C}^\dagger S\Delta = 0\big\}. \tag{30}$$

Proposition 10.5 (Orthogonal vertical-horizontal decomposition)**.** *For each $\mathbf{C} \in \mathrm{St}_S(M,N)$, the tangent space decomposes as the g^{St}-orthogonal direct sum*

$$T_\mathbf{C}\mathrm{St}_S(M,N) = \mathcal{V}_\mathbf{C} \oplus \mathcal{H}_\mathbf{C}. \tag{31}$$

Moreover, the vertical and horizontal projections are given explicitly by

$$\mathrm{Ver}_\mathbf{C}(X) = \mathbf{C}\omega_\mathbf{C}(X) = \mathbf{C}(\mathbf{C}^\dagger SX), \quad \mathrm{Hor}_\mathbf{C}(X) = X - \mathrm{Ver}_\mathbf{C}(X) = X - \mathbf{C}(\mathbf{C}^\dagger SX). \tag{32}$$

Equivalently, using the projector $P(\mathbf{C}) = \mathbf{C}\mathbf{C}^\dagger S$, the horizontal projection can be rewritten as

$$\mathrm{Hor}_\mathbf{C}(X) = (I - P(\mathbf{C}))X. \tag{33}$$

There exists a unique Riemannian metric g^{Gr} on $\mathrm{Gr}_S(M,N)$ such that π becomes a Riemannian submersion: for $\xi, \eta \in T_{[\mathbf{C}]}\mathrm{Gr}_S$, choose the unique horizontal lifts $\Delta_\xi, \Delta_\eta \in \mathcal{H}_\mathbf{C}$ satisfying $d\pi_\mathbf{C}(\Delta_\xi) = \xi$, $d\pi_\mathbf{C}(\Delta_\eta) = \eta$, and set

$$g^{\mathrm{Gr}}_{[\mathbf{C}]}(\xi,\eta) \coloneqq g^{\mathrm{St}}_\mathbf{C}(\Delta_\xi, \Delta_\eta).$$

Gauge Potential and Covariant Derivative. Consider a smooth curve $t \mapsto \mathbf{C}_t \in \mathrm{St}_\mathrm{S}(M,N)$. The geometry of the Stiefel manifold induces a natural connection, decomposing the tangent space into horizontal physical and vertical gauge subspaces. In matrix notation, the *gauge potential* $\mathcal{A}(t)$ is the projection of the velocity onto the vertical fiber:

$$\mathcal{A}(t) \coloneqq \mathbf{C}_t^\dagger \mathrm{S}\dot{\mathbf{C}}_t \in \mathfrak{u}(N). \tag{34}$$

The metric tensor is identified as $S_{\mu\nu}$. The components of the gauge potential $\mathcal{A}^i{}_j$ are obtained by contracting the metric with the ambient velocity:

$$\mathcal{A}^i{}_j \coloneqq \bar{\mathbf{C}}^\mu_i S_{\mu\nu} \dot{\mathbf{C}}^\nu_j, \qquad (\nabla_t \mathbf{C})^\mu_i \coloneqq \dot{\mathbf{C}}^\mu_i - \mathbf{C}^\mu_k \mathcal{A}^k{}_i. \tag{35}$$

Geometrically, $\mathcal{A}$ extracts the vertical (pure-gauge) component of $\dot{\mathbf{C}}_t$ along the fiber directions $\mathbf{C}_t\Omega$ ($\Omega \in \mathfrak{u}(N)$), and the covariant derivative $\nabla_t \mathbf{C}_t$ is defined as the corresponding horizontal component. Equivalently, if we introduce the S-orthogonal projector onto the occupied subspace,

$$\mathbf{P}^\mu{}_\nu \coloneqq \mathbf{C}^\mu_k \bar{\mathbf{C}}^\rho_k S_{\rho\nu},$$

then the horizontal projection can be written as

$$(\nabla_t \mathbf{C})^\mu{}_i = \big(\delta^\mu{}_\nu - \mathbf{P}^\mu{}_\nu\big)\dot{\mathbf{C}}^\nu{}_i, \quad \bar{\mathbf{C}}^\mu_i S_{\mu\nu} (\nabla_t \mathbf{C})^\nu{}_j = 0.$$

(We keep the following labeled equation slot for later reference.) Geometrically, this operation filters out the non-physical frame rotations (*i.e.*, the vertical correction) to ensure that $(\nabla_t \mathbf{C})^\mu_i$ represents

the intrinsic transport of the subspace itself. This covariant property is summarized in following result:

Lemma 10.6 (Gauge Covariance and Equivariance)**.** *Consider a smooth change of frame $\mathbf{U}_t \in \mathrm{U}(N)$ and the transformed curve $\widetilde{\mathbf{C}}_t \coloneqq \mathbf{C}_t\mathbf{U}_t$. The gauge potential $\mathcal{A}(t)$ follows the affine transformation law characteristic of a connection form:*

$$\widetilde{\mathcal{A}}(t) = \widetilde{\mathbf{C}}_t^\dagger S\dot{\widetilde{\mathbf{C}}}_t = \mathbf{U}_t^\dagger \mathcal{A}(t)\mathbf{U}_t + \mathbf{U}_t^\dagger \dot{\mathbf{U}}_t. \tag{36}$$

The inhomogeneous term $\mathbf{U}_t^\dagger\dot{\mathbf{U}}_t$ is exactly cancelled by the vertical correction inherent in the definition of ∇_t. Consequently, the covariant derivative transforms equivariantly under the right action of the gauge group:

$$\nabla_t\widetilde{\mathbf{C}}_t = (\nabla_t\mathbf{C}_t)\mathbf{U}_t, \tag{37}$$

implying that the kinetic energy metric is gauge-invariant, i.e., $\|\nabla_t\widetilde{\mathbf{C}}_t\|_S^2 = \|\nabla_t\mathbf{C}_t\|_S^2$. Thus, the dynamics are intrinsic to the subspace $[\mathbf{C}_t] \in \mathrm{Gr}_S(M,N)$ and independent of the choice of basis.

Musical isomorphisms and canonical momentum. To formulate the dynamics in the phase space, we require a mapping between the tangent bundle (velocities) and the cotangent bundle (momenta). The Riemannian metric S induces the canonical *musical isomorphisms*, which identify the vector space $\mathbb{C}^{M\times N}$ with its dual $(\mathbb{C}^{M\times N})^*$. Explicitly, the "flat" ($\flat$) and "sharp" ($\sharp$) maps are given by:

$$(\cdot)^\flat : X \mapsto SX, \quad (\cdot)^\sharp : \Pi \mapsto S^{-1}\Pi.$$

Using these isomorphisms, we define the *canonical momentum* Π_t conjugate to the generalized velocity. Since the dynamics is constrained to be horizontal (gauge-covariant), the momentum is strictly the dual of the horizontal velocity $\nabla_t\mathbf{C}_t$:

$$\Pi_t \coloneqq (\nabla_t\mathbf{C}_t)^\flat = S\nabla_t\mathbf{C}_t, \quad \Pi_t^\sharp = S^{-1}\Pi_t = \nabla_t\mathbf{C}_t. \tag{38}$$

A crucial property of this construction is that the momentum inherits the horizontal character of the velocity. Specifically, one verifies that $\mathbf{C}_t^\dagger\Pi_t = \mathbf{C}_t^\dagger S\nabla_t\mathbf{C}_t = 0$, which implies that Π_t contains no component along the gauge orbit. We now translate this frame dynamics into the evolution of the physical observable, the density matrix P_t. Let the occupied-subspace projector be $P_t = P(\mathbf{C}_t) = \mathbf{C}_t\mathbf{C}_t^\dagger S$. The horizontal evolution of the frame $\mathbf{C}_t$ induces a well-defined transport equation for the projector:

$$\nabla_t P_t \coloneqq (\nabla_t\mathbf{C}_t)\mathbf{C}_t^\dagger S + \mathbf{C}_t(\nabla_t\mathbf{C}_t)^\dagger S = \Pi_t^\sharp\mathbf{C}_t^\dagger S + \mathbf{C}_t(\Pi_t^\sharp)^\dagger S. \tag{39}$$

The right-hand side is manifestly invariant under the gauge transformation $\mathbf{C}_t \mapsto \mathbf{C}_t\mathbf{U}_t$, confirming that the dynamics depends strictly on the evolving physical state $[\mathbf{C}_t] \in \mathrm{Gr}_S$, not on the choice of basis. Finally, we remark on the geometric structure of the phase space. While the ambient space $\mathbb{C}^{M\times N}$ is complex, the physical configuration space $\mathrm{Gr}_S(M,N)$ carries a richer structure than just a Riemannian manifold; it is naturally a *Kähler manifold*. The Hermitian form defining the geometry decomposes into real and imaginary parts:

$$h_S(X,Y) = \mathrm{Tr}(X^\dagger SY) = g_S(X,Y) + i\omega_S(X,Y).$$

Here, the real part is the Riemannian metric used above, while the imaginary part, $\omega_S(X,Y) = \mathrm{Im}\,\mathrm{Tr}(X^\dagger SY)$, defines the *symplectic form*. This symplectic structure is essential for defining the Hamiltonian flow and ensures that the projected quantum dynamics preserves the geometric phase properties of the electronic wavefunction.

Symplectic Geometry and Lagrangian Mechanics. A *symplectic manifold* is a pair $(\mathcal{M},\omega)$, where $\mathcal{M}$ is a smooth manifold and $\omega \in \Omega^2(\mathcal{M})$ is a closed, *i.e.*, $d\omega = 0$, and non-degenerate differential 2-form. Given a smooth function $H : \mathcal{M} \to \mathbb{R}$, called the Hamiltonian, the corresponding *Hamiltonian vector field* X_H is uniquely determined by the intrinsic condition:

$$\iota_{X_H}\omega = dH. \tag{40}$$

The flow generated by X_H preserves the symplectic structure ω. The most prominent example of a symplectic manifold is the *cotangent bundle* T^*Q of a smooth configuration manifold Q. It

naturally carries the *canonical symplectic form*. Let Θ be the *Liouville 1-form*, defined pointwise at $(q,p) \in T^*Q$ by:

$$\Theta_{(q,p)}(\delta q, \delta p) \coloneqq \langle p, \delta q \rangle. \tag{41}$$

The canonical symplectic form is then defined as $\omega \coloneqq -d\Theta$. In local coordinates (q^i, p_i) on T^*Q, these forms are expressed as:

$$\Theta = \sum_i p_i dq^i, \quad \omega = \sum_i dq^i \wedge dp_i. \tag{42}$$

Consequently, Hamilton's equations of motion are simply represented as:

$$\dot{q}^i = \frac{\partial H}{\partial p_i}, \quad \dot{p}_i = -\frac{\partial H}{\partial q^i}. \tag{43}$$

To relate this geometric framework to classical mechanics, let Q be a smooth configuration manifold and $L : TQ \times \mathbb{R} \to \mathbb{R}$ be a *Lagrangian function*, mapping $(q, \dot{q}, t) \mapsto L(q, \dot{q}, t)$. A fundamental Hamilton's principle states that the physical path corresponds to a critical point of the action functional, which leads to the Euler-Lagrange equations. When the Lagrangian L is regular (*i.e.*, the Hessian with respect to velocities $\partial_{\dot{q}} L$ is invertible), the *Legendre transform* $L^\star : TQ \to T^*Q$ defined by

$$L^\star(q, \dot{q}) = (q, p), \quad \text{where} \quad p \coloneqq \frac{\partial L}{\partial \dot{q}}(q, \dot{q}, t), \tag{44}$$

is a local diffeomorphism onto its image. This ultimately defines the Hamiltonian function $H : T^*Q \times \mathbb{R} \to \mathbb{R}$:

$$H(q, p, t) \coloneqq \langle p, \dot{q} \rangle - L(q, \dot{q}, t), \tag{45}$$

where $\dot{q}$ is implicitly defined by $(q, p) = L^\star(q, \dot{q})$. Under these conditions, the Euler-Lagrange dynamics on TQ are equivalent to the Hamiltonian flow on the symplectic manifold (T^*Q, ω).

Lagrange-d'Alembert Principle on Electronic Configuration Space. For systems subject to external influences, we introduce a non-conservative generalized force field, defined as a map $F_{\mathrm{nc}} : TQ \times [0,T] \to T^*Q$. The *Lagrange-d'Alembert principle* (Marsden and Ratiu, 1999) generalizes Hamilton's principle: it states that a curve $q : [0,T] \to Q$ is a physical trajectory if, for every smooth variation $\delta q(t) \in T_{q(t)}Q$, the following variational equation holds:

$$\delta \int_0^T L(q, \dot{q}, t) dt + \int_0^T \langle F_{\mathrm{nc}}(q, \dot{q}, t), \delta q \rangle dt = 0. \tag{46}$$

Here δ is called the *first variation* (*i.e.*, Gâteaux derivative) of the action functional along an endpoint-fixed variation defined by following procedure. We first choose any smooth map $\Gamma : (-\varepsilon_0, \varepsilon_0) \times [0,T] \to Q$ for $q_\varepsilon(t) \coloneqq \Gamma(\varepsilon, t)$ with $q_0 = q$ and $q_\varepsilon(0) = q(0)$, $q_\varepsilon(T) = q(T)$, and the induced variation field is

$$\delta q(t) \coloneqq \partial_\varepsilon q_\varepsilon(t)\big|_{\varepsilon=0} \in T_{q(t)}Q, \quad \delta q(0) = \delta q(T) = 0,$$

Then, we have the induced velocity variation given as

$$\delta \dot{q}(t) \coloneqq \partial_\varepsilon \dot{q}_\varepsilon(t)\big|_{\varepsilon=0} = \frac{d}{dt}\delta q(t)$$

Accordingly, the equation (46) is read as

$$\delta \int_0^T L(q, \dot{q}, t) dt \coloneqq \left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} \int_0^T L\big(q_\varepsilon(t), \dot{q}_\varepsilon(t), t\big) dt = d\mathcal{S}(q)[\delta q], \;\; \mathcal{S}(q) \coloneqq \int_0^T L(q, \dot{q}, t) dt.$$

In our electronic setting, we specialize the aforementioned variational framework by lifting the dynamics to the time-dependent Stiefel manifold. With the nuclear path $t \mapsto \mathbf{r}_t$, the overlap matrix $\mathrm{S}_t \coloneqq \mathrm{S}(\mathbf{r}_t)$ evolves in time, defining the configuration space as:

$$Q_t \coloneqq \mathrm{St}_{\mathrm{S}_t}(M, N) = \left\{ \mathbf{C} \in \mathbb{C}^{M \times N} \mid \mathbf{C}^\dagger \mathrm{S}_t \mathbf{C} = \mathrm{I}_N \right\}. \tag{47}$$

The transition from the abstract configuration $q(t)$ to the electronic coefficient matrix $\mathbf{C}_t \in \mathrm{St}_{\mathrm{S}_t}$ requires a careful definition of velocity to account for gauge redundancy. We lift the generalized coordinates and velocities via the following mapping:

$$q(t) \mapsto \mathbf{C}_t, \quad \dot{q}(t) \mapsto \mathbf{V}_t \coloneqq \nabla_t \mathbf{C}_t, \quad \delta q(t) \mapsto \delta \mathbf{C}_t \in T_{\mathbf{C}_t} \mathrm{St}_{\mathrm{S}_t}(M, N) \tag{48}$$

Similarly, one can replace the Lagrangian L and non-conservative force field by introducing

$$L\big(q,\dot{q},t\big)\mapsto\mathcal{L}\big(\mathbf{C},\mathbf{V},\mathbf{P};\mathbf{r}_t\big),\quad F_{\mathrm{nc}}\big(q,\dot{q},t\big)\mapsto\mathbf{F}_{\mathrm{nc}}\big(\mathbf{C},\mathbf{V};\mathbf{r}_t\big)\in T^{*}_{\mathbf{C}}\mathrm{St}_{\mathrm{S}_t}(M,N)\tag{49}$$

Overall, the Lagrange-d'Alembert principle on electronic configuration space directly states the following

$$\delta\int_0^T\mathcal{L}\big(\mathbf{C}_t,\nabla_t\mathbf{C}_t,\mathbf{P}_t;\mathbf{r}_t\big)dt+\int_0^T\big\langle\mathbf{F}_{\mathrm{nc}}\big(\mathbf{C}_t,\nabla_t\mathbf{C}_t;\mathbf{r}_t\big),\,\delta\mathbf{C}_t\big\rangle dt=0,\quad\delta\mathbf{C}_0=\delta\mathbf{C}_T=\mathbf{0}.\tag{50}$$

By setting $\mathbf{F}_{\mathrm{nc}}=-\partial_{\mathbf{V}}\mathcal{D}$ and $\mathbf{V}=\nabla_t\mathbf{C}$, we obtain the simplified version of Lagrangian equation as follows:

$$\delta\int_0^T\mathcal{L}dt-\int_0^T\big\langle\partial_{\nabla_t\mathbf{C}}\mathcal{D},\,\delta\mathbf{C}_t\big\rangle dt=0.\tag{51}$$

11 Proofs

Proposition 11.1 (Lifted Dissipative Variational Principle)**.** *Let us consider the dissipative dynamics on the time-dependent Stiefel manifold governed by the following variational principle:*

$$\mathcal{L}(\mathbf{C}_t, \nabla_t\mathbf{C}_t, \mathbf{P}_t; \mathbf{r}_t) \coloneqq \frac{\mu_t}{2}\|\nabla_t\mathbf{C}_t\|^2_{\mathrm{S}_t} - \mathcal{E}(\mathbf{r}_t, \mathbf{P}_t), \quad \mathcal{D}(\mathbf{C}_t, \nabla_t\mathbf{C}_t; \mathbf{r}_t) \coloneqq \frac{\gamma}{2}\|\nabla_t\mathbf{C}_t\|^2_{\mathrm{S}_t}. \quad (52)$$

Then, the lifted version of Lagrange-d'Alembert equation in (50) *can be rewritten as follows:*

$$\partial_{\mathbf{C}}\mathcal{L} - \nabla_t\big(\partial_{\nabla_t\mathbf{C}}\mathcal{L}\big) - \partial_{\nabla_t\mathbf{C}}\mathcal{D} = \mathbf{0} \quad (53)$$

Here, $\mathbf{P}_t = \mathbf{C}_t\mathbf{C}_t^\dagger\mathrm{S}_t$ is the gauge-invariant projector, and $\langle\cdot,\cdot\rangle$ denotes the natural duality pairing between a covector and a tangent vector at $\mathbf{C}_t$.

Proof. In the first part of the proof, we develop local coordinate expression to rewrite the Lagrangian dynamical system explicitly. Let $\pi : \mathrm{St}_{\mathrm{S}_t}(M,N) \to \mathrm{Gr}_{\mathrm{S}_t}(M,N)$ denote the canonical projection $\pi(\mathbf{C}) = \mathbf{P}(\mathbf{C}) = \mathbf{C}\mathbf{C}^\dagger\mathrm{S}_t$. For the fixed chart $\psi_t : V \to \mathrm{Gr}_{\mathrm{S}_t}(M,N)$ and a smooth local section $\sigma_t : V \to \mathrm{St}_{\mathrm{S}_t}(M,N)$ satisfying $\pi\circ\sigma_t = \psi_t$, we write $\widetilde{\mathbf{C}}(x,t) \coloneqq \sigma_t(x)$. Then, the principal bundle $\mathrm{U}(N) \hookrightarrow \mathrm{St}_{\mathrm{S}_t}(M,N) \xrightarrow{\pi} \mathrm{Gr}_{\mathrm{S}_t}(M,N)$ admits a local trivialization

$$\Phi_t : V \times \mathrm{U}(N) \to \pi^{-1}\big(\psi_t(V)\big), \quad \Phi_t(x,\mathbf{U}) = \widetilde{\mathbf{C}}(x,t)\mathbf{U}.$$

Consequently, any smooth one-parameter family of Stiefel curves $\mathbf{C}_\varepsilon : [0,T] \to \pi^{-1}\big(\psi_t(V)\big)$ admits a bundle decomposition

$$\mathbf{C}_\varepsilon(t) = \widetilde{\mathbf{C}}\big(x_\varepsilon(t),t\big)\mathbf{U}_\varepsilon(t), \quad x_\varepsilon(t) \coloneqq \psi_t^{-1}\big(\pi(\mathbf{C}_\varepsilon(t))\big), \quad \mathbf{U}_\varepsilon(t) \coloneqq \widetilde{\mathbf{C}}\big(x_\varepsilon(t),t\big)^\dagger\mathrm{S}_t\mathbf{C}_\varepsilon(t).$$

If desired, one may further introduce local coordinates $\theta_\varepsilon(t)$ on $\mathrm{U}(N)$ and write $\mathbf{U}_\varepsilon(t) = \mathbf{U}\big(\theta_\varepsilon(t)\big)$, yielding the parametrization $\mathbf{C}_\varepsilon(t) = \mathbf{C}\big(x_\varepsilon(t),\theta_\varepsilon(t),t\big) \coloneqq \widetilde{\mathbf{C}}\big(x_\varepsilon(t),t\big)\mathbf{U}\big(\theta_\varepsilon(t)\big)$. Let us assume fixed endpoints in Stiefel manifold so that $\mathbf{C}_\varepsilon(0) = \mathbf{C}(0)$ and $\mathbf{C}_\varepsilon(T) = \mathbf{C}(T)$. Let $\mathbf{V}_t \coloneqq \nabla_t\mathbf{C}_t \in T_{\mathbf{C}_t}\mathrm{St}_{\mathrm{S}_t}$ be the covariant velocity and similarly $\mathbf{V}_{\varepsilon,t} \coloneqq \nabla_t\mathbf{C}_\varepsilon(t)$, so that $\delta\mathbf{V}_t \coloneqq \partial_\varepsilon\mathbf{V}_{\varepsilon,t}|_{\varepsilon=0}$. With these derived formulas in one hand, our next step is to explicitly derive the variational of action functional. For the action functional $\mathcal{S}[\mathbf{C}] \coloneqq \int_0^T \mathcal{L}(\mathbf{C}_t,\mathbf{V}_t,\mathbf{P}_t;\mathbf{r}_t)dt$ first variation can be written as

$$\begin{aligned}
\delta\int_0^T \mathcal{L}dt &\coloneqq \frac{d}{d\varepsilon}\Big|_{\varepsilon=0}\int_0^T \mathcal{L}\big(\mathbf{C}_\varepsilon(t),\mathbf{V}_\varepsilon(t),\mathbf{P}(\mathbf{C}_\varepsilon(t));\mathbf{r}_t\big)dt\\
&= \int_0^T \frac{d}{d\varepsilon}\Big|_{\varepsilon=0}\mathcal{L}\big(\mathbf{C}_\varepsilon(t),\mathbf{V}_\varepsilon(t),\mathbf{P}(\mathbf{C}_\varepsilon(t));\mathbf{r}_t\big)dt\\
&= \int_0^T \Big\langle\partial_{\mathbf{C}}\mathcal{L},\partial_\varepsilon\mathbf{C}_\varepsilon(t)\Big|_{\varepsilon=0}\Big\rangle + \Big\langle\partial_{\mathbf{V}}\mathcal{L},\partial_\varepsilon\mathbf{V}_\varepsilon(t)\Big|_{\varepsilon=0}\Big\rangle + \Big\langle\partial_{\mathbf{P}}\mathcal{L},\partial_\varepsilon\mathbf{P}(\mathbf{C}_\varepsilon(t))\Big|_{\varepsilon=0}\Big\rangle dt\\
&= \int_0^T \Big\langle\partial_{\mathbf{C}}\mathcal{L},\delta\mathbf{C}_t\Big\rangle + \Big\langle\partial_{\mathbf{V}}\mathcal{L},\delta\mathbf{V}_t\Big\rangle + \Big\langle\partial_{\mathbf{P}}\mathcal{L},D\mathbf{P}(\mathbf{C}_t)\big[\delta\mathbf{C}_t\big]\Big\rangle dt\\
&= \int_0^T \Big\langle\partial_{\mathbf{C}}\mathcal{L} + D\mathbf{P}(\mathbf{C}_t)^*\big[\partial_{\mathbf{P}}\mathcal{L}\big],\delta\mathbf{C}_t\Big\rangle + \Big\langle\partial_{\mathbf{V}}\mathcal{L},\delta\mathbf{V}_t\Big\rangle dt\\
&= \int_0^T \Big(\big\langle\partial_{\mathbf{C}}\mathcal{L},\delta\mathbf{C}_t\big\rangle + \big\langle\partial_{\mathbf{V}}\mathcal{L},\delta\mathbf{V}_t\big\rangle\Big)dt, \qquad (54)
\end{aligned}$$

where $\partial_{\mathbf{V}}\mathcal{L}$ denotes the fiber derivative and a covector in $T^*_{\mathbf{C}_t}\mathrm{St}_{\mathrm{S}_t}$ characterized by $D_{\mathbf{V}}\mathcal{L}[\mathbf{W}] = \langle\partial_{\mathbf{V}}\mathcal{L},\mathbf{W}\rangle$ for all $\mathbf{W} \in T_{\mathbf{C}_t}\mathrm{St}_{\mathrm{S}_t}$, and $\partial_{\mathbf{C}}\mathcal{L}$ is the corresponding covector for variations of $\mathbf{C}_t$ along the tangent space. For simplicity let $\mathbf{\Pi}_t \coloneqq \partial_{\mathbf{V}}\mathcal{L}$ be the variation of Lagrangian in $\mathbf{V}$-axis. Following

by the result obtained in (54), one can rewrite the

$$
\begin{aligned}
&\delta \int_0^T \mathcal{L} dt - \int_0^T \Big\langle \partial_{\nabla_t \mathbf{C}} \mathcal{D},\ \delta \mathbf{C}_t \Big\rangle dt \\
&= \int_0^T \Big(\big\langle \partial_{\mathbf{C}} \mathcal{L}, \delta \mathbf{C}_t \big\rangle + \big\langle \partial_{\mathbf{V}} \mathcal{L}, \delta \mathbf{V}_t \big\rangle \Big) dt - \int_0^T \Big\langle \partial_{\mathbf{V}} \mathcal{D},\ \delta \mathbf{C}_t \Big\rangle dt \\
&= \int_0^T \Big(\big\langle \partial_{\mathbf{C}} \mathcal{L}, \delta \mathbf{C}_t \big\rangle + \big\langle \partial_{\mathbf{V}} \mathcal{L}, \nabla_t \delta \mathbf{C}_t \big\rangle - \big\langle \partial_{\mathbf{V}} \mathcal{D}, \delta \mathbf{C}_t \big\rangle \Big) dt \\
&= \int_0^T \Big(\big\langle \partial_{\mathbf{C}} \mathcal{L} - \partial_{\mathbf{V}} \mathcal{D}, \delta \mathbf{C}_t \big\rangle + \big\langle \mathbf{\Pi}_t, \nabla_t \delta \mathbf{C}_t \big\rangle \Big) dt, \\
&= \int_0^T \Big(\big\langle \partial_{\mathbf{C}} \mathcal{L} - \partial_{\mathbf{V}} \mathcal{D}, \delta \mathbf{C}_t \big\rangle + \frac{d}{dt} \big\langle \mathbf{\Pi}_t, \delta \mathbf{C}_t \big\rangle - \big\langle \nabla_t \mathbf{\Pi}_t, \delta \mathbf{C}_t \big\rangle \Big) dt \\
&= \int_0^T \big\langle \partial_{\mathbf{C}} \mathcal{L} - \nabla_t \mathbf{\Pi}_t - \partial_{\mathbf{V}} \mathcal{D},\ \delta \mathbf{C}_t \big\rangle dt + \underbrace{\big\langle \mathbf{\Pi}_t, \delta \mathbf{C}_t \big\rangle \Big|_0^T}_{=\,0} \\
&= \int_0^T \big\langle \partial_{\mathbf{C}} \mathcal{L} - \nabla_t \big(\partial_{\mathbf{V}} \mathcal{L} \big) - \partial_{\nabla_t \mathbf{C}} \mathcal{D},\ \delta \mathbf{C}_t \big\rangle dt \\
&= 0.
\end{aligned}
$$

where we use the commutation of variation and covariant derivative $\delta \mathbf{V}_t \coloneqq \delta[\nabla_t \mathbf{C}_t] = \nabla_t \delta \mathbf{C}_t$ proved in Lemma 11.2 (colored in red). Following by the result, for every points $\mathbf{C}_t$ and $t \in [0, T]$, we obtain the desired result:

$$
\partial_{\mathbf{C}} \mathcal{L} - \nabla_t \big(\partial_{\mathbf{V}} \mathcal{L} \big) - \partial_{\nabla_t \mathbf{C}} \mathcal{D} = \mathbf{0} \quad \text{in } T^*_{\mathbf{C}_t} \mathrm{St}_{\mathrm{S}_t}(M, N), \tag{55}
$$

□

Lemma 11.2. *(Commutation of variation and covariant derivative) For a smooth variation surface $(t, \varepsilon) \mapsto \mathbf{C}_\varepsilon(t)$ on $\mathrm{St}_{\mathrm{S}_t}(M, N)$ with $\partial_\varepsilon \mathrm{S}_t = 0$, define the covariant velocity*

$$
\mathbf{V}_{\varepsilon,t} \coloneqq \nabla_t \mathbf{C}_\varepsilon(t), \quad \mathbf{V}_t \coloneqq \nabla_t \mathbf{C}_t, \quad \delta \mathbf{V}_t \coloneqq \partial_\varepsilon \mathbf{V}_{\varepsilon,t} \big|_{\varepsilon=0}.
$$

Then the induced velocity variation satisfies the identity

$$
\delta \mathbf{V}_t \coloneqq \delta \big[\nabla_t \mathbf{C}_t \big] = \nabla_t \delta \mathbf{C}_t, \quad \delta \mathbf{C}_t \coloneqq \partial_\varepsilon \mathbf{C}_\varepsilon(t) \big|_{\varepsilon=0}. \tag{56}
$$

Proof. In the local bundle coordinates, we assume that

$$
\delta x^\alpha(0) = 0, \quad \delta x^\alpha(T) = 0, \quad \delta \theta^\beta(0) = 0, \quad \delta \theta^\beta(T) = 0, \quad \delta x^\alpha \coloneqq \partial_\varepsilon x^\alpha_\varepsilon \big|_{\varepsilon=0}, \quad \delta \theta^\beta \coloneqq \partial_\varepsilon \theta^\beta_\varepsilon \big|_{\varepsilon=0}.
$$

The variation field decomposes into base and fiber contributions:

$$
\begin{aligned}
\delta \mathbf{C}_t &\coloneqq \partial_\varepsilon \mathbf{C}_\varepsilon(t) \Big|_{\varepsilon=0} = \partial_\varepsilon \Big(\widetilde{\mathbf{C}} \big(x_\varepsilon(t), t \big) \mathbf{U} \big(\theta_\varepsilon(t) \big) \Big) \Big|_{\varepsilon=0} \\
&= \Big(\partial_\varepsilon \widetilde{\mathbf{C}} \big(x_\varepsilon(t), t \big) \mathbf{U} \big(\theta_\varepsilon(t) \big) + \widetilde{\mathbf{C}} \big(x_\varepsilon(t), t \big) \partial_\varepsilon \mathbf{U} \big(\theta_\varepsilon(t) \big) \Big) \Big|_{\varepsilon=0} \\
&= \Big(\partial_\alpha \widetilde{\mathbf{C}} \big(x_\varepsilon(t), t \big) \partial_\varepsilon x^\alpha_\varepsilon(t) \mathbf{U} \big(\theta_\varepsilon(t) \big) + \widetilde{\mathbf{C}} \big(x_\varepsilon(t), t \big) \partial_\varepsilon \mathbf{U} \big(\theta_\varepsilon(t) \big) \Big) \Big|_{\varepsilon=0} \\
&= \partial_\alpha \widetilde{\mathbf{C}} \big(x(t), t \big) \delta x^\alpha(t) \mathbf{U} \big(\theta(t) \big) + \widetilde{\mathbf{C}} \big(x(t), t \big) \partial_\varepsilon \mathbf{U} \big(\theta_\varepsilon(t) \big) \Big|_{\varepsilon=0} \\
&= \partial_\alpha \widetilde{\mathbf{C}} \big(x(t), t \big) \delta x^\alpha(t) \mathbf{U}(t) + \widetilde{\mathbf{C}} \big(x(t), t \big) \mathbf{U}(t) \Big(\mathbf{U}(t)^\dagger \partial_\varepsilon \mathbf{U} \big(\theta_\varepsilon(t) \big) \Big|_{\varepsilon=0} \Big) \\
&= \partial_\alpha \widetilde{\mathbf{C}} \big(x(t), t \big) \delta x^\alpha(t) \mathbf{U}(t) + \widetilde{\mathbf{C}} \big(x(t), t \big) \mathbf{U}(t) \mathbf{\Omega}_t \\
&= \partial_\alpha \widetilde{\mathbf{C}} \big(x(t), t \big) \delta x^\alpha(t) \mathbf{U}(t) + \mathbf{C}_t \mathbf{\Omega}_t,
\end{aligned} \tag{57}
$$

where $\mathbf{U}(t) \coloneqq \mathbf{U}(\theta(t))$ and the fiber variation is written in right-trivialized form

$$\mathbf{\Omega}_t \coloneqq \mathbf{U}(t)^\dagger \partial_\varepsilon \mathbf{U}\big(\theta_\varepsilon(t)\big)\big|_{\varepsilon=0} \in \mathfrak{u}(N), \quad \mathbf{\Omega}_t^\dagger = -\mathbf{\Omega}_t.$$

Since S_t depends on t but not on ε, one has $\partial_\varepsilon \mathrm{S}_t = 0$. Taking the Gâteaux directional derivative with respect to ε at $\varepsilon = 0$ of the Stiefel constraint $\mathbf{C}_\varepsilon(t)^\dagger \mathrm{S}_t \mathbf{C}_\varepsilon(t) = \mathbf{I}_N$ yields

$$\begin{aligned}
&\partial_\varepsilon \Big(\mathbf{C}_\varepsilon(t)^\dagger \mathrm{S}_t \mathbf{C}_\varepsilon(t) \Big)\Big|_{\varepsilon=0} \\
&= \Big(\partial_\varepsilon \mathbf{C}_\varepsilon(t)\Big)^\dagger\Big|_{\varepsilon=0} \mathrm{S}_t \mathbf{C}_t + \mathbf{C}_t^\dagger \mathrm{S}_t \Big(\partial_\varepsilon \mathbf{C}_\varepsilon(t)\Big)\Big|_{\varepsilon=0} \\
&= (\delta \mathbf{C}_t)^\dagger \mathrm{S}_t \mathbf{C}_t + \mathbf{C}_t^\dagger \mathrm{S}_t \delta \mathbf{C}_t \\
&= \Big(\partial_\alpha \widetilde{\mathbf{C}}\big(x(t),t\big)\delta x^\alpha(t)\mathbf{U}(t) + \mathbf{C}_t \mathbf{\Omega}_t\Big)^\dagger \mathrm{S}_t \mathbf{C}_t + \mathbf{C}_t^\dagger \mathrm{S}_t \Big(\partial_\alpha \widetilde{\mathbf{C}}\big(x(t),t\big)\delta x^\alpha(t)\mathbf{U}(t) + \mathbf{C}_t \mathbf{\Omega}_t\Big) \\
&= \mathbf{U}(t)^\dagger \Big(\partial_\alpha \widetilde{\mathbf{C}}\big(x(t),t\big)\Big)^\dagger \mathrm{S}_t \widetilde{\mathbf{C}}\big(x(t),t\big)\mathbf{U}(t)\delta x^\alpha(t) + \mathbf{\Omega}_t^\dagger \mathbf{C}_t^\dagger \mathrm{S}_t \mathbf{C}_t \\
&\quad + \mathbf{U}(t)^\dagger \widetilde{\mathbf{C}}\big(x(t),t\big)^\dagger \mathrm{S}_t \Big(\partial_\alpha \widetilde{\mathbf{C}}\big(x(t),t\big)\Big)\mathbf{U}(t)\delta x^\alpha(t) + \mathbf{C}_t^\dagger \mathrm{S}_t \mathbf{C}_t \mathbf{\Omega}_t \\
&= \mathbf{U}(t)^\dagger \Big[\Big(\partial_\alpha \widetilde{\mathbf{C}}\big(x(t),t\big)\Big)^\dagger \mathrm{S}_t \widetilde{\mathbf{C}}\big(x(t),t\big) + \widetilde{\mathbf{C}}\big(x(t),t\big)^\dagger \mathrm{S}_t \Big(\partial_\alpha \widetilde{\mathbf{C}}\big(x(t),t\big)\Big)\Big]\mathbf{U}(t)\delta x^\alpha(t) \\
&\quad + \underbrace{\mathbf{\Omega}_t^\dagger + \mathbf{\Omega}_t}_{=\,\mathbf{0}} \\
&= \mathbf{U}(t)^\dagger \underbrace{\Big[\Big(\partial_\alpha \widetilde{\mathbf{C}}\big(x(t),t\big)\Big)^\dagger \mathrm{S}_t \widetilde{\mathbf{C}}\big(x(t),t\big) + \widetilde{\mathbf{C}}\big(x(t),t\big)^\dagger \mathrm{S}_t \Big(\partial_\alpha \widetilde{\mathbf{C}}\big(x(t),t\big)\Big)\Big]}_{=\,\mathbf{0}} \mathbf{U}(t)\delta x^\alpha(t) \\
&= 0.
\end{aligned} \tag{58}$$

In the last two lines remaining terms cancel since $\mathbf{\Omega}_t \in \mathfrak{u}(N)$, and the bracket vanishes since $x \mapsto \widetilde{\mathbf{C}}(x,t)$ takes values in $\mathrm{St}_{\mathrm{S}_t}$. Hence, the variation $\delta \mathbf{C}_t$ is well-defined on the tangent space

$$\delta \mathbf{C}_t \in T_{\mathbf{C}_t}\mathrm{St}_{\mathrm{S}_t}(M,N) = \Big\{ \Delta \in \mathbb{C}^{M\times N} : \mathbf{C}_t^\dagger \mathrm{S}_t \Delta + \Delta^\dagger \mathrm{S}_t \mathbf{C}_t = 0 \Big\}.$$

where covariant derivative can be further developed as $\nabla_t = \mathrm{Hor}_{\mathbf{C}_t} \circ \frac{d}{dt}$. Next, we take variations with respect to ε at $\varepsilon = 0$ for further discussion. Since the coordinate vector fields ∂_t and ∂_ε commute on the surface, we can interchange the derivatives. Moreover, S_t depends on t but not on ε, hence $\partial_\varepsilon S_{\rho\sigma}(t) = 0$:

$$\delta \mathbf{C}_{t\,i}^{\mu} = A \partial_\varepsilon \mathbf{C}_{\varepsilon\,i}^{\mu}(t)\big|_{\varepsilon=0}, \tag{59}$$

$$\delta \dot{\mathbf{C}}_{t\,i}^{\mu} = \partial_\varepsilon \dot{\mathbf{C}}_{\varepsilon\,i}^{\mu}(t)\big|_{\varepsilon=0} = \partial_t \delta \mathbf{C}_{t\,i}^{\mu}, \tag{60}$$

$$\delta A_{t\,i}^{j} = \partial_\varepsilon A_{\varepsilon\,i}^{j}(t)\big|_{\varepsilon=0}, \tag{61}$$

$$\partial_\varepsilon S_{\rho\sigma}(t) = 0. \tag{62}$$

Differentiating the gauge potential $A_{\varepsilon\,i}^{j}(t) = \mathbf{C}_{\varepsilon\,j}^{\rho} S_{\rho\sigma} \dot{\mathbf{C}}_{\varepsilon\,i}^{\sigma}$ by the product rule, and using $\partial_\varepsilon S_{\rho\sigma}(t) = 0$, yields the first-order variation of A:

$$\begin{aligned}
\delta A_{t\,i}^{j} &\coloneqq \partial_\varepsilon A_{\varepsilon\,i}^{j}(t)\Big|_{\varepsilon=0} \\
&= \partial_\varepsilon \Big[\mathbf{C}_{\varepsilon\,j}^{\rho}(t) S_{\rho\sigma}(t) \dot{\mathbf{C}}_{\varepsilon\,i}^{\sigma}(t)\Big]\Big|_{\varepsilon=0} \\
&= \Big(\partial_\varepsilon \mathbf{C}_{\varepsilon\,j}^{\rho}(t)\Big)\Big|_{\varepsilon=0} S_{\rho\sigma}(t) \dot{\mathbf{C}}_{t\,i}^{\sigma}(t) + \mathbf{C}_{t\,j}^{\rho}(t) \Big(\partial_\varepsilon S_{\rho\sigma}(t)\Big)\Big|_{\varepsilon=0} \dot{\mathbf{C}}_{t\,i}^{\sigma}(t) + \mathbf{C}_{t\,j}^{\rho}(t) S_{\rho\sigma}(t) \Big(\partial_\varepsilon \dot{\mathbf{C}}_{\varepsilon\,i}^{\sigma}(t)\Big)\Big|_{\varepsilon=0} \\
&= \delta \mathbf{C}_{t\,j}^{\rho} S_{\rho\sigma}(t) \dot{\mathbf{C}}_{t\,i}^{\sigma} + \mathbf{C}_{t\,j}^{\rho} \underbrace{\Big(\partial_\varepsilon S_{\rho\sigma}(t)\Big)\Big|_{\varepsilon=0}}_{=0} \dot{\mathbf{C}}_{t\,i}^{\sigma} + \mathbf{C}_{t\,j}^{\rho} S_{\rho\sigma}(t) \delta \dot{\mathbf{C}}_{t\,i}^{\sigma} \\
&= \delta \mathbf{C}_{t\,j}^{\rho} S_{\rho\sigma}(t) \dot{\mathbf{C}}_{t\,i}^{\sigma} + \mathbf{C}_{t\,j}^{\rho} S_{\rho\sigma}(t) \delta \dot{\mathbf{C}}_{t\,i}^{\sigma}.
\end{aligned} \tag{63}$$

Finally, we introduce the S_t-orthogonal projector onto the occupied subspace. In index form this projector will be used to collect the terms that are proportional to $\delta\dot{\mathbf{C}}_t$:

$$\mathbf{P}^{\mu}_{t\,\sigma} = \mathbf{C}^{\mu}_{t\,j}\mathbf{C}^{\rho}_{t\,j}S_{\rho\sigma}(t).$$

With these preparations, we expand the ε-variation of the covariant velocity. We differentiate the definition of $\nabla_t\mathbf{C}_\varepsilon(t)$ with respect to ε, apply the product rule, and then substitute the previously derived expression for $\delta A^j_{t\,i}$. This yields an explicit formula for $\delta\big(\nabla_t\mathbf{C}_t\big)$ in terms of $\delta\mathbf{C}_t$ and $\delta\dot{\mathbf{C}}_t$:

$$\begin{aligned}
\delta\big(\nabla_t\mathbf{C}_t\big)^{\mu}{}_i &= \partial_\varepsilon\Big[\big(\nabla_t\mathbf{C}_\varepsilon(t)\big)^{\mu}{}_i\Big]\Big|_{\varepsilon=0}\\
&= \partial_\varepsilon\Big[\dot{\mathbf{C}}^{\mu}_{\varepsilon\,i}(t) - \mathbf{C}^{\mu}_{\varepsilon\,j}(t)A^j_{\varepsilon i}(t)\Big]\Big|_{\varepsilon=0}\\
&= \partial_\varepsilon\dot{\mathbf{C}}^{\mu}_{\varepsilon\,i}(t)\Big|_{\varepsilon=0} - \partial_\varepsilon\mathbf{C}^{\mu}_{\varepsilon\,j}(t)\Big|_{\varepsilon=0}A^j_{t\,i}(t) - \mathbf{C}^{\mu}_{t\,j}(t)\partial_\varepsilon A^j_{\varepsilon i}(t)\Big|_{\varepsilon=0}\\
&= \delta\dot{\mathbf{C}}^{\mu}_{t\,i} - \delta\mathbf{C}^{\mu}_{t\,j}A^j_{t\,i} - \mathbf{C}^{\mu}_{t\,j}\delta A^j_{t\,i}\\
&= \delta\dot{\mathbf{C}}^{\mu}_{t\,i} - \delta\mathbf{C}^{\mu}_{t\,j}A^j_{t\,i} - \mathbf{C}^{\mu}_{t\,j}\Big(\delta\mathbf{C}^{\rho}_{t\,j}S_{\rho\sigma}\dot{\mathbf{C}}^{\sigma}_{t\,i} + \mathbf{C}^{\rho}_{t\,j}S_{\rho\sigma}\delta\dot{\mathbf{C}}^{\sigma}_{t\,i}\Big)\\
&= \delta\dot{\mathbf{C}}^{\mu}_{t\,i} - \delta\mathbf{C}^{\mu}_{t\,j}A^j_{t\,i} - \mathbf{C}^{\mu}_{t\,j}\delta\mathbf{C}^{\rho}_{t\,j}S_{\rho\sigma}\dot{\mathbf{C}}^{\sigma}_{t\,i} - \Big(\mathbf{C}^{\mu}_{t\,j}\mathbf{C}^{\rho}_{t\,j}S_{\rho\sigma}\Big)\delta\dot{\mathbf{C}}^{\sigma}_{t\,i}\\
&= \delta\dot{\mathbf{C}}^{\mu}_{t\,i} - \delta\mathbf{C}^{\mu}_{t\,j}A^j_{t\,i} - \mathbf{C}^{\mu}_{t\,j}\delta\mathbf{C}^{\rho}_{t\,j}S_{\rho\sigma}\dot{\mathbf{C}}^{\sigma}_{t\,i} - \mathbf{P}^{\mu}_{t\,\sigma}\delta\dot{\mathbf{C}}^{\sigma}_{t\,i}\\
&= \big(\delta^{\mu}{}_{\sigma} - \mathbf{P}^{\mu}_{t\,\sigma}\big)\delta\dot{\mathbf{C}}^{\sigma}_{t\,i} - \delta\mathbf{C}^{\mu}_{t\,j}A^j_{t\,i} - \mathbf{C}^{\mu}_{t\,j}\delta\mathbf{C}^{\rho}_{t\,j}S_{\rho\sigma}\dot{\mathbf{C}}^{\sigma}_{t\,i}.
\end{aligned} \tag{64}$$

We now compute $\nabla_t\delta\mathbf{C}_t$ directly from its defining formula, using the same S_t-connection and the same gauge potential A_t. Substituting the same expression for $\delta A^j_{t\,i}$ and collecting terms, we obtain an expanded form that is manifestly identical to the previous expression:

$$\begin{aligned}
\big(\nabla_t\delta\mathbf{C}_t\big)^{\mu}{}_i &= \delta\dot{\mathbf{C}}^{\mu}_{t\,i} - \delta\mathbf{C}^{\mu}_{t\,j}A^j_{t\,i} - \mathbf{C}^{\mu}_{t\,j}\delta A^j_{t\,i}\\
&= \delta\dot{\mathbf{C}}^{\mu}_{t\,i} - \delta\mathbf{C}^{\mu}_{t\,j}A^j_{t\,i} - \mathbf{C}^{\mu}_{t\,j}\Big(\delta\mathbf{C}^{\rho}_{t\,j}\mathrm{S}_{\rho\sigma}\dot{\mathbf{C}}^{\sigma}_{t\,i} + \mathbf{C}^{\rho}_{t\,j}\mathrm{S}_{\rho\sigma}\delta\dot{\mathbf{C}}^{\sigma}_{t\,i}\Big)\\
&= \delta\dot{\mathbf{C}}^{\mu}_{t\,i} - \delta\mathbf{C}^{\mu}_{t\,j}A^j_{t\,i} - \mathbf{C}^{\mu}_{t\,j}\delta\mathbf{C}^{\rho}_{t\,j}\mathrm{S}_{\rho\sigma}\dot{\mathbf{C}}^{\sigma}_{t\,i} - \Big(\mathbf{C}^{\mu}_{t\,j}\mathbf{C}^{\rho}_{t\,j}\mathrm{S}_{\rho\sigma}\Big)\delta\dot{\mathbf{C}}^{\sigma}_{t\,i}\\
&= \delta\dot{\mathbf{C}}^{\mu}_{t\,i} - \delta\mathbf{C}^{\mu}_{t\,j}A^j_{t\,i} - \mathbf{C}^{\mu}_{t\,j}\delta\mathbf{C}^{\rho}_{t\,j}\mathrm{S}_{\rho\sigma}\dot{\mathbf{C}}^{\sigma}_{t\,i} - \mathbf{P}^{\mu}_{t\,\sigma}\delta\dot{\mathbf{C}}^{\sigma}_{t\,i}\\
&= \big(\delta^{\mu}{}_{\sigma} - \mathbf{P}^{\mu}_{t\,\sigma}\big)\delta\dot{\mathbf{C}}^{\sigma}_{t\,i} - \delta\mathbf{C}^{\mu}_{t\,j}A^j_{t\,i} - \mathbf{C}^{\mu}_{t\,j}\delta\mathbf{C}^{\rho}_{t\,j}\mathrm{S}_{\rho\sigma}\dot{\mathbf{C}}^{\sigma}_{t\,i}.
\end{aligned} \tag{65}$$

Since the right-hand sides of the two expansions coincide term-by-term, we conclude that

$$\delta\big(\nabla_t\mathbf{C}_t\big)^{\mu}{}_i = \big(\nabla_t\delta\mathbf{C}_t\big)^{\mu}{}_i.$$

This proves the commutation identity in local coordinates and completes the proof. □

Lemma 11.3 (Horizontal Gradient as Projected Ambient Gradient)**.** *On the Stiefel manifold* $\mathrm{St_S}(M,N) = \{\mathbf{C} \in \mathbb{C}^{M\times N} : \mathbf{C}^\dagger \mathrm{S}\mathbf{C} = I_N\}$*, define the gauge-invariant projector* $\mathbf{P} := \mathbf{C}\mathbf{C}^\dagger \mathrm{S}$ *and assume the energy has the form* $\mathcal{E}(\mathbf{r},\mathbf{C}) = \widetilde{\mathcal{E}}(\mathbf{r},\mathbf{P})$ *with Fréchet derivative* $\delta\widetilde{\mathcal{E}} = \langle \mathbf{H}^\star, \delta\mathbf{P}\rangle$ *for some Hermitian* $\mathbf{H}^\star = \mathbf{H}^\star(\mathbf{r},\mathbf{P})$*. Then the horizontal Riemannian gradient of* $\mathcal{E}$ *at* $\mathbf{C}$*, with respect to the* S*-metric* $\langle X, Y\rangle_{\mathrm{S}} := \mathrm{ReTr}(X^\dagger \mathrm{S} Y)$ *and the standard principal* $\mathrm{U}(N)$*-gauge, is given by*

$$\mathrm{grad}^{\mathrm{Hor}}_{\mathbf{C}} \mathcal{E}(\mathbf{r},\mathbf{C}) = (\mathbf{I}-\mathbf{P})\mathrm{S}(\mathbf{r})^{-1}\mathbf{H}^\star\mathbf{C}. \tag{66}$$

Proof. To derive the expression for the horizontal gradient, we consider an arbitrary smooth path $\mathbf{C}_\varepsilon$ passing through $\mathbf{C}$ at $\varepsilon = 0$. Let us explicitly define the associated variations for the gauge-invariant projector and the wavefunction coefficients as follows:

$$\mathbf{P}_\varepsilon := \mathbf{P}(\mathbf{C}_\varepsilon) = \mathbf{C}_\varepsilon \mathbf{C}_\varepsilon^\dagger \mathrm{S}, \quad \delta\mathbf{P} := \frac{\partial}{\partial\varepsilon}\bigg|_{\varepsilon=0} \mathbf{P}_\varepsilon, \quad \delta\mathbf{C} := \frac{\partial}{\partial\varepsilon}\bigg|_{\varepsilon=0} \mathbf{C}_\varepsilon.$$

We first compute the variation of the projector $\delta\mathbf{P}$ in terms of $\delta\mathbf{C}$. By applying the chain rule directly to the definition of $\mathbf{P}_\varepsilon$, we obtain:

$$\begin{aligned}
\delta\mathbf{P} &= \frac{\partial}{\partial\varepsilon}\bigg|_{\varepsilon=0} \big(\mathbf{C}_\varepsilon \mathbf{C}_\varepsilon^\dagger \mathrm{S}\big) \\
&= \left(\frac{\partial \mathbf{C}_\varepsilon}{\partial\varepsilon}\mathbf{C}_\varepsilon^\dagger \mathrm{S} + \mathbf{C}_\varepsilon \frac{\partial \mathbf{C}_\varepsilon^\dagger}{\partial\varepsilon}\mathrm{S} + \mathbf{C}_\varepsilon \mathbf{C}_\varepsilon^\dagger \frac{\partial \mathrm{S}}{\partial\varepsilon}\right)\bigg|_{\varepsilon=0} \quad \text{(Leibniz rule)} \\
&= \left(\frac{\partial \mathbf{C}_\varepsilon}{\partial\varepsilon}\mathbf{C}_\varepsilon^\dagger \mathrm{S} + \mathbf{C}_\varepsilon \frac{\partial \mathbf{C}_\varepsilon^\dagger}{\partial\varepsilon}\mathrm{S}\right)\bigg|_{\varepsilon=0} \quad (\delta\mathrm{S} = 0) \\
&= \left(\frac{\partial \mathbf{C}_\varepsilon}{\partial\varepsilon}\mathbf{C}_\varepsilon^\dagger \mathrm{S} + \mathbf{C}_\varepsilon \Big(\frac{\partial \mathbf{C}_\varepsilon}{\partial\varepsilon}\Big)^\dagger \mathrm{S}\right)\bigg|_{\varepsilon=0} \quad \Big(\frac{\partial}{\partial\varepsilon}(\mathbf{C}_\varepsilon^\dagger) = \frac{\partial \mathbf{C}_\varepsilon}{\partial\varepsilon}\Big)^\dagger \\
&= \delta\mathbf{C}\mathbf{C}^\dagger \mathrm{S} + \mathbf{C}(\delta\mathbf{C})^\dagger \mathrm{S}.
\end{aligned}$$

This result relates the variation of the physical state to the variation of the representation. Next, we proceed to compute the variation of the energy functional. Substituting the expression for $\delta\mathbf{P}$ into the definition of the Fréchet derivative $\delta\widetilde{\mathcal{E}} = \langle \mathbf{H}^\star, \delta\mathbf{P}\rangle$, and exploiting the Hermiticity of $\mathbf{H}^\star$ alongside the cyclic property of the trace, we deduce:

$$\begin{aligned}
\delta\mathcal{E}(\mathbf{r},\mathbf{C}) &= \langle \mathbf{H}^\star, \delta\mathbf{P}\rangle \\
&= \mathrm{ReTr}\big(\mathbf{H}^\star(\delta\mathbf{C}\mathbf{C}^\dagger \mathrm{S} + \mathbf{C}(\delta\mathbf{C})^\dagger \mathrm{S})\big) \\
&= \mathrm{ReTr}\big(\mathbf{H}^\star \delta\mathbf{C}\mathbf{C}^\dagger \mathrm{S}\big) + \mathrm{ReTr}\big(\mathbf{H}^\star \mathbf{C}(\delta\mathbf{C})^\dagger \mathrm{S}\big) \\
&= \mathrm{ReTr}\big(\mathbf{C}^\dagger \mathrm{S}\mathbf{H}^\star \delta\mathbf{C}\big) + \mathrm{ReTr}\big((\delta\mathbf{C})^\dagger \mathrm{S}\mathbf{H}^\star\mathbf{C}\big) \quad \text{(cyclicity of trace)} \\
&= \mathrm{ReTr}\big((\mathbf{C}^\dagger \mathrm{S}\mathbf{H}^\star \delta\mathbf{C})^\dagger\big) + \mathrm{ReTr}\big((\delta\mathbf{C})^\dagger \mathrm{S}\mathbf{H}^\star\mathbf{C}\big) \quad (\text{since } \mathrm{ReTr}(X) = \mathrm{ReTr}(X^\dagger)) \\
&= \mathrm{ReTr}\big((\delta\mathbf{C})^\dagger (\mathbf{H}^\star)^\dagger \mathrm{S}^\dagger \mathbf{C}\big) + \mathrm{ReTr}\big((\delta\mathbf{C})^\dagger \mathrm{S}\mathbf{H}^\star\mathbf{C}\big) \\
&= \mathrm{ReTr}\big((\delta\mathbf{C})^\dagger \mathbf{H}^\star \mathrm{S}\mathbf{C}\big) + \mathrm{ReTr}\big((\delta\mathbf{C})^\dagger \mathrm{S}\mathbf{H}^\star\mathbf{C}\big) \quad (\mathbf{H}^\star = \mathbf{H}^{\star\dagger},\ \mathrm{S} = \mathrm{S}^\dagger) \\
&= \mathrm{ReTr}\Big((\delta\mathbf{C})^\dagger \mathrm{S}\Big(\mathrm{S}^{-1}\mathbf{H}^\star \mathrm{S} + \mathbf{H}^\star\Big)\mathbf{C}\Big) \\
&= 2\mathrm{ReTr}\Big((\delta\mathbf{C})^\dagger \mathrm{S}\mathrm{S}^{-1}\mathbf{H}^\star\mathbf{C}\Big) \quad \text{(identifying } \mathbf{H}^\star \text{ with the symmetrized representative)} \\
&= 2\langle \mathrm{S}^{-1}\mathbf{H}^\star\mathbf{C},\ \delta\mathbf{C}\rangle_{\mathrm{S}}.
\end{aligned}$$

From this relation, we can identify the unconstrained gradient in the ambient Euclidean space. By definition of the gradient with respect to the S-metric, $\langle X, Y\rangle_{\mathrm{S}} := \mathrm{ReTr}(X^\dagger \mathrm{S} Y)$, it follows that the ambient gradient is characterized by

$$\delta\mathcal{E}(\mathbf{r},\mathbf{C}) = \langle \mathrm{grad}^{\mathrm{Amb}}_{\mathbf{C}}\mathcal{E},\ \delta\mathbf{C}\rangle_{\mathrm{S}} = \mathrm{ReTr}\big((\mathrm{grad}^{\mathrm{Amb}}_{\mathbf{C}}\mathcal{E})^\dagger \mathrm{S}\delta\mathbf{C}\big) \quad (\forall \delta\mathbf{C} \in \mathbb{C}^{M\times N}),$$

and comparing with $\delta\mathcal{E} = 2\langle \mathrm{S}^{-1}\mathbf{H}^\star\mathbf{C}, \delta\mathbf{C}\rangle_{\mathrm{S}}$ gives

$$\mathrm{grad}^{\mathrm{Amb}}_{\mathbf{C}}\mathcal{E} = 2\mathrm{S}^{-1}\mathbf{H}^\star\mathbf{C}.$$

To isolate the physical gradient, we employ the projection operator onto the horizontal space. We assert that the operator $\Pi_{\mathrm{Hor}}(\mathbf{Z}) = (\mathbf{I} - \mathbf{P})\mathbf{Z}$ acts as the S-orthogonal projector onto $H_{\mathbf{C}}$ for any $\mathbf{Z} \in \mathbb{C}^{M\times N}$. This is verified by checking the orthogonality condition:

$$\begin{aligned}\mathbf{C}^\dagger \mathrm{S}(\mathbf{I} - \mathbf{P}) &= \mathbf{C}^\dagger \mathrm{S} - \mathbf{C}^\dagger \mathrm{S}\mathbf{C}\mathbf{C}^\dagger \mathrm{S} \\ &= \mathbf{C}^\dagger \mathrm{S} - (\mathbf{C}^\dagger \mathrm{S}\mathbf{C})\mathbf{C}^\dagger \mathrm{S} \\ &= \mathbf{C}^\dagger \mathrm{S} - \mathrm{I}_N \mathbf{C}^\dagger \mathrm{S} \\ &= 0,\end{aligned}$$

where we used the fact that $\mathbf{C}^\dagger \mathrm{S}\mathbf{C} = I_N$ and that $\mathbf{P} = \mathbf{C}\mathbf{C}^\dagger \mathrm{S}$ is self-adjoint with respect to S. This means that

$$\langle \mathbf{P}X, Y\rangle_{\mathrm{S}} = \mathrm{ReTr}(X^\dagger \mathbf{P}^\dagger \mathrm{S}Y) = \mathrm{ReTr}(X^\dagger \mathrm{S}\mathbf{P}Y) = \langle X, \mathbf{P}Y\rangle_{\mathrm{S}}.$$

Finally, the horizontal Riemannian gradient is uniquely characterized by the property that it represents the energy variation restricted to horizontal directions, i.e., $\delta\mathcal{E} = \langle \mathrm{grad}^{\mathrm{Hor}}_{\mathbf{C}}\mathcal{E}, \delta\mathbf{C}\rangle_{\mathrm{S}}$ for all $\delta\mathbf{C} \in H_{\mathbf{C}}$. Applying the projection to the ambient gradient derived earlier yields:

$$\begin{aligned}\mathrm{grad}^{\mathrm{Hor}}_{\mathbf{C}}\mathcal{E} &= \Pi_{\mathrm{Hor}}\big(\mathrm{grad}^{\mathrm{Amb}}_{\mathbf{C}}\mathcal{E}\big) = (\mathbf{I} - \mathbf{P})\big(2\mathrm{S}^{-1}\mathbf{H}^\star\mathbf{C}\big) \\ &= 2(\mathbf{I} - \mathbf{P})\mathrm{S}^{-1}\mathbf{H}^\star\mathbf{C}.\end{aligned}$$

Absorbing the conventional factor of 2 into the definition of $\mathbf{H}^\star$, this gives (66), as claimed. □

Lemma 11.4 (Equivalence between Euler–Lagrange–d'Alembert and the covariant momentum equation)**.** *Let us consider $t \mapsto \mathbf{r}_t$ and $t \mapsto \mathrm{S}_t$. Define the covariant velocity and momentum by*

$$\mathbf{V}_t \coloneqq \nabla_t \mathbf{C}_t, \quad \mathbf{\Pi}_t \coloneqq \mathbf{V}_t^\flat = \mathrm{S}_t \mathbf{V}_t, \quad \mathbf{\Pi}_t^\sharp = \mathrm{S}_t^{-1}\mathbf{\Pi}_t,$$

and consider the Lagrangian and Rayleigh dissipation

$$\mathcal{L}(\mathbf{C}_t, \mathbf{V}_t, \mathbf{P}_t; \mathbf{r}_t) = \frac{\mu_t}{2}\|\mathbf{V}_t\|^2_{\mathrm{S}_t} - \mathcal{E}(\mathbf{r}_t, \mathbf{P}_t), \quad \mathcal{D}(\mathbf{C}_t, \nabla_t\mathbf{C}_t; \mathbf{r}_t) = \frac{\gamma}{2}\|\mathbf{V}_t\|^2_{\mathrm{S}_t}.$$

Then the following two equations are equivalent:

$$\partial_{\mathbf{C}}\mathcal{L} - \nabla_t\big(\partial_{\mathbf{V}}\mathcal{L}\big) - \partial_{\nabla_t\mathbf{C}}\mathcal{D} = 0$$

and

$$\mu_t \nabla_t \mathbf{\Pi}_t + \gamma\mathbf{\Pi}_t + \big(\mathrm{grad}^{\mathrm{Hor}}_{\mathbf{C}}\mathcal{E}\big)^\flat\big(\mathbf{r}_t, \mathbf{P}_t\big) = -\dot{\mu}_t\mathbf{\Pi}_t.$$

To compute the fiber derivatives with respect to the velocity variable, we fix $\mathbf{C}_t$ and $\mathbf{P}_t$ and differentiate the kinetic and dissipation terms in the direction of an arbitrary variation $\mathbf{W}$. We work with the real Frobenius pairing on ambient matrices and its S_t-weighted counterpart:

$$\langle \mathbf{A}, \mathbf{B}\rangle_{\mathrm{F}} \coloneqq \Re\,\mathrm{Tr}\big(\mathbf{A}^\dagger\mathbf{B}\big), \quad \langle \mathbf{A}, \mathbf{B}\rangle_{\mathrm{S}_t} \coloneqq \Re\,\mathrm{Tr}\big(\mathbf{A}^\dagger \mathrm{S}_t\mathbf{B}\big).$$

With these pairings, the squared S_t-norm admits both a matrix-trace expression and a componentwise expression, which will be convenient for an explicit Gâteaux differentiation:

$$\|\mathbf{V}_t\|^2_{\mathrm{S}_t} = \langle \mathbf{V}_t, \mathbf{V}_t\rangle_{\mathrm{S}_t} = \Re\,\mathrm{Tr}\big(\mathbf{V}_t^\dagger \mathrm{S}_t\mathbf{V}_t\big) = \Re\Big(\overline{(\mathbf{V}_{t\,i}^{\mu})}(\mathrm{S}_t)_{\mu\nu}\mathbf{V}_{t\,i}^{\nu}\Big).$$

We also introduce the musical isomorphism induced by S_t, which converts the velocity vector to its associated covector: $(\mathbf{V}_t^\flat)^\mu{}_i = (\mathrm{S}_t)_{\mu\nu}\mathbf{V}_{t\,i}^{\nu}$. Let $\mathbf{W}$ be an arbitrary test variation of $\mathbf{V}_t$ and define the directional derivative $D_{\mathbf{V}}\mathcal{L}[\mathbf{W}]$ through the perturbation $\mathbf{V}_t + \varepsilon\mathbf{W}$. Expanding the quadratic form and keeping only the first-order terms in ε yields:

$$\mathcal{L}(\mathbf{V}_t) \coloneqq \frac{\mu_t}{2}\|\mathbf{V}_t\|^2_{\mathrm{S}_t} = \frac{\mu_t}{2}\Re\Big(\overline{(\mathbf{V}_{t\,i}^{\mu})}(\mathrm{S}_t)_{\mu\nu}\mathbf{V}_{t\,i}^{\nu}\Big), \tag{67}$$

Taking differentiation gives

$$\begin{aligned} D_{\mathbf{V}}\mathcal{L}[\mathbf{W}] &\coloneqq \frac{d}{d\varepsilon}\mathcal{L}(\mathbf{V}_t+\varepsilon\mathbf{W})\Big|_{\varepsilon=0} \\ &= \frac{\mu_t}{2}\frac{d}{d\varepsilon}\Re\Big(\overline{(\mathbf{V}^{\mu}_{t\,i}+\varepsilon\mathbf{W}^{\mu}{}_i)}(\mathrm{S}_t)_{\mu\nu}(\mathbf{V}^{\nu}_{t\,i}+\varepsilon\mathbf{W}^{\nu}{}_i)\Big)\Big|_{\varepsilon=0} \\ &= \frac{\mu_t}{2}\Re\Big(\overline{\mathbf{W}^{\mu}{}_i}(\mathrm{S}_t)_{\mu\nu}\mathbf{V}^{\nu}_{t\,i}+\overline{\mathbf{V}^{\mu}_{t\,i}}(\mathrm{S}_t)_{\mu\nu}\mathbf{W}^{\nu}{}_i\Big) \\ &= \mu_t\Re\Big(\overline{\mathbf{W}^{\mu}{}_i}(\mathrm{S}_t)_{\mu\nu}\mathbf{V}^{\nu}_{t\,i}\Big) \\ &= \mu_t\Re\,\mathrm{Tr}\big(\mathbf{W}^{\dagger}\mathrm{S}_t\mathbf{V}_t\big) = \mu_t\Re\,\mathrm{Tr}\big((\mathrm{S}_t\mathbf{V}_t)^{\dagger}\mathbf{W}\big) \\ &= \Big\langle \mu_t\mathrm{S}_t\mathbf{V}_t,\ \mathbf{W}\Big\rangle_{\mathrm{F}}. \end{aligned} \tag{68}$$

The last line exhibits $D_{\mathbf{V}}\mathcal{L}[\mathbf{W}]$ as a Frobenius pairing against $\mathbf{W}$. By the definition of the fiber derivative $\partial_{\mathbf{V}}\mathcal{L}$ as the unique covector satisfying $D_{\mathbf{V}}\mathcal{L}[\mathbf{W}] = \langle \partial_{\mathbf{V}}\mathcal{L}, \mathbf{W}\rangle_{\mathrm{F}}$ for all $\mathbf{W}$, we identify:

$$\partial_{\mathbf{V}}\mathcal{L} = \mu_t\mathrm{S}_t\mathbf{V}_t = \mu_t\mathbf{V}_t^{\flat}. \tag{69}$$

We now repeat the same calculation for the Rayleigh dissipation potential. Since $\mathcal{D}$ is the same quadratic form with coefficient γ, its directional derivative is obtained by the same expansion, and again can be written as a Frobenius pairing against $\mathbf{W}$:

$$\mathcal{D}(\mathbf{V}_t) \coloneqq \frac{\gamma}{2}\|\mathbf{V}_t\|^2_{\mathrm{S}_t} = \frac{\gamma}{2}\Re\Big(\overline{(\mathbf{V}^{\mu}_{t\,i})}(\mathrm{S}_t)_{\mu\nu}\mathbf{V}^{\nu}_{t\,i}\Big), \tag{70}$$

Thus, we have

$$\begin{aligned} D_{\mathbf{V}}\mathcal{D}[\mathbf{W}] &\coloneqq \frac{d}{d\varepsilon}\mathcal{D}(\mathbf{V}_t+\varepsilon\mathbf{W})\Big|_{\varepsilon=0} \\ &= \frac{\gamma}{2}\Re\Big(\overline{\mathbf{W}^{\mu}{}_i}(\mathrm{S}_t)_{\mu\nu}\mathbf{V}^{\nu}_{t\,i}+\overline{\mathbf{V}^{\mu}_{t\,i}}(\mathrm{S}_t)_{\mu\nu}\mathbf{W}^{\nu}{}_i\Big) \\ &= \gamma\Re\Big(\overline{\mathbf{W}^{\mu}{}_i}(\mathrm{S}_t)_{\mu\nu}\mathbf{V}^{\nu}_{t\,i}\Big) = \gamma\Re\,\mathrm{Tr}\big((\mathrm{S}_t\mathbf{V}_t)^{\dagger}\mathbf{W}\big) \\ &= \Big\langle \gamma\mathrm{S}_t\mathbf{V}_t,\ \mathbf{W}\Big\rangle_{\mathrm{F}}. \end{aligned} \tag{71}$$

As before, the characterization by duality implies that the dissipation force covector equals the coefficient times the S_t-flat of the velocity:

$$\partial_{\nabla_t\mathbf{C}}\mathcal{D} = \partial_{\mathbf{V}}\mathcal{D} = \gamma\mathrm{S}_t\mathbf{V}_t = \gamma\mathbf{V}_t^{\flat}. \tag{72}$$

We now define the differential of $\mathcal{E}$ at $\mathbf{C}_t$ by the Frobenius pairing

$$\delta\mathcal{E}_t \coloneqq \partial_{\varepsilon}\mathcal{E}\big(\mathbf{r}_t, \mathbf{P}(\mathbf{C}_{\varepsilon}(t))\big)\Big|_{\varepsilon=0} = \Big\langle \partial_{\mathbf{C}}\mathcal{E}\big(\mathbf{r}_t,\mathbf{P}_t\big),\ \delta\mathbf{C}_t\Big\rangle_{\mathrm{F}}, \quad \langle\mathbf{A},\mathbf{B}\rangle_{\mathrm{F}} \coloneqq \Re\,\mathrm{Tr}\big(\mathbf{A}^{\dagger}\mathbf{B}\big).$$

Since $\mathcal{E}(\mathbf{r}_t, \mathbf{P}(\mathbf{C}(x,\theta,t)))$ depends only on x, the chain rule in local coordinates gives

$$\begin{aligned} \delta\mathcal{E}_t &= \partial_{\varepsilon}\mathcal{E}\Big(\mathbf{r}_t, \mathbf{P}\big(\widetilde{\mathbf{C}}(x_{\varepsilon}(t),t)\mathbf{U}(\theta_{\varepsilon}(t))\big)\Big)\Big|_{\varepsilon=0} \\ &= \partial_{\varepsilon}\mathcal{E}\Big(\mathbf{r}_t, \widetilde{\mathbf{C}}\big(x_{\varepsilon}(t),t\big)\widetilde{\mathbf{C}}\big(x_{\varepsilon}(t),t\big)^{\dagger}\mathrm{S}_t\Big)\Big|_{\varepsilon=0} \\ &= \partial_{\alpha}\mathcal{E}\Big(\mathbf{r}_t, \widetilde{\mathbf{C}}\big(x(t),t\big)\widetilde{\mathbf{C}}\big(x(t),t\big)^{\dagger}\mathrm{S}_t\Big)\delta x^{\alpha}(t). \end{aligned} \tag{73}$$

On the other hand, substituting the decomposition of $\delta\mathbf{C}_t$ into the definition of $\partial_{\mathbf{C}}\mathcal{E}$ yields

$$\begin{aligned} \delta\mathcal{E}_t &= \Big\langle \partial_{\mathbf{C}}\mathcal{E}\big(\mathbf{r}_t,\mathbf{P}_t\big),\ \partial_{\alpha}\widetilde{\mathbf{C}}\big(x(t),t\big)\delta x^{\alpha}(t)\mathbf{U}(t) + \mathbf{C}_t\boldsymbol{\Omega}_t\Big\rangle_{\mathrm{F}} \\ &= \delta x^{\alpha}(t)\Big\langle \partial_{\mathbf{C}}\mathcal{E}\big(\mathbf{r}_t,\mathbf{P}_t\big),\ \partial_{\alpha}\widetilde{\mathbf{C}}\big(x(t),t\big)\mathbf{U}(t)\Big\rangle_{\mathrm{F}} + \Big\langle \partial_{\mathbf{C}}\mathcal{E}\big(\mathbf{r}_t,\mathbf{P}_t\big),\ \mathbf{C}_t\boldsymbol{\Omega}_t\Big\rangle_{\mathrm{F}}. \end{aligned} \tag{74}$$

Comparing (73) and (74), the variation $\delta\mathcal{E}_t$ has no dependence on $\mathbf{\Omega}_t$. Since $\mathbf{\Omega}_t \in \mathfrak{u}(N)$ is arbitrary, it follows that

$$\Big\langle \partial_{\mathbf{C}}\mathcal{E}\big(\mathbf{r}_t, \mathbf{P}_t\big),\ \mathbf{C}_t\mathbf{\Omega} \Big\rangle_{\mathrm{F}} = 0 \quad \forall \mathbf{\Omega} \in \mathfrak{u}(N). \tag{75}$$

Equation (75) states that the covector $\partial_{\mathbf{C}}\mathcal{E}(\mathbf{r}_t, \mathbf{P}_t)$ annihilates the vertical directions. Therefore $\partial_{\mathbf{C}}\mathcal{E}(\mathbf{r}_t, \mathbf{P}_t)$ belongs to the horizontal cotangent space. Next, we use the S_t-metric to identify horizontal covectors with horizontal vectors. Define the S_t-inner product and the musical isomorphism by

$$\langle \mathbf{A}, \mathbf{B} \rangle_{\mathrm{S}_t} \coloneqq \Re\,\mathrm{Tr}\big(\mathbf{A}^\dagger \mathrm{S}_t \mathbf{B}\big), \quad \mathbf{Z}^\flat \coloneqq \mathrm{S}_t\mathbf{Z}, \quad \langle \mathbf{Z}, \Delta \rangle_{\mathrm{S}_t} = \langle \mathbf{Z}^\flat, \Delta \rangle_{\mathrm{F}}.$$

Since $\partial_{\mathbf{C}}\mathcal{E}(\mathbf{r}_t, \mathbf{P}_t)$ is horizontal, there exists a unique horizontal vector $\mathrm{grad}^{\mathrm{Hor}}_{\mathbf{C}}\mathcal{E}(\mathbf{r}_t, \mathbf{P}_t)$ such that, for all admissible $\delta\mathbf{C}_t$,

$$\delta\mathcal{E}_t = \Big\langle \mathrm{grad}^{\mathrm{Hor}}_{\mathbf{C}}\mathcal{E}\big(\mathbf{r}_t, \mathbf{P}_t\big),\ \delta\mathbf{C}_t \Big\rangle_{\mathrm{S}_t} = \Big\langle \big(\mathrm{grad}^{\mathrm{Hor}}_{\mathbf{C}}\mathcal{E}\big)^\flat\big(\mathbf{r}_t, \mathbf{P}_t\big),\ \delta\mathbf{C}_t \Big\rangle_{\mathrm{F}}. \tag{76}$$

By uniqueness of the Riesz representation under $\langle \cdot, \cdot \rangle_{\mathrm{F}}$, we conclude

$$\partial_{\mathbf{C}}\mathcal{E}\big(\mathbf{r}_t, \mathbf{P}_t\big) = \big(\mathrm{grad}^{\mathrm{Hor}}_{\mathbf{C}}\mathcal{E}\big)^\flat\big(\mathbf{r}_t, \mathbf{P}_t\big).$$

Finally, because $\mathcal{L}(\mathbf{C}_t, \mathbf{V}_t, \mathbf{P}_t; \mathbf{r}_t) = \frac{\mu_t}{2}\|\mathbf{V}_t\|^2_{\mathrm{S}_t} - \mathcal{E}(\mathbf{r}_t, \mathbf{P}_t)$ and the partial derivative $\partial_{\mathbf{C}}\mathcal{L}$ is taken at fixed $\mathbf{V}_t$, the kinetic term contributes no $\mathbf{C}$-dependence in this partial derivative. Hence

$$\partial_{\mathbf{C}}\mathcal{L} = -\partial_{\mathbf{C}}\mathcal{E}\big(\mathbf{r}_t, \mathbf{P}_t\big) = -\big(\mathrm{grad}^{\mathrm{Hor}}_{\mathbf{C}}\mathcal{E}\big)^\flat\big(\mathbf{r}_t, \mathbf{P}_t\big),$$

Substituting the previously computed identities for the three terms $\partial_{\mathbf{C}}\mathcal{L}$, $\partial_{\mathbf{V}}\mathcal{L}$, and $\partial_{\nabla_t\mathbf{C}}\mathcal{D}$ into the Euler–Lagrange–d'Alembert equation, we obtain a closed covector equation on $T^*_{\mathbf{C}_t}\mathrm{St}_{\mathrm{S}_t}(M, N)$. In particular, using

$$\partial_{\mathbf{C}}\mathcal{L} = -\big(\mathrm{grad}^{\mathrm{Hor}}_{\mathbf{C}}\mathcal{E}\big)^\flat\big(\mathbf{r}_t, \mathbf{P}_t\big), \quad \partial_{\mathbf{V}}\mathcal{L} = \mu_t\mathbf{\Pi}_t, \quad \partial_{\nabla_t\mathbf{C}}\mathcal{D} = \gamma\mathbf{\Pi}_t,$$

we arrive at the expanded form

$$\begin{aligned}
\mathbf{0} &= \partial_{\mathbf{C}}\mathcal{L} - \nabla_t\big(\partial_{\mathbf{V}}\mathcal{L}\big) - \partial_{\nabla_t\mathbf{C}}\mathcal{D} \\
&= -\big(\mathrm{grad}^{\mathrm{Hor}}_{\mathbf{C}}\mathcal{E}\big)^\flat\big(\mathbf{r}_t, \mathbf{P}_t\big) - \nabla_t\big(\mu_t\mathbf{\Pi}_t\big) - \gamma\mathbf{\Pi}_t \\
&= -\big(\mathrm{grad}^{\mathrm{Hor}}_{\mathbf{C}}\mathcal{E}\big)^\flat\big(\mathbf{r}_t, \mathbf{P}_t\big) - \mu_t\nabla_t\mathbf{\Pi}_t - \dot{\mu}_t\mathbf{\Pi}_t - \gamma\mathbf{\Pi}_t.
\end{aligned} \tag{77}$$

This relation is naturally rewritten as a first-order system by introducing the momentum variable $\mathbf{\Pi}_t = \mathbf{V}_t^\flat = \mathrm{S}_t\mathbf{V}_t$ and applying the musical isomorphisms. The kinematic relation is obtained directly from the definition of $\mathbf{\Pi}_t$:

$$\mathbf{V}_t = \nabla_t\mathbf{C}_t = \mathbf{\Pi}_t^\sharp = \mathrm{S}_t^{-1}\mathbf{\Pi}_t, \tag{78}$$

while the remaining terms are grouped to isolate the covariant momentum evolution:

$$\mu_t\nabla_t\mathbf{\Pi}_t + \gamma\mathbf{\Pi}_t + \big(\mathrm{grad}^{\mathrm{Hor}}_{\mathbf{C}}\mathcal{E}\big)^\flat\big(\mathbf{r}_t, \mathbf{P}_t\big) = -\dot{\mu}_t\mathbf{\Pi}_t. \tag{79}$$

Together, these two equations provide the desired dual first-order dynamics in $(\mathbf{C}_t, \mathbf{\Pi}_t)$, with the extra term $-\dot{\mu}_t\mathbf{\Pi}_t$ appearing from the product rule applied to $\nabla_t(\mu_t\mathbf{\Pi}_t)$.

Proposition 11.5 (Gauge-Symmetric Dual Lagrangian Dynamic)**.** *For the nuclear trajectory $t \mapsto \mathbf{r}_t$, damping coefficients $\mu_t = e^{-at}$ and $\gamma > 0$, the following system of the physical states ensures energy minimization $\frac{d}{dt}\mathcal{E}(\mathbf{r}_t, \mathbf{C}_t) \approx -\|\mathrm{grad}_{\mathbf{C}}^{\mathrm{Hor}}\mathcal{E}(\mathbf{r}_t, \mathbf{C}_t)\|_{\mathrm{S}}^2 \leq 0$:*

$$(\mathbf{C}_t, \mathbf{\Pi}_t, \mathbf{P}_t)^\top \iff \begin{cases} \nabla_t \mathbf{C}_t = \mathbf{\Pi}_t^\sharp = \mathrm{S}_t^{-1}\mathbf{\Pi}, \\ \mu_t \nabla_t \mathbf{\Pi}_t + \gamma \mathbf{\Pi}_t + (\mathrm{grad}_{\mathbf{C}}^{\mathrm{Hor}}\mathcal{E})^\flat(\mathbf{r}_t, \mathbf{P}_t) = 0, \\ \nabla_t \mathbf{P}_t = \mathrm{S}_t^{-1}\mathbf{\Pi}_t\mathbf{C}_t^\dagger\mathrm{S}_t + \mathbf{C}_t\mathbf{\Pi}_t^\dagger, \end{cases} \tag{80}$$

where $\mathbf{H}^\star$ is self-consistent Hamiltonian matrix and the musical isomorphism of horizontal gradient is derived from the result in Lemma 11.3

Proof. We prove a fully exact dissipation identity for the dual energy

$$\mathscr{H}_t \coloneqq \mathcal{E}(\mathbf{r}_t, \mathbf{P}_t) + \frac{\mu_t}{2}\big\|\nabla_t \mathbf{C}_t\big\|_{\mathrm{S}_t}^2 = \mathcal{E}(\mathbf{r}_t, \mathbf{P}_t) + \frac{\mu_t}{2}\big\|\mathbf{\Pi}_t\big\|_{\mathrm{S}_t^{-1}}^2, \quad \mathbf{\Pi}_t \coloneqq \mathrm{S}_t\nabla_t\mathbf{C}_t.$$

In particular, we will show

$$\frac{d}{dt}\mathscr{H}_t = -\gamma\big\|\nabla_t \mathbf{C}_t\big\|_{\mathrm{S}_t}^2 + \frac{\dot{\mu}_t}{2}\big\|\nabla_t \mathbf{C}_t\big\|_{\mathrm{S}_t}^2 \leq 0,$$

First, we compute the total time derivative of $\mathcal{E}(\mathbf{r}_t, \mathbf{C}_t)$ directly from the definition of energy functional:

$$\mathcal{E}(\mathbf{r}, \mathbf{C}) = \mathrm{Tr}\big(\mathbf{H}(\mathbf{r})\mathbf{P}\big) + \frac{1}{2}\iint_{\mathbb{R}^3\times\mathbb{R}^3} \frac{\rho_{\mathbf{P}}(\mathbf{x})\rho_{\mathbf{P}}(\mathbf{y})}{\|\mathbf{x}-\mathbf{y}\|}d\mathbf{x}d\mathbf{y} + \mathbf{E}_{\mathrm{xc}}\big[\rho_{\mathbf{P}}\big]. \tag{81}$$

where we set $\mathbf{P}_t \coloneqq \mathbf{C}_t\mathbf{C}_t^\dagger\mathrm{S}(\mathbf{r}_t)$ and $\rho_t \coloneqq \rho_{\mathbf{P}_t}$. By taking differentiation on the one-electron trace term, product rule gives

$$\begin{aligned} \frac{d}{dt}\mathrm{Tr}\big(\mathbf{H}(\mathbf{r}_t)\mathbf{P}_t\big) &= \mathrm{Tr}\Big(\frac{d}{dt}\mathbf{H}(\mathbf{r}_t)\mathbf{P}_t\Big) + \mathrm{Tr}\Big(\mathbf{H}(\mathbf{r}_t)\frac{d}{dt}\mathbf{P}_t\Big) \\ &= \mathrm{Tr}\Big(\partial_{\mathbf{r}}\mathbf{H}(\mathbf{r}_t)\dot{\mathbf{r}}_t\mathbf{P}_t\Big) + \mathrm{Tr}\big(\mathbf{H}(\mathbf{r}_t)\dot{\mathbf{P}}_t\big), \end{aligned} \tag{82}$$

where $\partial_{\mathbf{r}}\mathbf{H}(\mathbf{r}_t)\dot{\mathbf{r}}_t$ denotes the directional derivative of $\mathbf{H}$. Differentiate the Hartree term and simplify it by symmetry. Let us define the Hartree potential

$$v_{\mathrm{H},t}(\mathbf{x}) \coloneqq \int_{\mathbb{R}^3} \frac{\rho_t(\mathbf{y})}{\|\mathbf{x}-\mathbf{y}\|}d\mathbf{y}.$$

Then, the temporal differentiation with respect to the integral gives

$$\begin{aligned} \frac{d}{dt}\frac{1}{2}\iint_{\mathbb{R}^3\times\mathbb{R}^3} \frac{\rho_t(\mathbf{x})\rho_t(\mathbf{y})}{\|\mathbf{x}-\mathbf{y}\|}d\mathbf{x}d\mathbf{y} &= \frac{1}{2}\iint \frac{\dot{\rho}_t(\mathbf{x})\rho_t(\mathbf{y}) + \rho_t(\mathbf{x})\dot{\rho}_t(\mathbf{y})}{\|\mathbf{x}-\mathbf{y}\|}d\mathbf{x}d\mathbf{y} \\ &= \iint \frac{\dot{\rho}_t(\mathbf{x})\rho_t(\mathbf{y})}{\|\mathbf{x}-\mathbf{y}\|}d\mathbf{x}d\mathbf{y} \\ &= \int_{\mathbb{R}^3} v_{\mathrm{H},t}(\mathbf{x})\dot{\rho}_t(\mathbf{x})d\mathbf{x}. \end{aligned} \tag{83}$$

Next, our goal is to differentiate the exchange-correlation term by the functional chain rule. Assume $\mathbf{E}_{\mathrm{xc}}[\rho]$ is Gâteaux differentiable, then for the differential $v_{\mathrm{xc},t}$, we have

$$v_{\mathrm{xc},t}(\mathbf{x}) \coloneqq \frac{\delta\mathbf{E}_{\mathrm{xc}}}{\delta\rho}[\rho_t](\mathbf{x}), \quad \frac{d}{dt}\mathbf{E}_{\mathrm{xc}}[\rho_t] = \int_{\mathbb{R}^3} v_{\mathrm{xc},t}(\mathbf{x})\dot{\rho}_t(\mathbf{x})d\mathbf{x}. \tag{84}$$

Next, we combine (82), (83), and (84) and define

$$v_{\mathrm{eff},t}(\mathbf{x}) \coloneqq v_{\mathrm{H},t}(\mathbf{x}) + v_{\mathrm{xc},t}(\mathbf{x}).$$

Consequently, temporal derivative of energy functional can be rewritten as

$$\frac{d}{dt}\mathcal{E}(\mathbf{r}_t, \mathbf{C}_t) = \mathrm{Tr}\Big(\partial_{\mathbf{r}}\mathbf{H}(\mathbf{r}_t)\dot{\mathbf{r}}_t\mathbf{P}_t\Big) + \mathrm{Tr}\big(\mathbf{H}(\mathbf{r}_t)\dot{\mathbf{P}}_t\big) + \int_{\mathbb{R}^3} v_{\mathrm{eff},t}(\mathbf{x})\dot{\rho}_t(\mathbf{x})d\mathbf{x}. \tag{85}$$

Next, we write $\dot{\rho}_t$ as the sum of a $\dot{\mathbf{P}}_t$ contribution and an explicit $\dot{\mathbf{r}}_t$ contribution. Using the AO basis $\phi_\mu(\mathbf{x};\mathbf{r})$, we write $\rho_{\mathbf{P}}(\mathbf{x}) = \mathbf{P}_{\mu\nu}\phi_\mu(\mathbf{x};\mathbf{r})\phi_\nu(\mathbf{x};\mathbf{r})$. By definition,

$$\begin{aligned}\dot{\rho}_t(\mathbf{x}) &= \dot{\mathbf{P}}_{t\mu\nu}\phi_\mu(\mathbf{x};\mathbf{r}_t)\phi_\nu(\mathbf{x};\mathbf{r}_t) + \mathbf{P}_{t\mu\nu}\frac{d}{dt}\Big(\phi_\mu(\mathbf{x};\mathbf{r}_t)\phi_\nu(\mathbf{x};\mathbf{r}_t)\Big) \\ &= \dot{\mathbf{P}}_{t\mu\nu}\phi_\mu(\mathbf{x};\mathbf{r}_t)\phi_\nu(\mathbf{x};\mathbf{r}_t) + \mathbf{P}_{t\mu\nu}\partial_{\mathbf{r}}\Big(\phi_\mu(\mathbf{x};\mathbf{r}_t)\phi_\nu(\mathbf{x};\mathbf{r}_t)\Big)\dot{\mathbf{r}}_t. \end{aligned} \tag{86}$$

For the $\mathbf{H}_t^\star = \mathbf{H}(\mathbf{r}_t) + \mathbf{V}_{\mathrm{eff},t}$, let us define the effective potential matrix

$$\mathbf{V}_{\mathrm{eff},t\mu\nu} \coloneqq \int_{\mathbb{R}^3} \phi_\mu(\mathbf{x};\mathbf{r}_t) v_{\mathrm{eff},t}(\mathbf{x})\phi_\nu(\mathbf{x};\mathbf{r}_t)d\mathbf{x},$$

By definition, substituting (86) into the last term of (85) yields

$$\begin{aligned}\int_{\mathbb{R}^3} v_{\mathrm{eff},t}(\mathbf{x})\dot{\rho}_t(\mathbf{x})d\mathbf{x} &= \dot{\mathbf{P}}_{t\mu\nu}\int_{\mathbb{R}^3}\phi_\mu(\mathbf{x};\mathbf{r}_t)v_{\mathrm{eff},t}(\mathbf{x})\phi_\nu(\mathbf{x};\mathbf{r}_t)d\mathbf{x} \\ &\quad + \mathbf{P}_{t\mu\nu}\int_{\mathbb{R}^3} v_{\mathrm{eff},t}(\mathbf{x})\partial_{\mathbf{r}}\Big(\phi_\mu(\mathbf{x};\mathbf{r}_t)\phi_\nu(\mathbf{x};\mathbf{r}_t)\Big)d\mathbf{x}\,\dot{\mathbf{r}}_t \\ &= \dot{\mathbf{P}}_{t\mu\nu}\mathbf{V}_{\mathrm{eff},t\mu\nu} + \Big\langle \partial_{\mathbf{r}}\mathcal{E}_{\mathrm{Hxc}}(\mathbf{r}_t,\mathbf{P}_t),\ \dot{\mathbf{r}}_t\Big\rangle. \end{aligned} \tag{87}$$

Inserting (87) into (85) gives the exact decomposition

$$\frac{d}{dt}\mathcal{E}(\mathbf{r}_t,\mathbf{C}_t) = \mathrm{Tr}\big(\mathbf{H}_t^\star\dot{\mathbf{P}}_t\big) + \Big\langle \partial_{\mathbf{r}}\mathcal{E}(\mathbf{r}_t,\mathbf{P}_t),\ \dot{\mathbf{r}}_t\Big\rangle, \tag{88}$$

where $\partial_{\mathbf{r}}\mathcal{E}$ collects the explicit $\mathbf{r}$-dependence coming from $\partial_{\mathbf{r}}\mathbf{H}$ and the basis dependence in $\rho_{\mathbf{P}}$ and $\mathrm{S}(\mathbf{r})$. Rewrite the $\dot{\mathbf{P}}_t$ term in terms of $\nabla_t\mathbf{C}_t$. Differentiate $\mathbf{P}_t = \mathbf{C}_t\mathbf{C}_t^\dagger\mathrm{S}(\mathbf{r}_t)$:

$$\dot{\mathbf{P}}_t = \dot{\mathbf{C}}_t\mathbf{C}_t^\dagger\mathrm{S}_t + \mathbf{C}_t\dot{\mathbf{C}}_t^\dagger\mathrm{S}_t + \mathbf{C}_t\mathbf{C}_t^\dagger\dot{\mathrm{S}}_t, \quad \mathrm{S}_t \coloneqq \mathrm{S}(\mathbf{r}_t). \tag{89}$$

Insert (89) into the first term of (88) and separate the $\dot{\mathrm{S}}_t$ contribution:

$$\begin{aligned}\mathrm{Tr}\big(\mathbf{H}_t^\star\dot{\mathbf{P}}_t\big) &= \mathrm{Tr}\big(\mathbf{H}_t^\star\dot{\mathbf{C}}_t\mathbf{C}_t^\dagger\mathrm{S}_t\big) + \mathrm{Tr}\big(\mathbf{H}_t^\star\mathbf{C}_t\dot{\mathbf{C}}_t^\dagger\mathrm{S}_t\big) + \mathrm{Tr}\big(\mathbf{H}_t^\star\mathbf{C}_t\mathbf{C}_t^\dagger\dot{\mathrm{S}}_t\big) \\ &= \mathrm{Tr}\big(\mathbf{C}_t^\dagger\mathrm{S}_t\mathbf{H}_t^\star\dot{\mathbf{C}}_t\big) + \mathrm{Tr}\big(\dot{\mathbf{C}}_t^\dagger\mathrm{S}_t\mathbf{H}_t^\star\mathbf{C}_t\big) + \mathrm{Tr}\big(\mathbf{H}_t^\star\mathbf{C}_t\mathbf{C}_t^\dagger\dot{\mathrm{S}}_t\big) \\ &= 2\Re\,\mathrm{Tr}\big(\dot{\mathbf{C}}_t^\dagger\mathrm{S}_t\mathbf{H}_t^\star\mathbf{C}_t\big) + \mathrm{Tr}\big(\mathbf{H}_t^\star\mathbf{C}_t\mathbf{C}_t^\dagger\dot{\mathrm{S}}_t\big). \end{aligned} \tag{90}$$

Define the gauge potential $\mathbf{A}_t \coloneqq \mathbf{C}_t^\dagger\mathrm{S}_t\dot{\mathbf{C}}_t \in \mathfrak{u}(N)$ and $\nabla_t\mathbf{C}_t \coloneqq \dot{\mathbf{C}}_t - \mathbf{C}_t\mathbf{A}_t$. Then

$$\begin{aligned}\Re\,\mathrm{Tr}\big(\dot{\mathbf{C}}_t^\dagger\mathrm{S}_t\mathbf{H}_t^\star\mathbf{C}_t\big) &= \Re\,\mathrm{Tr}\big((\nabla_t\mathbf{C}_t)^\dagger\mathrm{S}_t\mathbf{H}_t^\star\mathbf{C}_t\big) + \Re\,\mathrm{Tr}\big((\mathbf{C}_t\mathbf{A}_t)^\dagger\mathrm{S}_t\mathbf{H}_t^\star\mathbf{C}_t\big) \\ &= \Re\,\mathrm{Tr}\big((\nabla_t\mathbf{C}_t)^\dagger\mathrm{S}_t\mathbf{H}_t^\star\mathbf{C}_t\big) + \underbrace{\Re\,\mathrm{Tr}\big(\mathbf{A}_t^\dagger\mathbf{C}_t^\dagger\mathrm{S}_t\mathbf{H}_t^\star\mathbf{C}_t\big)}_{=0}. \end{aligned} \tag{91}$$

Since $\mathbf{A}_t^\dagger = -\mathbf{A}_t$ and $\mathbf{C}_t^\dagger\mathrm{S}_t\mathbf{H}_t^\star\mathbf{C}_t$ is Hermitian, the second trace in (91) is purely imaginary, hence its real part is zero. By combining results in (90) and (91), we derive

$$\mathrm{Tr}\big(\mathbf{H}_t^\star\dot{\mathbf{P}}_t\big) = 2\Re\,\mathrm{Tr}\big((\nabla_t\mathbf{C}_t)^\dagger\mathrm{S}_t\mathbf{H}_t^\star\mathbf{C}_t\big) + \mathrm{Tr}\big(\mathbf{H}_t^\star\mathbf{C}_t\mathbf{C}_t^\dagger\dot{\mathrm{S}}_t\big) \tag{92}$$

$$= \Big\langle \big(\mathrm{grad}_{\mathbf{C}}^{\mathrm{Hor}}\mathcal{E}\big)^\flat(\mathbf{r}_t,\mathbf{P}_t),\ \nabla_t\mathbf{C}_t\Big\rangle_{\mathrm{F}} + \mathrm{Tr}\big(\mathbf{H}_t^\star\mathbf{C}_t\mathbf{C}_t^\dagger\dot{\mathrm{S}}_t\big). \tag{93}$$

The first term in (93) is exactly the Frobenius pairing, and the remaining second term $\mathrm{Tr}\big(\mathbf{H}_t^\star\mathbf{C}_t\mathbf{C}_t^\dagger\dot{\mathrm{S}}_t\big)$ is an explicit $\mathbf{r}_t$ contribution and is absorbed into $\langle\partial_{\mathbf{r}}\mathcal{E},\dot{\mathbf{r}}_t\rangle$. Thus, substituting (93) into (88) yields the exact chain rule

$$\frac{d}{dt}\mathcal{E}(\mathbf{r}_t,\mathbf{P}_t) = \Big\langle \big(\mathrm{grad}_{\mathbf{C}}^{\mathrm{Hor}}\mathcal{E}\big)^\flat(\mathbf{r}_t,\mathbf{P}_t),\ \nabla_t\mathbf{C}_t\Big\rangle_{\mathrm{F}} + \Big\langle \partial_{\mathbf{r}}\mathcal{E}(\mathbf{r}_t,\mathbf{P}_t),\ \dot{\mathbf{r}}_t\Big\rangle. \tag{94}$$

Next, using the metric-compatibility of the induced covariant derivative, we have

$$\begin{aligned}
\frac{d}{dt}\frac{1}{2}\|\mathbf{V}_t\|_{\mathrm{S}_t}^2 &= \frac{d}{dt}\frac{1}{2}\Re\,\mathrm{Tr}\big(\mathbf{V}_t^\dagger \mathrm{S}_t \mathbf{V}_t\big)\\
&= \frac{1}{2}\Re\,\mathrm{Tr}\Big(\dot{\mathbf{V}}_t^\dagger \mathrm{S}_t \mathbf{V}_t + \mathbf{V}_t^\dagger \dot{\mathrm{S}}_t \mathbf{V}_t + \mathbf{V}_t^\dagger \mathrm{S}_t \dot{\mathbf{V}}_t\Big)\\
&= \Re\,\mathrm{Tr}\big(\dot{\mathbf{V}}_t^\dagger \mathrm{S}_t \mathbf{V}_t\big) + \frac{1}{2}\Re\,\mathrm{Tr}\big(\mathbf{V}_t^\dagger \dot{\mathrm{S}}_t \mathbf{V}_t\big)\\
&= \langle \dot{\mathbf{V}}_t, \mathbf{V}_t\rangle_{\mathrm{S}_t} + \frac{1}{2}\langle \mathbf{V}_t, \dot{\mathrm{S}}_t \mathbf{V}_t\rangle_{\mathrm{F}}\\
&= \langle \dot{\mathbf{V}}_t - \mathbf{V}_t\mathbf{A}_t, \mathbf{V}_t\rangle_{\mathrm{S}_t} + \frac{1}{2}\langle \mathbf{V}_t, \dot{\mathrm{S}}_t \mathbf{V}_t\rangle_{\mathrm{F}} + \Re\,\mathrm{Tr}\big(\mathbf{A}_t^\dagger \mathbf{V}_t^\dagger \mathrm{S}_t \mathbf{V}_t\big)\\
&= \langle \dot{\mathbf{V}}_t - \mathbf{V}_t\mathbf{A}_t, \mathbf{V}_t\rangle_{\mathrm{S}_t} + \frac{1}{2}\langle \mathbf{V}_t, \dot{\mathrm{S}}_t \mathbf{V}_t\rangle_{\mathrm{F}} + \Re\,\mathrm{Tr}\big(-\mathbf{A}_t\mathbf{M}_t\big)\\
&= \langle \dot{\mathbf{V}}_t - \mathbf{V}_t\mathbf{A}_t, \mathbf{V}_t\rangle_{\mathrm{S}_t} + \frac{1}{2}\langle \mathbf{V}_t, \dot{\mathrm{S}}_t \mathbf{V}_t\rangle_{\mathrm{F}}\\
&= \langle \mathrm{Hor}_{\mathbf{C}_t}(\dot{\mathbf{V}}_t - \mathbf{V}_t\mathbf{A}_t), \mathbf{V}_t\rangle_{\mathrm{S}_t} + \frac{1}{2}\langle \mathbf{V}_t, \dot{\mathrm{S}}_t \mathbf{V}_t\rangle_{\mathrm{F}}\\
&= \langle \nabla_t \mathbf{V}_t, \mathbf{V}_t\rangle_{\mathrm{S}_t} + \frac{1}{2}\langle \mathbf{V}_t, \dot{\mathrm{S}}_t \mathbf{V}_t\rangle_{\mathrm{F}}. \qquad (95)
\end{aligned}$$

Equivalently, in momentum form,

$$\frac{d}{dt}\frac{1}{2}\|\mathbf{\Pi}_t\|_{\mathrm{S}_t^{-1}}^2 = \Big\langle \nabla_t \mathbf{\Pi}_t,\ \mathbf{\Pi}_t^\sharp \Big\rangle_{\mathrm{F}} - \frac{1}{2}\Big\langle \mathbf{\Pi}_t,\ \big(\tfrac{d}{dt}\mathrm{S}_t^{-1}\big)\mathbf{\Pi}_t \Big\rangle_{\mathrm{F}}. \qquad (96)$$

Both formulas are exact. From the result in (80), we have

$$\mathbf{V}_t = \nabla_t \mathbf{C}_t = \mathbf{\Pi}_t^\sharp, \quad \mu_t \nabla_t \mathbf{\Pi}_t + \gamma \mathbf{\Pi}_t + \big(\mathrm{grad}_{\mathbf{C}}^{\mathrm{Hor}} \mathcal{E}\big)^\flat(\mathbf{r}_t, \mathbf{P}_t) = 0.$$

Pair the momentum equation with $\mathbf{\Pi}_t^\sharp = \mathbf{V}_t$ under $\langle\cdot,\cdot\rangle_{\mathrm{F}}$:

$$\begin{aligned}
0 &= \Big\langle \mu_t \nabla_t \mathbf{\Pi}_t + \gamma \mathbf{\Pi}_t + \big(\mathrm{grad}_{\mathbf{C}}^{\mathrm{Hor}} \mathcal{E}\big)^\flat,\ \mathbf{\Pi}_t^\sharp \Big\rangle_{\mathrm{F}}\\
&= \mu_t \Big\langle \nabla_t \mathbf{\Pi}_t, \mathbf{\Pi}_t^\sharp \Big\rangle_{\mathrm{F}} + \gamma \Big\langle \mathbf{\Pi}_t, \mathbf{\Pi}_t^\sharp \Big\rangle_{\mathrm{F}} + \Big\langle \big(\mathrm{grad}_{\mathbf{C}}^{\mathrm{Hor}} \mathcal{E}\big)^\flat, \nabla_t \mathbf{C}_t \Big\rangle_{\mathrm{F}}. \qquad (97)
\end{aligned}$$

Since

$$\Big\langle \mathbf{\Pi}_t, \mathbf{\Pi}_t^\sharp \Big\rangle_{\mathrm{F}} = \|\mathbf{\Pi}_t\|_{\mathrm{S}_t^{-1}}^2 = \|\nabla_t \mathbf{C}_t\|_{\mathrm{S}_t}^2,$$

equation (97) gives the exact identity

$$\Big\langle \big(\mathrm{grad}_{\mathbf{C}}^{\mathrm{Hor}} \mathcal{E}\big)^\flat, \nabla_t \mathbf{C}_t \Big\rangle_{\mathrm{F}} = -\mu_t \Big\langle \nabla_t \mathbf{\Pi}_t, \mathbf{\Pi}_t^\sharp \Big\rangle_{\mathrm{F}} - \gamma \|\nabla_t \mathbf{C}_t\|_{\mathrm{S}_t}^2. \qquad (98)$$

Collecting the results together we have:

$$\frac{d}{dt}\mathscr{H}_t = \frac{d}{dt}\mathcal{E}(\mathbf{r}_t, \mathbf{P}_t) + \frac{\dot{\mu}_t}{2}\|\mathbf{V}_t\|_{\mathrm{S}_t}^2 + \mu_t \frac{d}{dt}\frac{1}{2}\|\mathbf{V}_t\|_{\mathrm{S}_t}^2. \qquad (99)$$

We use (98) to replace the pairing $\langle (\mathrm{grad}_{\mathbf{C}}^{\mathrm{Hor}} \mathcal{E})^\flat, \nabla_t \mathbf{C}_t\rangle_{\mathrm{F}}$, and use the exact kinetic identity $\mu_t \frac{d}{dt}\frac{1}{2}\|\mathbf{V}_t\|_{\mathrm{S}_t}^2 = \mu_t \langle \nabla_t \mathbf{\Pi}_t, \mathbf{\Pi}_t^\sharp\rangle_{\mathrm{F}}$ when S_t is treated as fixed in the electronic subproblem. Then (99) becomes

$$\begin{aligned}
\frac{d}{dt}\mathscr{H}_t &= \Big\langle \big(\mathrm{grad}_{\mathbf{C}}^{\mathrm{Hor}} \mathcal{E}\big)^\flat, \nabla_t \mathbf{C}_t \Big\rangle_{\mathrm{F}} + \frac{\dot{\mu}_t}{2}\|\mathbf{V}_t\|_{\mathrm{S}_t}^2 + \mu_t \Big\langle \nabla_t \mathbf{\Pi}_t, \mathbf{\Pi}_t^\sharp \Big\rangle_{\mathrm{F}} + \Big\langle \partial_{\mathbf{r}} \mathcal{E}(\mathbf{r}_t, \mathbf{P}_t), \dot{\mathbf{r}}_t \Big\rangle\\
&= \Big(-\mu_t \Big\langle \nabla_t \mathbf{\Pi}_t, \mathbf{\Pi}_t^\sharp \Big\rangle_{\mathrm{F}} - \gamma \|\mathbf{V}_t\|_{\mathrm{S}_t}^2\Big) + \frac{\dot{\mu}_t}{2}\|\mathbf{V}_t\|_{\mathrm{S}_t}^2 + \mu_t \Big\langle \nabla_t \mathbf{\Pi}_t, \mathbf{\Pi}_t^\sharp \Big\rangle_{\mathrm{F}} + \Big\langle \partial_{\mathbf{r}} \mathcal{E}(\mathbf{r}_t, \mathbf{P}_t), \dot{\mathbf{r}}_t \Big\rangle\\
&= -\gamma \|\mathbf{V}_t\|_{\mathrm{S}_t}^2 + \frac{\dot{\mu}_t}{2}\|\mathbf{V}_t\|_{\mathrm{S}_t}^2 + \Big\langle \partial_{\mathbf{r}} \mathcal{E}(\mathbf{r}_t, \mathbf{P}_t), \dot{\mathbf{r}}_t \Big\rangle. \qquad (100)
\end{aligned}$$

While the nuclear trajectory is frozen during the electronic relaxation step, namely $\dot{\mathbf{r}}_t = \mathbf{0}$ and hence $\dot{\mathrm{S}}_t = \mathbf{0}$, then (100) reduces to the exact dissipation identity

$$\frac{d}{dt}\Big[\mathcal{E}(\mathbf{r}, \mathbf{P}_t) + \frac{\mu_t}{2}\|\nabla_t \mathbf{C}_t\|_{\mathrm{S}}^2\Big] = -\Big(\gamma + \frac{a}{2}\mu_t\Big)\|\nabla_t \mathbf{C}_t\|_{\mathrm{S}}^2 \le 0. \qquad (101)$$

for $\mu_t = e^{-at}$ we have $\dot{\mu}_t = -a\mu_t$. This provides an exact monotone decay law for a Lyapunov function of the dual dynamics. □

Lemma 11.6 (Explicit evaluation of a valid L_ρ for the KS functional in a finite AO basis)**.** *Fix a nuclear geometry* $\mathbf{r}$*, and let* $f(\mathbf{P}) \coloneqq \mathcal{E}(\mathbf{r}, \mathbf{P})$ *be the Kohn–Sham electronic energy in a finite AO basis, with* ϵ_{xc} *of class* C^3 *on the range of admissible densities. Then, for every* $\mathbf{P} \in \mathrm{Gr}_{\mathrm{S}}(M, N)$*,*

$$\big\|\nabla(\mathrm{Hess} f)(\mathbf{P})\big\|_{\mathrm{op}} \le L_\rho, \tag{102}$$

where L_ρ *is the explicit constant in* (109)*, determined by* $\mathbf{H}$*, the two-electron integrals,* $\mathrm{S}(\mathbf{r})$*, the AO basis,* N*, and the first three derivatives of* ϵ_{xc}*. Consequently,* $\|\mathbf{B}'(s)\|_{\mathrm{op}} \le L_\rho \, \|\gamma'(s)\|_{\mathcal{M}}$ *along any Grassmann geodesic* $\gamma(s)$ *with transported Hessian operator* $\mathbf{B}(s)$*.*

Proof. As in the previous settings, we abbreviate $\mathrm{S} \coloneqq \mathrm{S}(\mathbf{r})$ and work throughout in the real Hilbert space of complex matrices

$$\mathbb{V} \coloneqq \mathbb{C}^{M\times M}, \quad \langle \mathbf{A}, \mathbf{B}\rangle_{\mathrm{F}} \coloneqq \Re\mathbf{Tr}(\mathbf{A}^\dagger \mathbf{B}), \quad \|\mathbf{A}\|_{\mathrm{F}}^2 \coloneqq \langle \mathbf{A}, \mathbf{A}\rangle_{\mathrm{F}}.$$

All operator norms $\|\cdot\|_{\mathrm{op}}$ below are induced by $\|\cdot\|_{\mathrm{F}}$, both for linear and for multilinear maps. In the AO basis $\{\phi_\mu(\cdot;\mathbf{r})\}_{\mu=1}^M$, the energy in the statement reads

$$f(\mathbf{P}) = \mathrm{Tr}(\mathbf{H}\mathbf{P}) + \frac{1}{2}\sum_{\mu,\nu,\kappa,\lambda} (\mu\nu|\kappa\lambda)\mathbf{P}_{\mu\nu}\mathbf{P}_{\kappa\lambda} + \int_{\mathbb{R}^3} \epsilon_{\mathrm{xc}}\big(\rho_{\mathbf{P}}(\mathbf{x})\big)d\mathbf{x}, \tag{103}$$

$$\rho_{\mathbf{P}}(\mathbf{x}) = \sum_{\mu,\nu} \mathbf{P}_{\mu\nu}\phi_\mu(\mathbf{x};\mathbf{r})\phi_\nu(\mathbf{x};\mathbf{r}), \tag{104}$$

and $\mathrm{Gr}_{\mathrm{S}}(M,N)$ carries the Grassmann metric induced by the Frobenius pairing on the corresponding S-orthogonal projectors. Here L_ρ is assembled from the AO product matrix and the condition number of S,

$$\mathbf{A}_{\mu\nu}(\mathbf{x}) \coloneqq \phi_\mu(\mathbf{x};\mathbf{r})\phi_\nu(\mathbf{x};\mathbf{r}), \quad \kappa(\mathrm{S}) \coloneqq \frac{\lambda_{\max}(\mathrm{S})}{\lambda_{\min}(\mathrm{S})},$$

the density bound

$$\rho_{\max} \coloneqq \sqrt{\kappa(\mathrm{S})}\sqrt{N}\sup_{\mathbf{x}\in\mathbb{R}^3}\big\|\mathbf{A}(\mathbf{x})\big\|_{\mathrm{F}}, \tag{105}$$

and the computable scalar constants

$$A_1 \coloneqq \int_{\mathbb{R}^3}\|\mathbf{A}(\mathbf{x})\|_{\mathrm{F}}d\mathbf{x}, \quad A_2 \coloneqq \int_{\mathbb{R}^3}\|\mathbf{A}(\mathbf{x})\|_{\mathrm{F}}^2 d\mathbf{x}, \quad A_3 \coloneqq \int_{\mathbb{R}^3}\|\mathbf{A}(\mathbf{x})\|_{\mathrm{F}}^3 d\mathbf{x}, \tag{106}$$

$$M_k \coloneqq \max_{u\in[0,\rho_{\max}]}\big|\epsilon_{\mathrm{xc}}^{(k)}(u)\big|, \quad k\in\{1,2,3\}, \tag{107}$$

$$\|\mathcal{J}\|_{\mathrm{op}} \coloneqq \sup_{\|\Delta\mathbf{P}\|_{\mathrm{F}}=1}\|\mathbf{J}(\Delta\mathbf{P})\|_{\mathrm{F}}, \quad \mathbf{J}(\mathbf{P})_{\mu\nu} \coloneqq \sum_{\kappa,\lambda}(\mu\nu|\kappa\lambda)\mathbf{P}_{\kappa\lambda}. \tag{108}$$

By (111) below, the range of admissible densities in the statement is $[0,\rho_{\max}]$; accordingly, the regularity assumption is $\epsilon_{\mathrm{xc}} \in C^3([0,\rho_{\max}])$, and the explicit constant of the statement is

$$L_\rho \coloneqq \kappa(\mathrm{S})^{3/2}M_3A_3 + 6\kappa(\mathrm{S})\Big(\|\mathcal{J}\|_{\mathrm{op}} + M_2A_2\Big) + 12\kappa(\mathrm{S})^{1/2}\Big(\|\mathbf{H}\|_{\mathrm{F}} + \|\mathcal{J}\|_{\mathrm{op}}\sqrt{\kappa(\mathrm{S})}\sqrt{N} + M_1A_1\Big). \tag{109}$$

Our first step is to show a uniform pointwise bound. Let $\mathbf{P} = \mathbf{C}\mathbf{C}^\dagger\mathrm{S}$ be any admissible rank-N S-projector (i.e. $\mathbf{C}^\dagger\mathrm{S}\mathbf{C} = \mathbf{I}_N$), and introduce the S-orthogonalization

$$\widetilde{\mathbf{C}} \coloneqq \mathrm{S}^{1/2}\mathbf{C} \in \mathbb{C}^{M\times N}, \quad \widetilde{\mathbf{C}}^\dagger\widetilde{\mathbf{C}} = \mathbf{I}_N, \quad \widetilde{\mathbf{P}} \coloneqq \widetilde{\mathbf{C}}\widetilde{\mathbf{C}}^\dagger.$$

Then $\widetilde{\mathbf{P}}$ is an orthogonal projector of rank N, which is also Hermitian and idempotent, and

$$\mathbf{P} = \mathbf{C}\mathbf{C}^\dagger\mathrm{S} = \mathrm{S}^{-1/2}\,\widetilde{\mathbf{C}}\widetilde{\mathbf{C}}^\dagger\,\mathrm{S}^{1/2} = \mathrm{S}^{-1/2}\,\widetilde{\mathbf{P}}\,\mathrm{S}^{1/2}.$$

Hence, using $\|\mathbf{A}\mathbf{B}\mathbf{C}\|_{\mathrm{F}} \le \|\mathbf{A}\|_2\|\mathbf{B}\|_{\mathrm{F}}\|\mathbf{C}\|_2$ and $\|\widetilde{\mathbf{P}}\|_{\mathrm{F}}^2 = \mathbf{Tr}(\widetilde{\mathbf{P}}^2) = \mathbf{Tr}(\widetilde{\mathbf{P}}) = N$,

$$\|\mathbf{P}\|_{\mathrm{F}} \le \|\mathrm{S}^{-1/2}\|_2\,\|\widetilde{\mathbf{P}}\|_{\mathrm{F}}\,\|\mathrm{S}^{1/2}\|_2 = \sqrt{\kappa(\mathrm{S})}\,\sqrt{N}. \tag{110}$$

Now fix $\mathbf{x} \in \mathbb{R}^3$. By (104), the density is the Frobenius pairing of $\mathbf{A}(\mathbf{x})$ with $\mathbf{P}$,

$$\rho_{\mathbf{P}}(\mathbf{x}) = \sum_{\mu,\nu} \mathbf{P}_{\mu\nu}\, \mathbf{A}_{\mu\nu}(\mathbf{x}) = \langle \mathbf{A}(\mathbf{x}), \mathbf{P} \rangle_{\mathrm{F}}.$$

Therefore, by Cauchy–Schwarz in $\langle \cdot, \cdot \rangle_{\mathrm{F}}$ and (110),

$$|\rho_{\mathbf{P}}(\mathbf{x})| \le \|\mathbf{A}(\mathbf{x})\|_{\mathrm{F}}\, \|\mathbf{P}\|_{\mathrm{F}} \le \sqrt{\kappa(\mathrm{S})}\, \sqrt{N}\, \|\mathbf{A}(\mathbf{x})\|_{\mathrm{F}} \le \rho_{\max}, \tag{111}$$

with $\rho_{\max}$ as in (105). In particular, $\rho_{\mathbf{P}}(\mathbf{x}) \in [0, \rho_{\max}]$ for all admissible densities, so the constants $M_k = \max_{u \in [0,\rho_{\max}]} |\epsilon_{\mathrm{xc}}^{(k)}(u)|$ are valid uniformly on the manifold. Next, we decompose $f = f_{1\mathrm{e}} + f_{\mathrm{H}} + f_{\mathrm{xc}}$ with

$$f_{1\mathrm{e}}(\mathbf{P}) = \mathbf{Tr}(\mathbf{H}\mathbf{P}), \quad f_{\mathrm{H}}(\mathbf{P}) = \tfrac{1}{2} \sum_{\mu,\nu,\kappa,\lambda} (\mu\nu|\kappa\lambda)\, \mathbf{P}_{\mu\nu}\mathbf{P}_{\kappa\lambda}, \quad f_{\mathrm{xc}}(\mathbf{P}) = \int_{\mathbb{R}^3} \epsilon_{\mathrm{xc}}\big(\rho_{\mathbf{P}}(\mathbf{x})\big)\, d\mathbf{x}.$$

For any $\Delta\mathbf{P} \in \mathbb{V}$,

$$Df_{1\mathrm{e}}(\mathbf{P})[\Delta\mathbf{P}] = \Re\mathbf{Tr}(\mathbf{H}^{\dagger}\Delta\mathbf{P}) = \langle \mathbf{H}, \Delta\mathbf{P} \rangle_{\mathrm{F}},$$

so the Euclidean gradient is $\nabla f_{1\mathrm{e}}(\mathbf{P}) = \mathbf{H}$, and

$$\|Df_{1\mathrm{e}}(\mathbf{P})\|_{\mathrm{op}} = \|\mathbf{H}\|_{\mathrm{F}}, \quad \|D^2 f_{1\mathrm{e}}(\mathbf{P})\|_{\mathrm{op}} = 0, \quad \|D^3 f_{1\mathrm{e}}(\mathbf{P})\|_{\mathrm{op}} = 0. \tag{112}$$

Next, recall the linear operator $\mathbf{J} : \mathbb{V} \to \mathbb{V}$ of (108). Then $f_{\mathrm{H}}(\mathbf{P}) = \frac{1}{2}\langle \mathbf{J}(\mathbf{P}), \mathbf{P} \rangle_{\mathrm{F}}$ and, for any $\Delta\mathbf{P}$, a direct computation gives

$$Df_{\mathrm{H}}(\mathbf{P})[\Delta\mathbf{P}] = \tfrac{1}{2}\Big(\langle \mathbf{J}(\Delta\mathbf{P}), \mathbf{P} \rangle_{\mathrm{F}} + \langle \mathbf{J}(\mathbf{P}), \Delta\mathbf{P} \rangle_{\mathrm{F}} \Big) = \langle \mathbf{J}(\mathbf{P}), \Delta\mathbf{P} \rangle_{\mathrm{F}}, \tag{113}$$

where the last equality uses that $\mathbf{J}$ is self-adjoint with respect to $\langle \cdot, \cdot \rangle_{\mathrm{F}}$, which follows from the usual permutational symmetry of two-electron Coulomb integrals. Hence $\nabla f_{\mathrm{H}}(\mathbf{P}) = \mathbf{J}(\mathbf{P})$ and $D(\nabla f_{\mathrm{H}})(\mathbf{P}) = \mathbf{J}$ is constant, so

$$\|Df_{\mathrm{H}}(\mathbf{P})\|_{\mathrm{op}} = \|\mathbf{J}(\mathbf{P})\|_{\mathrm{F}} \le \|\mathcal{J}\|_{\mathrm{op}}\, \|\mathbf{P}\|_{\mathrm{F}} \le \|\mathcal{J}\|_{\mathrm{op}}\, \sqrt{\kappa(\mathrm{S})}\, \sqrt{N}, \tag{114}$$

$$\|D^2 f_{\mathrm{H}}(\mathbf{P})\|_{\mathrm{op}} = \|\mathcal{J}\|_{\mathrm{op}}, \quad \|D^3 f_{\mathrm{H}}(\mathbf{P})\|_{\mathrm{op}} = 0. \tag{115}$$

For the exchange–correlation term, set $\rho(\mathbf{x}) \coloneqq \rho_{\mathbf{P}}(\mathbf{x}) = \langle \mathbf{A}(\mathbf{x}), \mathbf{P} \rangle_{\mathrm{F}}$. For $\Delta\mathbf{P} \in \mathbb{V}$, the chain rule gives

$$Df_{\mathrm{xc}}(\mathbf{P})[\Delta\mathbf{P}] = \int_{\mathbb{R}^3} \epsilon_{\mathrm{xc}}'\big(\rho(\mathbf{x})\big)\, D\rho_{\mathbf{P}}(\mathbf{x})[\Delta\mathbf{P}]\, d\mathbf{x} = \int_{\mathbb{R}^3} \epsilon_{\mathrm{xc}}'\big(\rho(\mathbf{x})\big)\, \langle \mathbf{A}(\mathbf{x}), \Delta\mathbf{P} \rangle_{\mathrm{F}}\, d\mathbf{x}. \tag{116}$$

Thus $\nabla f_{\mathrm{xc}}(\mathbf{P})$ is the Riesz representative

$$\nabla f_{\mathrm{xc}}(\mathbf{P}) = \int_{\mathbb{R}^3} \epsilon_{\mathrm{xc}}'\big(\rho(\mathbf{x})\big)\, \mathbf{A}(\mathbf{x})\, d\mathbf{x}, \tag{117}$$

and by (111) and the definition of M_1 and A_1,

$$\|\nabla f_{\mathrm{xc}}(\mathbf{P})\|_{\mathrm{F}} \le \int_{\mathbb{R}^3} \big|\epsilon_{\mathrm{xc}}'\big(\rho(\mathbf{x})\big)\big|\, \|\mathbf{A}(\mathbf{x})\|_{\mathrm{F}}\, d\mathbf{x} \le M_1 A_1. \tag{118}$$

Differentiating (116) once more, for $\Delta\mathbf{P}_1, \Delta\mathbf{P}_2 \in \mathbb{V}$,

$$D^2 f_{\mathrm{xc}}(\mathbf{P})[\Delta\mathbf{P}_1, \Delta\mathbf{P}_2] = \int_{\mathbb{R}^3} \epsilon_{\mathrm{xc}}''\big(\rho(\mathbf{x})\big)\, \langle \mathbf{A}(\mathbf{x}), \Delta\mathbf{P}_1 \rangle_{\mathrm{F}}\, \langle \mathbf{A}(\mathbf{x}), \Delta\mathbf{P}_2 \rangle_{\mathrm{F}}\, d\mathbf{x}. \tag{119}$$

Equivalently, the Euclidean Hessian operator $\nabla^2 f_{\mathrm{xc}}(\mathbf{P}) : \mathbb{V} \to \mathbb{V}$ satisfies

$$\nabla^2 f_{\mathrm{xc}}(\mathbf{P})[\Delta\mathbf{P}] = \int_{\mathbb{R}^3} \epsilon_{\mathrm{xc}}''\big(\rho(\mathbf{x})\big)\, \langle \mathbf{A}(\mathbf{x}), \Delta\mathbf{P} \rangle_{\mathrm{F}}\, \mathbf{A}(\mathbf{x})\, d\mathbf{x},$$

and hence, by Cauchy–Schwarz and $|\epsilon_{\mathrm{xc}}''| \le M_2$,

$$\begin{aligned} \|\nabla^2 f_{\mathrm{xc}}(\mathbf{P})[\Delta\mathbf{P}]\|_{\mathrm{F}} &\le \int_{\mathbb{R}^3} \big|\epsilon_{\mathrm{xc}}''\big(\rho(\mathbf{x})\big)\big|\, \big|\langle \mathbf{A}(\mathbf{x}), \Delta\mathbf{P} \rangle_{\mathrm{F}}\big|\, \|\mathbf{A}(\mathbf{x})\|_{\mathrm{F}}\, d\mathbf{x} \\ &\le M_2 \int_{\mathbb{R}^3} \|\mathbf{A}(\mathbf{x})\|_{\mathrm{F}}^2\, d\mathbf{x}\, \|\Delta\mathbf{P}\|_{\mathrm{F}} = M_2 A_2\, \|\Delta\mathbf{P}\|_{\mathrm{F}}. \end{aligned} \tag{120}$$

Therefore

$$\|D^2 f_{\mathrm{xc}}(\mathbf{P})\|_{\mathrm{op}} = \|\nabla^2 f_{\mathrm{xc}}(\mathbf{P})\|_{\mathrm{op}} \le M_2 A_2. \tag{121}$$

Finally, differentiating (119) once more yields the third derivative: for $\Delta\mathbf{P}_1, \Delta\mathbf{P}_2, \Delta\mathbf{P}_3 \in \mathbb{V}$,

$$D^3 f_{\mathrm{xc}}(\mathbf{P})[\Delta\mathbf{P}_1, \Delta\mathbf{P}_2, \Delta\mathbf{P}_3] = \int_{\mathbb{R}^3} \epsilon_{\mathrm{xc}}^{(3)}\big(\rho(\mathbf{x})\big) \prod_{j=1}^{3} \langle \mathbf{A}(\mathbf{x}), \Delta\mathbf{P}_j\rangle_{\mathrm{F}}\, d\mathbf{x}, \tag{122}$$

so, using $|\epsilon_{\mathrm{xc}}^{(3)}| \le M_3$ and $|\langle \mathbf{A}(\mathbf{x}), \Delta\mathbf{P}_j\rangle_{\mathrm{F}}| \le \|\mathbf{A}(\mathbf{x})\|_{\mathrm{F}}\, \|\Delta\mathbf{P}_j\|_{\mathrm{F}}$,

$$\big|D^3 f_{\mathrm{xc}}(\mathbf{P})[\Delta\mathbf{P}_1, \Delta\mathbf{P}_2, \Delta\mathbf{P}_3]\big| \le M_3 \int_{\mathbb{R}^3} \|\mathbf{A}(\mathbf{x})\|_{\mathrm{F}}^3\, d\mathbf{x} \prod_{j=1}^{3} \|\Delta\mathbf{P}_j\|_{\mathrm{F}} = M_3 A_3 \prod_{j=1}^{3} \|\Delta\mathbf{P}_j\|_{\mathrm{F}}. \tag{123}$$

Hence

$$\|D^3 f_{\mathrm{xc}}(\mathbf{P})\|_{\mathrm{op}} \le M_3 A_3. \tag{124}$$

Collecting (112), (115), (118), (121), and (124), we obtain the uniform Euclidean bounds

$$\|Df(\mathbf{P})\|_{\mathrm{op}} \le \|\mathbf{H}\|_{\mathrm{F}} + \|\mathcal{J}\|_{\mathrm{op}} \sqrt{\kappa(\mathrm{S})}\sqrt{N} + M_1 A_1, \tag{125}$$

$$\|D^2 f(\mathbf{P})\|_{\mathrm{op}} \le \|\mathcal{J}\|_{\mathrm{op}} + M_2 A_2, \tag{126}$$

$$\|D^3 f(\mathbf{P})\|_{\mathrm{op}} \le M_3 A_3. \tag{127}$$

In the projector coordinate chart with Frobenius-induced metric, $\mathrm{Gr}_{\mathrm{S}}(M, N)$ is represented by the orthogonal projectors $\widetilde{\mathbf{P}} = \mathrm{S}^{1/2}\mathbf{P}\,\mathrm{S}^{-1/2}$, on which the energy is the function $\widetilde{\mathbf{P}} \mapsto f(\mathrm{S}^{-1/2}\widetilde{\mathbf{P}}\,\mathrm{S}^{1/2})$. Because $\widetilde{\mathbf{P}} \mapsto \mathrm{S}^{-1/2}\widetilde{\mathbf{P}}\,\mathrm{S}^{1/2}$ is linear, for $k \in \{1, 2, 3\}$ and $\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_k \in \mathbb{V}$,

$$D^k\big[f(\mathrm{S}^{-1/2}\,\cdot\,\mathrm{S}^{1/2})\big](\widetilde{\mathbf{P}})[\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_k] = D^k f(\mathbf{P})\big[\mathrm{S}^{-1/2}\boldsymbol{\eta}_1\mathrm{S}^{1/2}, \dots, \mathrm{S}^{-1/2}\boldsymbol{\eta}_k\mathrm{S}^{1/2}\big]. \tag{128}$$

By the estimate used in (110),

$$\|\mathrm{S}^{-1/2}\boldsymbol{\eta}\,\mathrm{S}^{1/2}\|_{\mathrm{F}} \le \|\mathrm{S}^{-1/2}\|_2\, \|\boldsymbol{\eta}\|_{\mathrm{F}}\, \|\mathrm{S}^{1/2}\|_2 = \sqrt{\kappa(\mathrm{S})}\, \|\boldsymbol{\eta}\|_{\mathrm{F}}, \quad \boldsymbol{\eta} \in \mathbb{V},$$

and (128) yields

$$\big\|D^k\big[f(\mathrm{S}^{-1/2}\,\cdot\,\mathrm{S}^{1/2})\big](\widetilde{\mathbf{P}})\big\|_{\mathrm{op}} \le \kappa(\mathrm{S})^{k/2}\, \big\|D^k f(\mathbf{P})\big\|_{\mathrm{op}}, \quad k \in \{1, 2, 3\}. \tag{129}$$

In the rest of the proof we use the notation of this chart: $\mathbf{P}$ denotes the orthogonal projector $\widetilde{\mathbf{P}}$, which satisfies $\mathbf{P} = \mathbf{P}^\dagger = \mathbf{P}^2$, and f denotes $f(\mathrm{S}^{-1/2}\,\cdot\,\mathrm{S}^{1/2})$, whose derivatives obey (125)–(127) up to the factors in (129). The tangent projection is the linear map

$$\Pi_{\mathbf{P}}(\mathbf{Z}) \coloneqq \mathbf{P}\mathbf{Z}(\mathbf{I} - \mathbf{P}) + (\mathbf{I} - \mathbf{P})\mathbf{Z}\mathbf{P}, \quad \|\Pi_{\mathbf{P}}\|_{\mathrm{op}} = 1, \tag{130}$$

and a short calculation shows that the first and second Fréchet derivatives of $\Pi_{\mathbf{P}}$ are

$$D\Pi_{\mathbf{P}}[\boldsymbol{\eta}](\mathbf{Z}) = \boldsymbol{\eta}\,\mathbf{Z}(\mathbf{I} - 2\mathbf{P}) + (\mathbf{I} - 2\mathbf{P})\mathbf{Z}\,\boldsymbol{\eta}, \tag{131}$$

$$D^2\Pi_{\mathbf{P}}[\boldsymbol{\eta}, \boldsymbol{\xi}](\mathbf{Z}) = -2\boldsymbol{\eta}\,\mathbf{Z}\,\boldsymbol{\xi} - 2\boldsymbol{\xi}\,\mathbf{Z}\,\boldsymbol{\eta}. \tag{132}$$

Since $\mathbf{I} - 2\mathbf{P}$ has eigenvalues ± 1, one has $\|\mathbf{I} - 2\mathbf{P}\|_2 = 1$ and therefore, using $\|\mathbf{A}\mathbf{B}\|_{\mathrm{F}} \le \|\mathbf{A}\|_2 \|\mathbf{B}\|_{\mathrm{F}}$ and $\|\boldsymbol{\eta}\|_2 \le \|\boldsymbol{\eta}\|_{\mathrm{F}}$,

$$\begin{aligned} \|D\Pi_{\mathbf{P}}[\boldsymbol{\eta}](\mathbf{Z})\|_{\mathrm{F}} &\le \|\boldsymbol{\eta}\,\mathbf{Z}(\mathbf{I} - 2\mathbf{P})\|_{\mathrm{F}} + \|(\mathbf{I} - 2\mathbf{P})\mathbf{Z}\,\boldsymbol{\eta}\|_{\mathrm{F}} \\ &\le \|\boldsymbol{\eta}\|_2 \|\mathbf{Z}\|_{\mathrm{F}} + \|\mathbf{Z}\|_{\mathrm{F}} \|\boldsymbol{\eta}\|_2 \le 2\|\boldsymbol{\eta}\|_{\mathrm{F}}\, \|\mathbf{Z}\|_{\mathrm{F}}, \end{aligned} \tag{133}$$

$$\begin{aligned} \|D^2\Pi_{\mathbf{P}}[\boldsymbol{\eta}, \boldsymbol{\xi}](\mathbf{Z})\|_{\mathrm{F}} &\le 2\|\boldsymbol{\eta}\,\mathbf{Z}\,\boldsymbol{\xi}\|_{\mathrm{F}} + 2\|\boldsymbol{\xi}\,\mathbf{Z}\,\boldsymbol{\eta}\|_{\mathrm{F}} \\ &\le 2\|\boldsymbol{\eta}\|_2 \|\mathbf{Z}\|_{\mathrm{F}} \|\boldsymbol{\xi}\|_2 + 2\|\boldsymbol{\xi}\|_2 \|\mathbf{Z}\|_{\mathrm{F}} \|\boldsymbol{\eta}\|_2 \le 4\|\boldsymbol{\eta}\|_{\mathrm{F}}\, \|\boldsymbol{\xi}\|_{\mathrm{F}}\, \|\mathbf{Z}\|_{\mathrm{F}}. \end{aligned} \tag{134}$$

Equivalently, for all directions $\boldsymbol{\eta}, \boldsymbol{\xi} \in \mathbb{V}$, tangent or not,

$$\|D\Pi_{\mathbf{P}}[\boldsymbol{\eta}]\|_{\mathrm{op}} \le 2\|\boldsymbol{\eta}\|_{\mathrm{F}}, \quad \|D^2\Pi_{\mathbf{P}}[\boldsymbol{\eta}, \boldsymbol{\xi}]\|_{\mathrm{op}} \le 4\|\boldsymbol{\eta}\|_{\mathrm{F}}\, \|\boldsymbol{\xi}\|_{\mathrm{F}}. \tag{135}$$

Let $\mathbf{g}(\mathbf{P}) \coloneqq \nabla f(\mathbf{P}) \in \mathbb{V}$ denote the Euclidean gradient of Riesz representatives. The Riemannian gradient on the Grassmann chart is $\operatorname{grad} f(\mathbf{P}) = \Pi_{\mathbf{P}}\mathbf{g}(\mathbf{P})$, and the Riemannian Hessian is the $(1,1)$-tensor

$$\operatorname{Hess} f(\mathbf{P})[\boldsymbol{\xi}] \;=\; \Pi_{\mathbf{P}}\Big(D(\operatorname{grad} f)(\mathbf{P})[\boldsymbol{\xi}]\Big) \quad (\boldsymbol{\xi} \in T_{\mathbf{P}}\mathrm{Gr}).$$

Using $\operatorname{grad} f = \Pi_{\mathbf{P}} \circ \mathbf{g}$ and differentiating yields

$$\begin{aligned}\operatorname{Hess} f(\mathbf{P})[\boldsymbol{\xi}] &= \Pi_{\mathbf{P}}\Big(D\Pi_{\mathbf{P}}[\boldsymbol{\xi}]\mathbf{g}(\mathbf{P})\Big) + \Pi_{\mathbf{P}}\Big(D\mathbf{g}(\mathbf{P})[\boldsymbol{\xi}]\Big)\\ &= \Pi_{\mathbf{P}}\Big(D\Pi_{\mathbf{P}}[\boldsymbol{\xi}]\mathbf{g}(\mathbf{P})\Big) + \Pi_{\mathbf{P}}\Big(\nabla^2 f(\mathbf{P})[\boldsymbol{\xi}]\Big),\end{aligned} \tag{136}$$

where $\nabla^2 f(\mathbf{P}) : \mathbb{V} \to \mathbb{V}$ is the Euclidean Hessian operator. Since the right-hand side of (136) is linear in $\boldsymbol{\xi}$ and well defined for every $\boldsymbol{\xi} \in \mathbb{V}$, we write $\operatorname{Hess} f(\mathbf{P})[\mathbf{Z}]$ for this expression when $\mathbf{Z} \in \mathbb{V}$; it lies in the range of $\Pi_{\mathbf{P}}$.

Now fix $\boldsymbol{\eta}, \boldsymbol{\xi} \in T_{\mathbf{P}}\mathrm{Gr}$ and compute the covariant derivative $(\nabla_{\boldsymbol{\eta}}\operatorname{Hess} f)(\mathbf{P})[\boldsymbol{\xi}]$. Because the Levi–Civita connection is the orthogonal projection of the ambient derivative, we extend $\boldsymbol{\xi}$ to the tangent field $\mathbf{Q} \mapsto \Pi_{\mathbf{Q}}\boldsymbol{\xi}$ on Gr. Differentiating $\Pi_{\mathbf{Q}} \circ \Pi_{\mathbf{Q}} = \Pi_{\mathbf{Q}}$ along a curve in Gr through $\mathbf{P}$ with velocity $\boldsymbol{\eta}$ gives

$$D\Pi_{\mathbf{P}}[\boldsymbol{\eta}] \circ \Pi_{\mathbf{P}} + \Pi_{\mathbf{P}} \circ D\Pi_{\mathbf{P}}[\boldsymbol{\eta}] = D\Pi_{\mathbf{P}}[\boldsymbol{\eta}],$$

and composing with $\Pi_{\mathbf{P}}$ on the left and on the right yields

$$2\,\Pi_{\mathbf{P}} \circ D\Pi_{\mathbf{P}}[\boldsymbol{\eta}] \circ \Pi_{\mathbf{P}} = \Pi_{\mathbf{P}} \circ D\Pi_{\mathbf{P}}[\boldsymbol{\eta}] \circ \Pi_{\mathbf{P}}, \quad \text{that is,} \quad \Pi_{\mathbf{P}} \circ D\Pi_{\mathbf{P}}[\boldsymbol{\eta}] \circ \Pi_{\mathbf{P}} = 0.$$

Since $\Pi_{\mathbf{P}}\boldsymbol{\xi} = \boldsymbol{\xi}$, this gives

$$\Pi_{\mathbf{P}}\big(D\Pi_{\mathbf{P}}[\boldsymbol{\eta}]\,\boldsymbol{\xi}\big) = 0. \tag{137}$$

Thus the covariant derivative of the extension along $\boldsymbol{\eta}$ vanishes at $\mathbf{P}$, whereas its ambient derivative $D\Pi_{\mathbf{P}}[\boldsymbol{\eta}]\,\boldsymbol{\xi}$ is a normal vector, nonzero in general. By the product rule,

$$\begin{aligned}(\nabla_{\boldsymbol{\eta}}\operatorname{Hess} f)(\mathbf{P})[\boldsymbol{\xi}] &= \Pi_{\mathbf{P}}\, D\Big(\operatorname{Hess} f(\cdot)\big[\Pi_{(\cdot)}\boldsymbol{\xi}\big]\Big)(\mathbf{P})[\boldsymbol{\eta}]\\ &= \Pi_{\mathbf{P}}\, D\big(\operatorname{Hess} f(\cdot)[\boldsymbol{\xi}]\big)(\mathbf{P})[\boldsymbol{\eta}] + \operatorname{Hess} f(\mathbf{P})\big[D\Pi_{\mathbf{P}}[\boldsymbol{\eta}]\,\boldsymbol{\xi}\big],\end{aligned} \tag{138}$$

where the projection in front of the second term is omitted because $\operatorname{Hess} f(\mathbf{P})$ takes values in the range of $\Pi_{\mathbf{P}}$. Differentiating (136) in the first term and using $\|\Pi_{\mathbf{P}}\|_{\mathrm{op}} = 1$ gives

$$\begin{aligned}\big\|(\nabla_{\boldsymbol{\eta}}\operatorname{Hess} f)(\mathbf{P})[\boldsymbol{\xi}]\big\|_{\mathrm{F}} \le{}& \big\|D\Pi_{\mathbf{P}}[\boldsymbol{\eta}]\big(\nabla^2 f(\mathbf{P})[\boldsymbol{\xi}]\big)\big\|_{\mathrm{F}} + \big\|D(\nabla^2 f)(\mathbf{P})[\boldsymbol{\eta}][\boldsymbol{\xi}]\big\|_{\mathrm{F}}\\ &+ \big\|D\Pi_{\mathbf{P}}[\boldsymbol{\eta}]\big(D\Pi_{\mathbf{P}}[\boldsymbol{\xi}]\mathbf{g}(\mathbf{P})\big)\big\|_{\mathrm{F}} + \big\|D\big(D\Pi_{\mathbf{P}}[\boldsymbol{\xi}]\mathbf{g}(\mathbf{P})\big)[\boldsymbol{\eta}]\big\|_{\mathrm{F}}\\ &+ \big\|\operatorname{Hess} f(\mathbf{P})\big[D\Pi_{\mathbf{P}}[\boldsymbol{\eta}]\,\boldsymbol{\xi}\big]\big\|_{\mathrm{F}}.\end{aligned} \tag{139}$$

For the fourth term we use the product rule:

$$\begin{aligned}D\big(D\Pi_{\mathbf{P}}[\boldsymbol{\xi}]\mathbf{g}(\mathbf{P})\big)[\boldsymbol{\eta}] &= D^2\Pi_{\mathbf{P}}[\boldsymbol{\eta},\boldsymbol{\xi}]\mathbf{g}(\mathbf{P}) + D\Pi_{\mathbf{P}}[\boldsymbol{\xi}]\big(D\mathbf{g}(\mathbf{P})[\boldsymbol{\eta}]\big)\\ &= D^2\Pi_{\mathbf{P}}[\boldsymbol{\eta},\boldsymbol{\xi}]\mathbf{g}(\mathbf{P}) + D\Pi_{\mathbf{P}}[\boldsymbol{\xi}]\big(\nabla^2 f(\mathbf{P})[\boldsymbol{\eta}]\big).\end{aligned}$$

For the last term, (136) with $D\Pi_{\mathbf{P}}[\boldsymbol{\eta}]\,\boldsymbol{\xi}$ in place of $\boldsymbol{\xi}$, $\|\Pi_{\mathbf{P}}\|_{\mathrm{op}} = 1$, and (135) give

$$\begin{aligned}\big\|\operatorname{Hess} f(\mathbf{P})\big[D\Pi_{\mathbf{P}}[\boldsymbol{\eta}]\,\boldsymbol{\xi}\big]\big\|_{\mathrm{F}} &\le \Big(2\|\mathbf{g}(\mathbf{P})\|_{\mathrm{F}} + \|\nabla^2 f(\mathbf{P})\|_{\mathrm{op}}\Big)\,\big\|D\Pi_{\mathbf{P}}[\boldsymbol{\eta}]\,\boldsymbol{\xi}\big\|_{\mathrm{F}}\\ &\le \Big(4\|\mathbf{g}(\mathbf{P})\|_{\mathrm{F}} + 2\|\nabla^2 f(\mathbf{P})\|_{\mathrm{op}}\Big)\,\|\boldsymbol{\eta}\|_{\mathrm{F}}\,\|\boldsymbol{\xi}\|_{\mathrm{F}}.\end{aligned} \tag{140}$$

Combining the last two displays with (139) and applying (135) yields

$$\begin{aligned}\big\|(\nabla_{\boldsymbol{\eta}}\operatorname{Hess} f)(\mathbf{P})[\boldsymbol{\xi}]\big\|_{\mathrm{F}} &\le \|D(\nabla^2 f)(\mathbf{P})\|_{\mathrm{op}}\|\boldsymbol{\eta}\|_{\mathrm{F}}\,\|\boldsymbol{\xi}\|_{\mathrm{F}}\\ &\quad + \|D\Pi_{\mathbf{P}}[\boldsymbol{\eta}]\|_{\mathrm{op}}\,\|\nabla^2 f(\mathbf{P})\|_{\mathrm{op}}\,\|\boldsymbol{\xi}\|_{\mathrm{F}} + \|D\Pi_{\mathbf{P}}[\boldsymbol{\xi}]\|_{\mathrm{op}}\,\|\nabla^2 f(\mathbf{P})\|_{\mathrm{op}}\,\|\boldsymbol{\eta}\|_{\mathrm{F}}\\ &\quad + \|D\Pi_{\mathbf{P}}[\boldsymbol{\eta}]\|_{\mathrm{op}}\,\|D\Pi_{\mathbf{P}}[\boldsymbol{\xi}]\|_{\mathrm{op}}\,\|\mathbf{g}(\mathbf{P})\|_{\mathrm{F}} + \|D^2\Pi_{\mathbf{P}}[\boldsymbol{\eta},\boldsymbol{\xi}]\|_{\mathrm{op}}\,\|\mathbf{g}(\mathbf{P})\|_{\mathrm{F}}\\ &\quad + \Big(4\|\mathbf{g}(\mathbf{P})\|_{\mathrm{F}} + 2\|\nabla^2 f(\mathbf{P})\|_{\mathrm{op}}\Big)\,\|\boldsymbol{\eta}\|_{\mathrm{F}}\,\|\boldsymbol{\xi}\|_{\mathrm{F}}\\ &\le \|D^3 f(\mathbf{P})\|_{\mathrm{op}}\,\|\boldsymbol{\eta}\|_{\mathrm{F}}\,\|\boldsymbol{\xi}\|_{\mathrm{F}} + 6\|\nabla^2 f(\mathbf{P})\|_{\mathrm{op}}\,\|\boldsymbol{\eta}\|_{\mathrm{F}}\,\|\boldsymbol{\xi}\|_{\mathrm{F}} + 12\|\mathbf{g}(\mathbf{P})\|_{\mathrm{F}}\,\|\boldsymbol{\eta}\|_{\mathrm{F}}\,\|\boldsymbol{\xi}\|_{\mathrm{F}}.\end{aligned} \tag{141}$$

Taking the supremum over $\|\boldsymbol{\eta}\|_{\mathrm{F}} = \|\boldsymbol{\xi}\|_{\mathrm{F}} = 1$ gives the *geometric reduction*

$$\|\nabla(\mathrm{Hess} f)(\mathbf{P})\|_{\mathrm{op}} \leq \|D^3 f(\mathbf{P})\|_{\mathrm{op}} + 6\|D^2 f(\mathbf{P})\|_{\mathrm{op}} + 12\|Df(\mathbf{P})\|_{\mathrm{op}}. \tag{142}$$

Combining (142) with (129) and (125)–(127) yields, uniformly in $\mathbf{P}$,

$$\begin{aligned}\|\nabla(\mathrm{Hess} f)(\mathbf{P})\|_{\mathrm{op}} \leq \kappa(\mathrm{S})^{3/2} M_3 A_3 + 6\kappa(\mathrm{S})\big(\|\mathcal{J}\|_{\mathrm{op}} + M_2 A_2\big) \\ + 12\kappa(\mathrm{S})^{1/2}\big(\|\mathbf{H}\|_{\mathrm{F}} + \|\mathcal{J}\|_{\mathrm{op}}\sqrt{\kappa(\mathrm{S})}\sqrt{N} + M_1 A_1\big) = L_\rho,\end{aligned} \tag{143}$$

which proves (102). Finally, let $\gamma(s)$ be any Grassmann geodesic and let $\mathbf{B}(s)$ be the transported Hessian operator along γ,

$$\mathbf{B}(s) \coloneqq \mathcal{T}_{s\to 0} \circ \mathrm{Hess} f(\gamma(s)) \circ \mathcal{T}_{0\to s},$$

where $\mathcal{T}$ is parallel transport. Differentiating and using that $\mathcal{T}$ is an isometry gives the standard estimate

$$\begin{aligned}\|\mathbf{B}'(s)\|_{\mathrm{op}} &= \big\|\mathcal{T}_{s\to 0} \circ (\nabla_{\gamma'(s)}\mathrm{Hess} f)(\gamma(s)) \circ \mathcal{T}_{0\to s}\big\|_{\mathrm{op}} \\ &= \|(\nabla_{\gamma'(s)}\mathrm{Hess} f)(\gamma(s))\|_{\mathrm{op}} \leq \|\nabla(\mathrm{Hess} f)(\gamma(s))\|_{\mathrm{op}}\, \|\gamma'(s)\|_{\mathcal{M}},\end{aligned}$$

and inserting (102) yields the claimed inequality $\|\mathbf{B}'(s)\|_{\mathrm{op}} \leq L_\rho \|\gamma'(s)\|_{\mathcal{M}}$. □

Lemma 11.7 (Local error bound near a stable BO minimizer)**.** *Assume $\mathbf{P}_\star(\mathbf{r})$ is an isolated stable minimizer of $\mathbf{P} \mapsto \mathcal{E}(\mathbf{r}, \mathbf{P})$ on $\mathrm{Gr}_{\mathrm{S}}(M, N)$, and assume $\mathcal{E}(\mathbf{r}, \mathbf{P})$ is C^2 in a neighborhood of $\mathbf{P}_\star(\mathbf{r})$. Then there exist $\delta, c > 0$ such that, for every admissible projector $\mathbf{P} = \mathbf{C}\mathbf{C}^\dagger \mathrm{S}$ satisfying $\|\mathbf{P} - \mathbf{P}_\star(\mathbf{r})\|_{\mathrm{F}} \le \delta$, one has*

$$\|\mathbf{P} - \mathbf{P}_\star(\mathbf{r})\|_{\mathrm{F}} \le \frac{2\sqrt{2\kappa(\mathrm{S}(\mathbf{r}))}}{m} \big\|\mathrm{grad}^{\mathrm{Hor}}_{\mathbf{C}} \mathcal{E}(\mathbf{r}, \mathbf{C})\big\|_{\mathrm{S}}.$$

Proof. For simplicity, let us abbreviate $\mathrm{S} = \mathrm{S}(\mathbf{r})$ and denote $\mathcal{M} \coloneqq \mathrm{Gr}_{\mathrm{S}}(M, N)$, $f(\mathbf{P}) \coloneqq \mathcal{E}(\mathbf{r}, \mathbf{P})$. Let $\mathbf{P}_\star \coloneqq \mathbf{P}_\star(\mathbf{r})$ be the isolated stable minimizer of energy functional. and $\pi : \mathrm{St}_{\mathrm{S}}(M, N) \to \mathcal{M}$ be the canonical projection $\pi(\mathbf{C}) = \mathbf{P}(\mathbf{C}) \coloneqq \mathbf{C}\mathbf{C}^\dagger \mathrm{S}$. Since stability of the solution $\mathbf{P}^\star$ directly implies a uniform positive definiteness of the Riemannian Hessian at the minimizer, and f is C^2 near $\mathbf{P}_\star$, the first-order optimality condition holds: $\mathrm{grad}_{\mathbf{P}} f(\mathbf{P}_\star) = \mathbf{0}$, and the second-order sufficient condition holds with a strict spectral gap. There exists a constant $m > 0$ such that for every tangent vector $\boldsymbol{\xi} \in T_{\mathbf{P}_\star}\mathcal{M}$,

$$\big\langle \mathrm{Hess}_{\mathbf{P}} f(\mathbf{P}_\star)[\boldsymbol{\xi}],\ \boldsymbol{\xi} \big\rangle_{\mathcal{M}} \ge m \|\boldsymbol{\xi}\|^2_{\mathcal{M}}. \tag{144}$$

Here $\langle \cdot, \cdot \rangle_{\mathcal{M}}$ and $\| \cdot \|_{\mathcal{M}}$ denote the Riemannian metric and norm on $\mathcal{M}$ induced from $(\mathrm{St}_{\mathrm{S}}, \langle \cdot, \cdot \rangle_{\mathrm{S}})$. From the inequality (144), we derive a lower bound on the operator action in the following step. Let $\mathbf{H}_\star \coloneqq \mathrm{Hess}_{\mathbf{P}} f(\mathbf{P}_\star) : T_{\mathbf{P}_\star}\mathcal{M} \to T_{\mathbf{P}_\star}\mathcal{M}$. Since $\mathbf{H}_\star$ is self-adjoint with respect to $\langle \cdot, \cdot \rangle_{\mathcal{M}}$, Cauchy–Schwarz directly gives

$$\begin{aligned}
\big\|\mathbf{H}_\star[\boldsymbol{\xi}]\big\|_{\mathcal{M}} \|\boldsymbol{\xi}\|_{\mathcal{M}} &= \sqrt{\big\langle \mathbf{H}_\star[\boldsymbol{\xi}], \mathbf{H}_\star[\boldsymbol{\xi}] \big\rangle_{\mathcal{M}}} \sqrt{\big\langle \boldsymbol{\xi}, \boldsymbol{\xi} \big\rangle_{\mathcal{M}}} \\
&\ge \big| \big\langle \mathbf{H}_\star[\boldsymbol{\xi}], \boldsymbol{\xi} \big\rangle_{\mathcal{M}} \big| \\
&\ge m \big\langle \boldsymbol{\xi}, \boldsymbol{\xi} \big\rangle_{\mathcal{M}} = m \|\boldsymbol{\xi}\|^2_{\mathcal{M}},
\end{aligned}$$

for some $m \ge 0$. While f is C^2 in a neighborhood of $\mathbf{P}_\star$, the covariant derivative of the Hessian is continuous. Fix any radius $\rho > 0$ smaller than the injectivity radius of $\mathcal{M}$ at $\mathbf{P}_\star$ and small enough so that the closed geodesic ball $\overline{B}_\rho(\mathbf{P}_\star) \coloneqq \{\mathbf{P} \in \mathcal{M} :\ d_{\mathcal{M}}(\mathbf{P}, \mathbf{P}_\star) \le \rho\}$ lies in a coordinate neighborhood where f is C^2. Since $\overline{B}_\rho(\mathbf{P}_\star)$ is compact and the gradient of Hessian is continuous, there exists $L_\rho < \infty$ such that

$$\sup_{\mathbf{P} \in \overline{B}_\rho(\mathbf{P}_\star)} \big\|\nabla(\mathrm{Hess}_{\mathbf{P}} f)(\mathbf{P})\big\|_{\mathrm{op}} \le L_\rho. \tag{145}$$

Here $\| \cdot \|_{\mathrm{op}}$ denotes the operator norm induced by $\| \cdot \|_{\mathcal{M}}$. In order to improve the inequality, we first take arbitrary point $\mathbf{P} \in \overline{B}_\rho(\mathbf{P}_\star)$ and let $\boldsymbol{\xi}$ be a mapped point via Riemannian logarithm:

$$\boldsymbol{\xi} \coloneqq \mathrm{Log}_{\mathbf{P}_\star}(\mathbf{P}) \in T_{\mathbf{P}_\star}\mathcal{M}, \quad \|\boldsymbol{\xi}\|_{\mathcal{M}} = d_{\mathcal{M}}(\mathbf{P}, \mathbf{P}_\star).$$

Let $\gamma : [0, 1] \to \mathcal{M}$ be the geodesic on manifolds starting from $\gamma(0) = \mathbf{P}_\star$ to $\gamma(1) = \mathbf{P}$ where initial vector is given as $\gamma'(0) = \boldsymbol{\xi}$. By definition, we have $\gamma(s) \coloneqq \mathrm{Exp}_{\mathbf{P}_\star}(s\boldsymbol{\xi})$. Let $\mathcal{T}_{s \to 0} : T_{\gamma(s)}\mathcal{M} \to T_{\mathbf{P}_\star}\mathcal{M}$ denote parallel transport along γ. While $\mathcal{T}_{s \to 0}$ is parallel transport, one can find

$$\frac{d}{ds} \mathcal{T}_{s \to 0} \mathrm{grad}_{\mathbf{P}} f(\gamma(s)) = \mathcal{T}_{s \to 0} \nabla_{\gamma'(s)} \big(\mathrm{grad}_{\mathbf{P}} f\big)(\gamma(s)) = \mathcal{T}_{s \to 0} \mathrm{Hess}_{\mathbf{P}} f(\gamma(s))[\gamma'(s)].$$

Let $X(s)$, η be an arbitrary tangent vector and vector fields on $T_{\gamma(s)}\mathcal{M}$ and $T_{\gamma(0)}\mathcal{M}$, respectively. Then, the definition of the Riemannian norm gives

$$\begin{aligned}
\left\langle \frac{d}{ds}\Big(\mathcal{T}_{s \to 0} X(s)\Big), \eta \right\rangle_{\gamma(0)} &= \frac{d}{ds} \left\langle \mathcal{T}_{s \to 0} X(s), \eta \right\rangle_{\gamma(0)} \\
&= \frac{d}{ds} \left\langle X(s), \mathcal{T}_{0 \to s} \eta \right\rangle_{\gamma(s)} \\
&= \left\langle \nabla_{\gamma'(s)} X(s), \mathcal{T}_{0 \to s} \eta \right\rangle_{\gamma(s)} + \left\langle X(s), \nabla_{\gamma'(s)} \big(\mathcal{T}_{0 \to s} \eta\big) \right\rangle_{\gamma(s)} \\
&= \left\langle \nabla_{\gamma'(s)} X(s), \mathcal{T}_{0 \to s} \eta \right\rangle_{\gamma(s)} \\
&= \left\langle \mathcal{T}_{s \to 0} \nabla_{\gamma'(s)} X(s), \eta \right\rangle_{\gamma(0)}.
\end{aligned}$$

By setting $X(s) = \mathrm{grad}_{\mathbf{P}} f(\gamma(s))$, the direct consequence follows

$$\frac{d}{ds}\Big(\mathcal{T}_{s\to 0}\mathrm{grad}_{\mathbf{P}} f(\gamma(s))\Big) = \mathcal{T}_{s\to 0}\nabla_{\gamma'(s)}\big(\mathrm{grad}_{\mathbf{P}} f\big)(\gamma(s)) \tag{146}$$

$$= \mathcal{T}_{s\to 0}\mathrm{Hess}_{\mathbf{P}} f(\gamma(s))[\gamma'(s)]. \tag{147}$$

By taking integration over the interval $[0,1]$, the optimality condition $\mathrm{grad}_{\mathbf{P}} f(\mathbf{P}_\star) = \mathbf{0}$ gives

$$\mathcal{T}_{1\to 0}\mathrm{grad}_{\mathbf{P}} f(\mathbf{P}) = \int_0^1 \mathcal{T}_{s\to 0}\mathrm{Hess}_{\mathbf{P}} f(\gamma(s))[\gamma'(s)]ds. \tag{148}$$

While γ is a geodesic, $\gamma'(s)$ is parallel along γ. Thus, $\gamma'(s) = \mathcal{T}_{0\to s}\boldsymbol{\xi}$ and $\|\gamma'(s)\|_{\mathcal{M}} = \|\boldsymbol{\xi}\|_{\mathcal{M}}$ for all $s \in [0,1]$. Let us consider the transported Hessian operator on $T_{\mathbf{P}_\star}\mathcal{M}$,

$$\mathbf{B}(s) \coloneqq \mathcal{T}_{s\to 0} \circ \mathrm{Hess}_{\mathbf{P}} f(\gamma(s)) \circ \mathcal{T}_{0\to s}, \quad \mathcal{T}_{1\to 0}\mathrm{grad}_{\mathbf{P}} f(\mathbf{P}) = \int_0^1 \mathbf{B}(s)[\boldsymbol{\xi}]ds.$$

Split $\mathbf{B}(s)$ into $\mathbf{B}(0)$ plus deviation:

$$\int_0^1 \mathbf{B}(s)[\boldsymbol{\xi}]ds = \mathbf{B}(0)[\boldsymbol{\xi}] + \int_0^1 \big(\mathbf{B}(s) - \mathbf{B}(0)\big)[\boldsymbol{\xi}]ds.$$

Note $\mathbf{B}(0) = \mathrm{Hess}_{\mathbf{P}} f(\mathbf{P}_\star) = \mathbf{H}_\star$. Taking norms and using that parallel transport is an isometry yields

$$\begin{aligned}\big\|\mathrm{grad}_{\mathbf{P}} f(\mathbf{P})\big\|_{\mathcal{M}} &= \big\|\mathcal{T}_{1\to 0}\mathrm{grad}_{\mathbf{P}} f(\mathbf{P})\big\|_{\mathcal{M}} \\ &\geq \big\|\mathbf{H}_\star[\boldsymbol{\xi}]\big\|_{\mathcal{M}} - \int_0^1 \big\|\mathbf{B}(s) - \mathbf{B}(0)\big\|_{\mathrm{op}}\|\boldsymbol{\xi}\|_{\mathcal{M}}ds. \end{aligned}\tag{149}$$

By using the identity $\frac{d}{ds}\mathcal{T}_{s\to 0} = -\mathcal{T}_{s\to 0} \circ \nabla_{\gamma'(s)}$ with $\frac{d}{ds}\mathcal{T}_{0\to s} = \nabla_{\gamma'(s)} \circ \mathcal{T}_{0\to s}$, one obtains

$$\mathbf{B}'(s) = \mathcal{T}_{s\to 0}\nabla_{\gamma'(s)}\big(\mathrm{Hess}_{\mathbf{P}} f\big)(\gamma(s))\mathcal{T}_{0\to s}.$$

By using the result, one can improve the lower bound of gradient norm in the following:

$$\begin{aligned}\big\|\mathrm{grad}_{\mathbf{P}} f(\mathbf{P})\big\|_{\mathcal{M}} &\geq m\|\boldsymbol{\xi}\|_{\mathcal{M}} - \int_0^1 \|\mathbf{B}(s) - \mathbf{B}(0)\|_{\mathrm{op}}\,\|\boldsymbol{\xi}\|_{\mathcal{M}}\,ds \\ &\geq m\|\boldsymbol{\xi}\|_{\mathcal{M}} - \int_0^1 \left(\int_0^s \|\mathbf{B}'(\tau)\|_{\mathrm{op}}\,d\tau\right)\|\boldsymbol{\xi}\|_{\mathcal{M}}\,ds \\ &\geq m\|\boldsymbol{\xi}\|_{\mathcal{M}} - \int_0^1 \left(\int_0^s \big\|\nabla(\mathrm{Hess} f)(\gamma(\tau))\big\|_{\mathrm{op}}\,\|\gamma'(\tau)\|_{\mathcal{M}}\,d\tau\right)\|\boldsymbol{\xi}\|_{\mathcal{M}}\,ds \\ &\geq m\|\boldsymbol{\xi}\|_{\mathcal{M}} - \int_0^1 \left(\int_0^s L_\rho\|\boldsymbol{\xi}\|_{\mathcal{M}}\,d\tau\right)\|\boldsymbol{\xi}\|_{\mathcal{M}}\,ds \\ &= m\|\boldsymbol{\xi}\|_{\mathcal{M}} - \int_0^1 L_\rho s\|\boldsymbol{\xi}\|_{\mathcal{M}}^2\,ds \\ &= m\|\boldsymbol{\xi}\|_{\mathcal{M}} - \frac{L_\rho}{2}\|\boldsymbol{\xi}\|_{\mathcal{M}}^2. \end{aligned}\tag{150}$$

Next, we choose ρ smaller if necessary to satisfy $m - \frac{L_\rho}{2}\rho \geq \frac{m}{2}$. Then, for every $\mathbf{P} \in \overline{B}_\rho(\mathbf{P}_\star)$, $\|\boldsymbol{\xi}\|_{\mathcal{M}} \leq \rho$, the following follows directly from (150):

$$\big\|\mathrm{grad}_{\mathbf{P}} f(\mathbf{P})\big\|_{\mathcal{M}} \geq \frac{m}{2}\|\boldsymbol{\xi}\|_{\mathcal{M}} = \frac{m}{2}d_{\mathcal{M}}(\mathbf{P},\mathbf{P}_\star). \tag{151}$$

Equivalently, one can conclude

$$d_{\mathcal{M}}(\mathbf{P},\mathbf{P}_\star) \leq \frac{2}{m}\big\|\mathrm{grad}_{\mathbf{P}} f(\mathbf{P})\big\|_{\mathcal{M}}. \tag{152}$$

Next, we convert the geodesic distance to ambient Frobenius distance of projectors. Introduce the similarity transform to the standard Grassmann manifold:

$$\widetilde{\mathbf{P}} \coloneqq \mathrm{S}^{1/2}\mathbf{P}\mathrm{S}^{-1/2}, \quad \widetilde{\mathbf{P}}_\star \coloneqq \mathrm{S}^{1/2}\mathbf{P}_\star\mathrm{S}^{-1/2}.$$

Then $\widetilde{\mathbf{P}}$ and $\widetilde{\mathbf{P}}_\star$ are rank-N orthogonal projectors in the usual sense. Moreover, the map $\mathbf{P} \mapsto \widetilde{\mathbf{P}}$ is an isometry between $(\mathcal{M}, d_{\mathcal{M}})$ and the standard Grassmann, because $\mathbf{C} \mapsto \mathrm{S}^{1/2}\mathbf{C}$ is an isometry from $(\mathrm{St_S}, \langle\cdot,\cdot\rangle_{\mathrm{S}})$ to $(\mathrm{St_I}, \langle\cdot,\cdot\rangle_{\mathrm{F}})$ and the quotient metric is induced by Riemannian submersion. Hence, we have $d_{\mathcal{M}}(\mathbf{P}, \mathbf{P}_\star) = d_{\mathrm{Gr}}(\widetilde{\mathbf{P}}, \widetilde{\mathbf{P}}_\star)$. For orthogonal projectors, let $\theta_1, \ldots, \theta_N \in [0, \pi/2]$ denote the principal angles between the two subspaces. A standard canonical-angle normal form yields

$$\|\widetilde{\mathbf{P}} - \widetilde{\mathbf{P}}_\star\|_{\mathrm{F}}^2 = 2\sum_{i=1}^{N} \sin^2\theta_i, \quad d_{\mathrm{Gr}}(\widetilde{\mathbf{P}}, \widetilde{\mathbf{P}}_\star)^2 = \sum_{i=1}^{N} \theta_i^2.$$

Since $\sin\theta_i \le \theta_i$ for $\theta_i \in [0, \pi/2]$, one obtains

$$\|\widetilde{\mathbf{P}} - \widetilde{\mathbf{P}}_\star\|_{\mathrm{F}}^2 = 2\sum_{i=1}^{N} \sin^2\theta_i \le 2\sum_{i=1}^{N} \theta_i^2 = 2d_{\mathrm{Gr}}(\widetilde{\mathbf{P}}, \widetilde{\mathbf{P}}_\star)^2 \le \sqrt{2} d_{\mathcal{M}}(\mathbf{P}, \mathbf{P}_\star),$$

By using the property $\mathbf{P} - \mathbf{P}_\star = \mathrm{S}^{-1/2}(\widetilde{\mathbf{P}} - \widetilde{\mathbf{P}}_\star)\mathrm{S}^{1/2}$, one induces

$$\|\mathbf{P} - \mathbf{P}_\star\|_{\mathrm{F}} \le \|\mathrm{S}^{-1/2}\|_2 \|\mathrm{S}^{1/2}\|_2 \|\widetilde{\mathbf{P}} - \widetilde{\mathbf{P}}_\star\|_{\mathrm{F}} = \sqrt{\kappa(\mathrm{S})}\|\widetilde{\mathbf{P}} - \widetilde{\mathbf{P}}_\star\|_{\mathrm{F}},$$

where $\kappa(\mathrm{S}) \coloneqq \lambda_{\max}(\mathrm{S})/\lambda_{\min}(\mathrm{S})$. Combining this result gives

$$\|\mathbf{P} - \mathbf{P}_\star\|_{\mathrm{F}} \le \sqrt{2\kappa(\mathrm{S})} d_{\mathcal{M}}(\mathbf{P}, \mathbf{P}_\star). \tag{153}$$

As $f \circ \pi(\mathbf{C}) = \mathcal{E}(\mathbf{r}, \mathbf{C})$ is gauge-invariant, its differential annihilates vertical directions, hence its Riemannian gradient on $\mathrm{St_S}$ is horizontal. The map $D\pi(\mathbf{C})$ restricted to the horizontal space is an isometry. Therefore, the horizontal gradient is the unique horizontal lift of the Grassmann gradient:

$$D\pi(\mathbf{C})\big[\mathrm{grad}_{\mathbf{C}}^{\mathrm{Hor}} \mathcal{E}(\mathbf{r}, \mathbf{C})\big] = \mathrm{grad}_{\mathbf{P}} f(\mathbf{P}), \quad \big\|\mathrm{grad}_{\mathbf{C}}^{\mathrm{Hor}} \mathcal{E}(\mathbf{r}, \mathbf{C})\big\|_{\mathrm{S}} = \big\|\mathrm{grad}_{\mathbf{P}} f(\mathbf{P})\big\|_{\mathcal{M}}.$$

Insert (152) into (153) and use the norm identity above:

$$\begin{aligned}
\big\|\mathbf{P} - \mathbf{P}_\star\big\|_{\mathrm{F}} &\le \sqrt{2\kappa(\mathrm{S})} d_{\mathcal{M}}\big(\mathbf{P}, \mathbf{P}_\star\big) && (154)\\
&\le \sqrt{2\kappa(\mathrm{S})} \frac{2}{m} \big\|\mathrm{grad}_{\mathbf{P}} f(\mathbf{P})\big\|_{\mathcal{M}} && (155)\\
&= \frac{2\sqrt{2\kappa(\mathrm{S})}}{m} \big\|\mathrm{grad}_{\mathbf{C}}^{\mathrm{Hor}} \mathcal{E}(\mathbf{r}, \mathbf{C})\big\|_{\mathrm{S}}. && (156)
\end{aligned}$$

The main argument requires $\mathbf{P} \in \overline{B}_\rho(\mathbf{P}_\star)$, and hence we choose $\delta > 0$ such that $\|\mathbf{P} - \mathbf{P}_\star\|_{\mathrm{F}} \le \delta$ implies $\mathbf{P} \in \overline{B}_\rho(\mathbf{P}_\star)$. This is possible because $\mathbf{P} \mapsto \widetilde{\mathbf{P}}$ is continuous and $\|\widetilde{\mathbf{P}} - \widetilde{\mathbf{P}}_\star\|_{\mathrm{F}} \to 0$ whenever $\|\mathbf{P} - \mathbf{P}_\star\|_{\mathrm{F}} \to 0$. This completes the proof. □

Proposition 11.8 (Dual electronic dynamics can be used as BOMD)**.** *Assume that for each nuclear geometry $\mathbf{r} \in \Omega$ there exists a unique stable Born–Oppenheimer minimizer $\mathbf{P}_\star(\mathbf{r})$ of the electronic energy $\mathcal{E}(\mathbf{r}, \mathbf{P})$ over admissible rank-N projectors, and define $\mathbf{E}^\star(\mathbf{r}) \coloneqq \mathcal{E}(\mathbf{r}, \mathbf{P}_\star(\mathbf{r}))$ and $\mathbf{E}_{\mathrm{BO}}(\mathbf{r}) \coloneqq \mathbf{E}_{\mathrm{nn}}(\mathbf{r}) + \mathbf{E}^\star(\mathbf{r})$. Let $(\mathbf{C}_t, \mathbf{\Pi}_t, \mathbf{P}_t)$ be a proposed Lagrangian system of electronic information along $\mathbf{r}_t$, drive nuclei by the instantaneous force*

$$\mathbf{m}_I \ddot{\mathbf{r}}_I(t) = -\nabla_{\mathbf{r}_I}\Big(\mathbf{E}_{\mathrm{nn}}(\mathbf{r}_t) + \mathcal{E}(\mathbf{r}_t, \mathbf{P}_t)\Big).$$

Then, the residual

$$\mathbf{R}_I(t) := -\nabla_{\mathbf{r}_I}\mathcal{E}(\mathbf{r}_t, \mathbf{P}_t) + \nabla_{\mathbf{r}_I}\mathbf{E}^\star(\mathbf{r}_t),$$

and moreover $\|\mathbf{R}_I(t)\| \to 0$ whenever the electronic consistency error vanishes ($\mathrm{grad}^{\mathrm{Hor}}_{\mathbf{C}}\mathcal{E}(\mathbf{r}_t, \mathbf{C}_t) \to 0$).

Proof. We start from the imposed nuclear equation and rewrite it by adding and subtracting the Born–Oppenheimer force. For each nucleus I, one has

$$\mathbf{m}_I \ddot{\mathbf{r}}_I(t) = -\nabla_{\mathbf{r}_I}\Big(\mathbf{E}_{\mathrm{nn}}(\mathbf{r}_t) + \mathcal{E}(\mathbf{r}_t, \mathbf{P}_t)\Big) = -\nabla_{\mathbf{r}_I}\mathbf{E}_{\mathrm{nn}}(\mathbf{r}_t) - \nabla_{\mathbf{r}_I}\mathcal{E}(\mathbf{r}_t, \mathbf{P}_t). \tag{157}$$

Insert $0 = \nabla_{\mathbf{r}_I}\mathbf{E}^\star(\mathbf{r}_t) - \nabla_{\mathbf{r}_I}\mathbf{E}^\star(\mathbf{r}_t)$ and regroup:

$$\begin{aligned}
\mathbf{m}_I \ddot{\mathbf{r}}_I(t) &= -\nabla_{\mathbf{r}_I}\mathbf{E}_{\mathrm{nn}}(\mathbf{r}_t) - \nabla_{\mathbf{r}_I}\mathcal{E}(\mathbf{r}_t, \mathbf{P}_t) + \nabla_{\mathbf{r}_I}\mathbf{E}^\star(\mathbf{r}_t) - \nabla_{\mathbf{r}_I}\mathbf{E}^\star(\mathbf{r}_t) \\
&= -\nabla_{\mathbf{r}_I}\Big(\mathbf{E}_{\mathrm{nn}}(\mathbf{r}_t) + \mathbf{E}^\star(\mathbf{r}_t)\Big) + \Big(-\nabla_{\mathbf{r}_I}\mathcal{E}(\mathbf{r}_t, \mathbf{P}_t) + \nabla_{\mathbf{r}_I}\mathbf{E}^\star(\mathbf{r}_t)\Big) \\
&= -\nabla_{\mathbf{r}_I}\mathbf{E}_{\mathrm{BO}}(\mathbf{r}_t) + \mathbf{R}_I(t),
\end{aligned} \tag{158}$$

where $\mathbf{E}_{\mathrm{BO}} = \mathbf{E}_{\mathrm{nn}} + \mathbf{E}^\star$ and $\mathbf{R}_I$ is exactly the residual stated in the proposition. This already establishes the BOMD form with an explicit correction term $\mathbf{R}_I(t)$. To connect this expression to the true Born–Oppenheimer force, it suffices to relate $\nabla_{\mathbf{r}}\mathbf{E}^\star(\mathbf{r})$ to the minimized electronic force and then quantify how $\mathbf{R}_I(t)$ depends on the electronic tracking error. Thus, the optimal energy can be rewritten as

$$\mathbf{E}^\star(\mathbf{r}) = \min_{\mathbf{C} \in \mathrm{St}_{\mathrm{S}(\mathbf{r})}(M,N)} \mathcal{E}\big(\mathbf{r}, \mathbf{P}(\mathbf{C}; \mathbf{r})\big) = \min_{[\mathbf{C}] \in \mathrm{Gr}_{\mathrm{S}(\mathbf{r})}(M,N)} \mathcal{E}(\mathbf{r}, [\mathbf{C}]).$$

Let $\mathbf{C}_\star(\mathbf{r}) \in \mathrm{St}_{\mathrm{S}(\mathbf{r})}(M, N)$ be any minimizer which is unique only up to the gauge $\mathbf{C}_\star(\mathbf{r}) \sim \mathbf{C}_\star(\mathbf{r})\mathbf{U}$ for $\mathbf{U} \in \mathrm{U}(N)$. it induces the projector minimizer

$$\mathbf{P}_\star(\mathbf{r}) = \mathbf{P}(\mathbf{C}_\star(\mathbf{r}); \mathbf{r}) = \mathbf{C}_\star(\mathbf{r})\mathbf{C}_\star(\mathbf{r})^\dagger \mathrm{S}(\mathbf{r}).$$

Consider the directional increment $\mathbf{r}_\varepsilon := \mathbf{r} + \varepsilon\delta\mathbf{r}$ and note that for every ε, by optimality of $\mathbf{P}_\star(\mathbf{r})$ at $\mathbf{r}$ one has the upper bound

$$\begin{aligned}
\mathbf{E}^\star(\mathbf{r}_\varepsilon) - \mathbf{E}^\star(\mathbf{r}) &= \min_{\mathbf{P}} \mathcal{E}(\mathbf{r}_\varepsilon, \mathbf{P}) - \mathcal{E}(\mathbf{r}, \mathbf{P}_\star(\mathbf{r})) \\
&\le \mathcal{E}(\mathbf{r}_\varepsilon, \mathbf{P}_\star(\mathbf{r})) - \mathcal{E}(\mathbf{r}, \mathbf{P}_\star(\mathbf{r})).
\end{aligned} \tag{159}$$

Divide by ε and let $\varepsilon \to 0$ (using differentiability of $\mathcal{E}$ in $\mathbf{r}$):

$$\begin{aligned}
\limsup_{\varepsilon \to 0} \frac{\mathbf{E}^\star(\mathbf{r}_\varepsilon) - \mathbf{E}^\star(\mathbf{r})}{\varepsilon} &\le \lim_{\varepsilon \to 0} \frac{\mathcal{E}(\mathbf{r}_\varepsilon, \mathbf{P}_\star(\mathbf{r})) - \mathcal{E}(\mathbf{r}, \mathbf{P}_\star(\mathbf{r}))}{\varepsilon} \\
&= \nabla_{\mathbf{r}}\mathcal{E}(\mathbf{r}, \mathbf{P}_\star(\mathbf{r})) \cdot \delta\mathbf{r}.
\end{aligned} \tag{160}$$

Similarly, by optimality of $\mathbf{P}_\star(\mathbf{r}_\varepsilon)$ at $\mathbf{r}_\varepsilon$ one has

$$\begin{aligned}
\mathbf{E}^\star(\mathbf{r}_\varepsilon) - \mathbf{E}^\star(\mathbf{r}) &= \mathcal{E}(\mathbf{r}_\varepsilon, \mathbf{P}_\star(\mathbf{r}_\varepsilon)) - \min_{\mathbf{P}} \mathcal{E}(\mathbf{r}, \mathbf{P}) \\
&\ge \mathcal{E}(\mathbf{r}_\varepsilon, \mathbf{P}_\star(\mathbf{r}_\varepsilon)) - \mathcal{E}(\mathbf{r}, \mathbf{P}_\star(\mathbf{r}_\varepsilon)).
\end{aligned} \tag{161}$$

Again divide by ε and let $\varepsilon \to 0$. Using continuity of $\mathbf{P}_\star(\mathbf{r})$ (which follows from "unique stable minimizer" under standard regularity assumptions), $\mathbf{P}_\star(\mathbf{r}_\varepsilon) \to \mathbf{P}_\star(\mathbf{r})$ as $\varepsilon \to 0$, and differentiability of $\mathcal{E}$ in $\mathbf{r}$ yields

$$\begin{aligned}
\liminf_{\varepsilon \to 0} \frac{\mathbf{E}^\star(\mathbf{r}_\varepsilon) - \mathbf{E}^\star(\mathbf{r})}{\varepsilon} &\ge \lim_{\varepsilon \to 0} \frac{\mathcal{E}(\mathbf{r}_\varepsilon, \mathbf{P}_\star(\mathbf{r}_\varepsilon)) - \mathcal{E}(\mathbf{r}, \mathbf{P}_\star(\mathbf{r}_\varepsilon))}{\varepsilon} \\
&= \lim_{\varepsilon \to 0} \Big(\nabla_{\mathbf{r}}\mathcal{E}(\mathbf{r}, \mathbf{P}_\star(\mathbf{r}_\varepsilon)) \cdot \delta\mathbf{r}\Big) \\
&= \nabla_{\mathbf{r}}\mathcal{E}(\mathbf{r}, \mathbf{P}_\star(\mathbf{r})) \cdot \delta\mathbf{r}.
\end{aligned} \tag{162}$$

Combining (160) and (162) shows that the directional derivative exists and equals

$$D_{\mathbf{r}}\mathbf{E}^\star(\mathbf{r})[\delta\mathbf{r}] = \nabla_{\mathbf{r}}\mathcal{E}(\mathbf{r}, \mathbf{P}_\star(\mathbf{r})) \cdot \delta\mathbf{r} \quad \forall\delta\mathbf{r},$$

hence

$$\nabla_{\mathbf{r}}\mathbf{E}^\star(\mathbf{r}) = \nabla_{\mathbf{r}}\mathcal{E}(\mathbf{r}, \mathbf{P}_\star(\mathbf{r})). \tag{163}$$

Following by the definition of $\mathbf{R}_I(t)$, we have:

$$\begin{aligned}\mathbf{R}_I(t) &= -\nabla_{\mathbf{r}_I}\mathcal{E}(\mathbf{r}_t, \mathbf{P}_t) + \nabla_{\mathbf{r}_I}\mathbf{E}^\star(\mathbf{r}_t) \\ &= -\nabla_{\mathbf{r}_I}\mathcal{E}(\mathbf{r}_t, \mathbf{P}_t) + \nabla_{\mathbf{r}_I}\mathcal{E}(\mathbf{r}_t, \mathbf{P}_\star(\mathbf{r}_t)) \\ &= \nabla_{\mathbf{r}_I}\mathcal{E}(\mathbf{r}_t, \mathbf{P}_\star(\mathbf{r}_t)) - \nabla_{\mathbf{r}_I}\mathcal{E}(\mathbf{r}_t, \mathbf{P}_t).\end{aligned} \tag{164}$$

To bound the residual $\mathbf{R}_I(t)$ in terms of the projector tracking error, we now show that, for each nuclear geometry $\mathbf{r}$ and each nucleus index I, the following map is locally Lipschitz in $\mathbf{P}$ with respect to $\|\cdot\|_F$ on a neighborhood of the Born–Oppenheimer minimizer $\mathbf{P}_\star(\mathbf{r})$:

$$\mathcal{F}_I(\mathbf{P}) \coloneqq \nabla_{\mathbf{r}_I}\mathcal{E}(\mathbf{r}, \mathbf{P}) \in \mathbb{R}^3$$

Since $\rho_{\mathbf{P}}$ depends linearly on $\mathbf{P}$, the Hartree term is a quadratic polynomial in the entries of $\mathbf{P}$ with coefficients given by two-electron integrals, and the first term is linear in $\mathbf{P}$. For standard smooth exchange–correlation models, the functional $E_{\mathrm{xc}}[\rho]$ is C^2 on the set of densities encountered along the trajectory, hence the composition $\mathbf{P} \mapsto E_{\mathrm{xc}}[\rho_{\mathbf{P}}]$ is C^2 as a function on the finite-dimensional matrix space. Consequently, for each $\mathbf{r}$ the partial derivative $\mathbf{P} \mapsto \nabla_{\mathbf{r}_I}\mathcal{E}(\mathbf{r}, \mathbf{P})$ is C^1 on some open neighborhood $\mathcal{U}$ of $\mathbf{P}_\star(\mathbf{r})$ in the ambient space of matrices.

Now, we take any admissible $\mathbf{P}, \mathbf{Q} \in \mathcal{U}$ and set the line segment

$$\mathbf{P}_s \coloneqq \mathbf{Q} + s(\mathbf{P} - \mathbf{Q}), \quad s \in [0, 1].$$

By the fundamental theorem of calculus applied component-wise to the C^1 map $\mathcal{F}_I : \mathcal{U} \to \mathbb{R}^3$, we have

$$\mathcal{F}_I(\mathbf{P}) - \mathcal{F}_I(\mathbf{Q}) = \int_0^1 \frac{d}{ds}\mathcal{F}_I(\mathbf{P}_s)ds = \int_0^1 D_{\mathbf{P}}\mathcal{F}_I(\mathbf{P}_s)[\mathbf{P} - \mathbf{Q}]ds, \tag{165}$$

where $D_{\mathbf{P}}\mathcal{F}_I(\mathbf{P}_s)$ denotes the Fréchet derivative with respect to $\mathbf{P}$. Write $\mathcal{F}_I = (\mathcal{F}_{I,1}, \mathcal{F}_{I,2}, \mathcal{F}_{I,3})$ for the Cartesian components. For each $\alpha \in \{1, 2, 3\}$, the scalar map $\mathcal{F}_{I,\alpha}(\mathbf{P}) = \partial_{r_{I,\alpha}}\mathcal{E}(\mathbf{r}, \mathbf{P})$ is C^1 in $\mathbf{P}$, so its derivative is a bounded linear functional. With the Frobenius pairing $\langle A, B\rangle \coloneqq \mathrm{Re}\,\mathrm{Tr}(A^\dagger B)$ on matrices, Riesz representation in finite dimension gives an unique matrix $\mathbf{G}_{I,\alpha}(\mathbf{r}, \mathbf{P}_s) \in \mathbb{C}^{M\times M}$ such that for all $\Delta\mathbf{P}$,

$$D_{\mathbf{P}}\mathcal{F}_{I,\alpha}(\mathbf{P}_s)[\Delta\mathbf{P}] = \big\langle \mathbf{G}_{I,\alpha}(\mathbf{r}, \mathbf{P}_s), \Delta\mathbf{P}\big\rangle = \mathrm{Re}\,\mathrm{Tr}\left(\mathbf{G}_{I,\alpha}(\mathbf{r}, \mathbf{P}_s)^\dagger \Delta\mathbf{P}\right). \tag{166}$$

Applying Cauchy–Schwarz for the Frobenius inner product yields, for each α,

$$\big|D_{\mathbf{P}}\mathcal{F}_{I,\alpha}(\mathbf{P}_s)[\mathbf{P} - \mathbf{Q}]\big| \le \big\|\mathbf{G}_{I,\alpha}(\mathbf{r}, \mathbf{P}_s)\big\|_F \big\|\mathbf{P} - \mathbf{Q}\big\|_F. \tag{167}$$

Therefore, taking the Euclidean norm in $\mathbb{R}^3$ and using $\|\mathbf{v}\|^2 = \sum_\alpha |v_\alpha|^2$,

$$\begin{aligned}\big\|D_{\mathbf{P}}\mathcal{F}_I(\mathbf{P}_s)[\mathbf{P} - \mathbf{Q}]\big\| &= \left(\sum_{\alpha=1}^3 \big|D_{\mathbf{P}}\mathcal{F}_{I,\alpha}(\mathbf{P}_s)[\mathbf{P} - \mathbf{Q}]\big|^2\right)^{1/2} \\ &\le \left(\sum_{\alpha=1}^3 \big\|\mathbf{G}_{I,\alpha}(\mathbf{r}, \mathbf{P}_s)\big\|_F^2\right)^{1/2} \big\|\mathbf{P} - \mathbf{Q}\big\|_F.\end{aligned} \tag{168}$$

Combining (165) and (168) gives

$$\big\|\mathcal{F}_I(\mathbf{P}) - \mathcal{F}_I(\mathbf{Q})\big\| \le \int_0^1 \big\|\mathbf{G}_I(\mathbf{r}, \mathbf{P}_s)\big\|_{F,2} ds \big\|\mathbf{P} - \mathbf{Q}\big\|_F. \tag{169}$$

It remains to show that $\big\|\mathbf{G}_I(\mathbf{r}, \mathbf{P}_s)\big\|_{F,2}$ is bounded on the relevant set. Because $\mathbf{P} \mapsto \nabla_{\mathbf{r}_I}\mathcal{E}(\mathbf{r}, \mathbf{P})$ is C^1 on $\mathcal{U}$, the map $\mathbf{P} \mapsto \mathbf{G}_I(\mathbf{r}, \mathbf{P})$ is continuous.

We choose a closed ball $\overline{\mathcal{U}}_\delta \coloneqq \{\mathbf{P} : \|\mathbf{P} - \mathbf{P}_\star(\mathbf{r})\|_F \le \delta\} \subset \mathcal{U}$ to be compact in finite dimension, hence the continuous function $\mathbf{P} \mapsto \|\mathbf{G}_I(\mathbf{r}, \mathbf{P})\|_{F,2}$ attains its maximum on $\overline{\mathcal{U}}_\delta$. Our goal is then to show the explicit formulation of Lipschitz constant in the following:

$$L_I(\mathbf{r}) \coloneqq \max_{\mathbf{P} \in \overline{\mathcal{U}}_\delta} \big\|\mathbf{G}_I(\mathbf{r}, \mathbf{P})\big\|_{F,2} < \infty.$$

Since $\mathbf{P}_s \in \overline{\mathcal{U}}_\delta$ for all $s \in [0,1]$ whenever $\mathbf{P}, \mathbf{Q} \in \overline{\mathcal{U}}_\delta$, (169) yields the desired Lipschitz estimate

$$\big\|\nabla_{\mathbf{r}_I} \mathcal{E}(\mathbf{r}, \mathbf{P}) - \nabla_{\mathbf{r}_I} \mathcal{E}(\mathbf{r}, \mathbf{Q})\big\| \le L_I(\mathbf{r}) \|\mathbf{P} - \mathbf{Q}\|_F. \tag{170}$$

If, moreover, the nuclear path $\mathbf{r}_t$ remains in a compact set Ω and the mixed derivative is continuous in $(\mathbf{r}, \mathbf{P})$, one may take $L_I \coloneqq \sup_{\mathbf{r} \in \Omega} L_I(\mathbf{r}) < \infty$ to obtain a uniform constant along the trajectory. Then applying (170) to (164) yields

$$\begin{aligned} \|\mathbf{R}_I(t)\| &= \big\|\nabla_{\mathbf{r}_I} \mathcal{E}(\mathbf{r}_t, \mathbf{P}_\star(\mathbf{r}_t)) - \nabla_{\mathbf{r}_I} \mathcal{E}(\mathbf{r}_t, \mathbf{P}_t)\big\| \\ &\le L_I \|\mathbf{P}_\star(\mathbf{r}_t) - \mathbf{P}_t\|_F. \end{aligned} \tag{171}$$

Fix a nuclear geometry $\mathbf{r}$ and a nucleus index I. Let the ambient matrix space be

$$\mathbb{V} \coloneqq \mathbb{C}^{M \times M}, \quad \langle \mathbf{A}, \mathbf{B} \rangle_F \coloneqq \operatorname{Re} \operatorname{Tr}(\mathbf{A}^\dagger \mathbf{B}), \quad \|\mathbf{A}\|_F^2 \coloneqq \langle \mathbf{A}, \mathbf{A} \rangle_F.$$

For $\alpha \in \{1,2,3\}$ define the scalar component

$$\mathcal{F}_{I,\alpha}(\mathbf{P}) \coloneqq \partial_{r_{I,\alpha}} \mathcal{E}(\mathbf{r}, \mathbf{P}), \quad \mathcal{F}_I(\mathbf{P}) = (\mathcal{F}_{I,1}, \mathcal{F}_{I,2}, \mathcal{F}_{I,3}) \in \mathbb{R}^3.$$

By the standing regularity assumption in the proof, the map $\mathbf{P} \mapsto \mathcal{F}_I(\mathbf{P})$ is C^1 on an open neighborhood $\mathcal{U}$ of $\mathbf{P}_\star(\mathbf{r})$. Equivalently, for each α the Fréchet derivative $D_{\mathbf{P}} \mathcal{F}_{I,\alpha}(\mathbf{P}) : \mathbb{V} \to \mathbb{R}$ exists for all $\mathbf{P} \in \mathcal{U}$ and depends continuously on $\mathbf{P}$ in operator norm, i.e.

$$\mathbf{P}_n \to \mathbf{P} \quad \Longrightarrow \quad \big\|D_{\mathbf{P}} \mathcal{F}_{I,\alpha}(\mathbf{P}_n) - D_{\mathbf{P}} \mathcal{F}_{I,\alpha}(\mathbf{P})\big\|_{\mathrm{op}} \to 0, \quad \|\ell\|_{\mathrm{op}} \coloneqq \sup_{\|\Delta \mathbf{P}\|_F = 1} |\ell[\Delta \mathbf{P}]|.$$

For the Frobenius Riesz map $\mathcal{R} : \mathbb{V} \longrightarrow \mathbb{V}^*$, one can show that

$$\mathcal{R}(\mathbf{A})[\Delta \mathbf{P}] \coloneqq \langle \mathbf{A}, \Delta \mathbf{P} \rangle_F = \operatorname{Re} \operatorname{Tr}(\mathbf{A}^\dagger \Delta \mathbf{P}). \tag{172}$$

Then $\mathcal{R}$ is a linear bijection of finite-dimensional Riesz representation, and it is an isometry:

$$\begin{aligned} \|\mathcal{R}(\mathbf{A})\|_{\mathrm{op}} &= \sup_{\|\Delta \mathbf{P}\|_F = 1} |\langle \mathbf{A}, \Delta \mathbf{P} \rangle_F| \le \sup_{\|\Delta \mathbf{P}\|_F = 1} \|\mathbf{A}\|_F \|\Delta \mathbf{P}\|_F = \|\mathbf{A}\|_F, \\ \|\mathcal{R}(\mathbf{A})\|_{\mathrm{op}} &\ge |\langle \mathbf{A}, \mathbf{A}/\|\mathbf{A}\|_F \rangle_F| = \|\mathbf{A}\|_F. \end{aligned}$$

Hence, we have

$$\|\mathcal{R}(\mathbf{A})\|_{\mathrm{op}} = \|\mathbf{A}\|_F, \quad \|\mathcal{R}^{-1}(\ell)\|_F = \|\ell\|_{\mathrm{op}}.$$

Define $\mathbf{G}_{I,\alpha}(\mathbf{r}, \mathbf{P}) \in \mathbb{V}$ by pulling back the derivative through $\mathcal{R}^{-1}$:

$$\begin{aligned} \mathbf{G}_{I,\alpha}(\mathbf{r}, \mathbf{P}) &\coloneqq \mathcal{R}^{-1}\big(D_{\mathbf{P}} \mathcal{F}_{I,\alpha}(\mathbf{P})\big), \\ D_{\mathbf{P}} \mathcal{F}_{I,\alpha}(\mathbf{P})[\Delta \mathbf{P}] &= \mathcal{R}\big(\mathbf{G}_{I,\alpha}(\mathbf{r}, \mathbf{P})\big)[\Delta \mathbf{P}] = \operatorname{Re} \operatorname{Tr}\Big(\mathbf{G}_{I,\alpha}(\mathbf{r}, \mathbf{P})^\dagger \Delta \mathbf{P}\Big), \quad \forall \Delta \mathbf{P} \in \mathbb{V}. \end{aligned}$$

Because $\mathbf{P} \mapsto D_{\mathbf{P}} \mathcal{F}_{I,\alpha}(\mathbf{P})$ is continuous and $\mathcal{R}^{-1}$ is continuous, the composition $\mathbf{P} \mapsto \mathbf{G}_{I,\alpha}(\mathbf{r}, \mathbf{P})$ is continuous on $\mathcal{U}$:

$$\begin{aligned} \mathbf{P}_n \to \mathbf{P} \Longrightarrow \big\|\mathbf{G}_{I,\alpha}(\mathbf{r}, \mathbf{P}_n) - \mathbf{G}_{I,\alpha}(\mathbf{r}, \mathbf{P})\big\|_F &= \Big\|\mathcal{R}^{-1}\Big(D_{\mathbf{P}} \mathcal{F}_{I,\alpha}(\mathbf{P}_n) - D_{\mathbf{P}} \mathcal{F}_{I,\alpha}(\mathbf{P})\Big)\Big\|_F \\ &= \Big\|D_{\mathbf{P}} \mathcal{F}_{I,\alpha}(\mathbf{P}_n) - D_{\mathbf{P}} \mathcal{F}_{I,\alpha}(\mathbf{P})\Big\|_{\mathrm{op}} \to 0. \end{aligned}$$

Collect the three matrices into $\mathbf{G}_I(\mathbf{r}, \mathbf{P}) \coloneqq (\mathbf{G}_{I,1}, \mathbf{G}_{I,2}, \mathbf{G}_{I,3})$ and define

$$\big\|\mathbf{G}_I(\mathbf{r}, \mathbf{P})\big\|_{F,2} \coloneqq \left(\sum_{\alpha=1}^{3} \big\|\mathbf{G}_{I,\alpha}(\mathbf{r}, \mathbf{P})\big\|_F^2\right)^{1/2}.$$

Since each $\mathbf{P} \mapsto \|\mathbf{G}_{I,\alpha}(\mathbf{r}, \mathbf{P})\|_F$ is continuous, the map

$$\varphi(\mathbf{P}) \coloneqq \big\|\mathbf{G}_I(\mathbf{r}, \mathbf{P})\big\|_{F,2}$$

is continuous on $\mathcal{U}$ as a composition of continuous operations:

$$\begin{aligned}\mathbf{P}_n \to \mathbf{P} &\iff \|\mathbf{G}_{I,\alpha}(\mathbf{r},\mathbf{P}_n)\|_F \to \|\mathbf{G}_{I,\alpha}(\mathbf{r},\mathbf{P})\|_F \\ &\iff \sum_{\alpha=1}^{3} \|\mathbf{G}_{I,\alpha}(\mathbf{r},\mathbf{P}_n)\|_F^2 \to \sum_{\alpha=1}^{3} \|\mathbf{G}_{I,\alpha}(\mathbf{r},\mathbf{P})\|_F^2, \\ &\iff \varphi(\mathbf{P}_n) = \left(\sum_{\alpha=1}^{3} \|\mathbf{G}_{I,\alpha}(\mathbf{r},\mathbf{P}_n)\|_F^2\right)^{1/2} \to \left(\sum_{\alpha=1}^{3} \|\mathbf{G}_{I,\alpha}(\mathbf{r},\mathbf{P})\|_F^2\right)^{1/2} = \varphi(\mathbf{P}).\end{aligned}$$

While $\mathcal{U}$ is open and $\mathbf{P}_\star(\mathbf{r}) \in \mathcal{U}$, there exists $\delta > 0$ such that the Frobenius ball is contained in $\mathcal{U}$:

$$\exists \delta > 0 \quad \text{s.t.} \quad \{\mathbf{P} : \ \|\mathbf{P} - \mathbf{P}_\star(\mathbf{r})\|_F < \delta\} \subset \mathcal{U},$$

and we set the closed ball

$$\overline{\mathcal{U}}_\delta \coloneqq \{\mathbf{P} \in \mathbb{V} : \ \|\mathbf{P} - \mathbf{P}_\star(\mathbf{r})\|_F \leq \delta\} \subset \mathcal{U}.$$

To see that $\overline{\mathcal{U}}_\delta$ is compact, identify $\mathbb{V}$ with $\mathbb{R}^{2M^2}$ via the real-linear isometry

$$\begin{aligned}\Psi : \mathbb{C}^{M\times M} &\longrightarrow \mathbb{R}^{2M^2}, \\ \mathbf{P} = (P_{\mu\nu})_{\mu,\nu=1}^{M} &\longmapsto \big(\operatorname{Re} P_{11}, \operatorname{Im} P_{11}, \ldots, \operatorname{Re} P_{MM}, \operatorname{Im} P_{MM}\big).\end{aligned}$$

Then the Frobenius norm becomes the Euclidean norm in $\mathbb{R}^{2M^2}$:

$$\|\mathbf{P}\|_F^2 = \sum_{\mu,\nu=1}^{M} |P_{\mu\nu}|^2 = \sum_{\mu,\nu=1}^{M} \Big((\operatorname{Re} P_{\mu\nu})^2 + (\operatorname{Im} P_{\mu\nu})^2\Big) = \|\Psi(\mathbf{P})\|_2^2,$$

hence

$$\Psi(\overline{\mathcal{U}}_\delta) = \Big\{x \in \mathbb{R}^{2M^2} : \ \|x - \Psi(\mathbf{P}_\star(\mathbf{r}))\|_2 \leq \delta\Big\}.$$

The right-hand side is a closed and bounded subset of $\mathbb{R}^{2M^2}$, and therefore compact by the Heine-Borel theorem. Since Ψ is a homeomorphism and a linear isometry, $\overline{\mathcal{U}}_\delta$ is compact in $\mathbb{V}$. Because φ is continuous on the compact set $\overline{\mathcal{U}}_\delta$, the extreme value theorem implies that φ attains a maximum:

$$\exists \widehat{\mathbf{P}} \in \overline{\mathcal{U}}_\delta \quad \text{s.t.} \quad \varphi(\widehat{\mathbf{P}}) = \max_{\mathbf{P}\in\overline{\mathcal{U}}_\delta} \varphi(\mathbf{P}).$$

For the defined Lipschitz constant

$$L_I(\mathbf{r}) \coloneqq \max_{\mathbf{P}\in\overline{\mathcal{U}}_\delta} \big\|\mathbf{G}_I(\mathbf{r},\mathbf{P})\big\|_{F,2} = \big\|\mathbf{G}_I(\mathbf{r},\widehat{\mathbf{P}})\big\|_{F,2} = \left(\sum_{\alpha=1}^{3} \big\|\mathbf{G}_{I,\alpha}(\mathbf{r},\widehat{\mathbf{P}})\big\|_F^2\right)^{1/2} < \infty$$

Therefore, $\|\mathbf{R}_I(t)\| \to 0$ follows once we show $\|\mathbf{P}_t - \mathbf{P}_\star(\mathbf{r}_t)\|_F \to 0$. Combining this result with Lemma 11.7, the proof is complete. □

Algorithm 1 Electronic update driven by gauge-invariant SCF residual

Require: Current electronic variables $(\mathbf{C}_n, \mathbf{V}_n)$, nuclear geometries $(\mathbf{r}_n, \mathbf{r}_{n+1})$, step size h, parameters (μ, γ, κ)
Require: Routines: $\mathrm{S}(\mathbf{r})$, $\mathbf{H}_{\mathbf{r}}[\mathbf{P}]$
Ensure: Updated electronic variables $(\mathbf{C}_{n+1}, \mathbf{V}_{n+1})$
function INVSQRT($\mathbf{G}$)
  Compute eigendecomposition $\mathbf{G} = \mathbf{Q}\Lambda\mathbf{Q}^\dagger$ with diagonal $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$
  Set $\Lambda^{-1/2} = \mathrm{diag}(\lambda_1^{-1/2}, \ldots, \lambda_N^{-1/2})$
  return $\mathbf{Q}\Lambda^{-1/2}\mathbf{Q}^\dagger$
end function
function PROJECT($\mathbf{C}, \mathbf{S}, \mathbf{X}$)
  return $\mathbf{X} - \mathbf{C}(\mathbf{C}^\dagger\mathbf{S}\mathbf{X})$
end function
function RETRACT($\widetilde{\mathbf{C}}, \mathbf{S}$)
  $\mathbf{G} \leftarrow \widetilde{\mathbf{C}}^\dagger\mathbf{S}\widetilde{\mathbf{C}}$
  return $\widetilde{\mathbf{C}}$INVSQRT($\mathbf{G}$)
end function
function POLARUNITARY($\mathbf{M}$)
  Compute SVD $\mathbf{M} = \mathbf{U}_M\Sigma\mathbf{V}_M^\dagger$
  return $\mathbf{U}_M\mathbf{V}_M^\dagger$
end function
$\mathbf{S}_n \leftarrow \mathrm{S}(\mathbf{r}_n)$
$\mathbf{C}_n \leftarrow$ RETRACT$(\mathbf{C}_n, \mathbf{S}_n)$
$\mathbf{V}_n \leftarrow$ PROJECT$(\mathbf{C}_n, \mathbf{S}_n, \mathbf{V}_n)$
$\mathbf{P}_n \leftarrow \mathbf{C}_n\mathbf{C}_n^\dagger\mathbf{S}_n$
$\mathbf{H}_n \leftarrow \mathbf{H}_{\mathbf{r}_n}[\mathbf{P}_n]$
Solve the linear system $\mathbf{S}_n\mathbf{X}_n = \mathbf{H}_n\mathbf{C}_n$
$\mathcal{R}_n \leftarrow$ PROJECT$(\mathbf{C}_n, \mathbf{S}_n, \mathbf{X}_n)$
$\alpha \leftarrow h\gamma/\mu$
$\mathbf{V}_{n+1}^* \leftarrow (\mathbf{V}_n - (h\kappa/\mu)\mathcal{R}_n)/(1+\alpha)$
$\mathbf{V}_{n+1}^* \leftarrow$ PROJECT$(\mathbf{C}_n, \mathbf{S}_n, \mathbf{V}_{n+1}^*)$
$\widetilde{\mathbf{C}}_{n+1} \leftarrow \mathbf{C}_n + h\mathbf{V}_{n+1}^*$
$\mathbf{S}_{n+1} \leftarrow \mathrm{S}(\mathbf{r}_{n+1})$
$\mathbf{C}_{n+1} \leftarrow$ RETRACT$(\widetilde{\mathbf{C}}_{n+1}, \mathbf{S}_{n+1})$
$\mathbf{M} \leftarrow \mathbf{C}_n^\dagger\mathbf{S}_{n+1}\mathbf{C}_{n+1}$
$\mathbf{U} \leftarrow$ POLARUNITARY$(\mathbf{M})$
$\mathbf{C}_{n+1} \leftarrow \mathbf{C}_{n+1}\mathbf{U}^\dagger$
$\mathbf{V}_{n+1}^* \leftarrow \mathbf{V}_{n+1}^*\mathbf{U}^\dagger$
$\mathbf{V}_{n+1} \leftarrow$ PROJECT$(\mathbf{C}_{n+1}, \mathbf{S}_{n+1}, \mathbf{V}_{n+1}^*)$
return $(\mathbf{C}_{n+1}, \mathbf{V}_{n+1})$

12 Algorithm

Algorithm 1 implements a discrete realization of the gauge-symmetric dual dynamics. At each step, the overlap matrix $\mathbf{S}_n = \mathrm{S}(\mathbf{r}_n)$ and projector $\mathbf{P}_n = \mathbf{C}_n\mathbf{C}_n^\dagger\mathbf{S}_n$ are constructed, and $\mathbf{H}_n = \mathbf{H}_{\mathbf{r}_n}[\mathbf{P}_n]$ is evaluated. The residual is obtained by solving $\mathbf{S}_n\mathbf{X}_n = \mathbf{H}_n\mathbf{C}_n$ followed by projection onto $\mathrm{Hor}_{\mathbf{C}_n}$. The auxiliary variable $\mathbf{V}_n$, which discretizes the covariant momentum, is updated via a damped residual correction with parameters (μ, γ, κ), and the orbitals are advanced as $\widetilde{\mathbf{C}}_{n+1} = \mathbf{C}_n + h\mathbf{V}_{n+1}^*$. Retraction onto $\mathrm{St}_{\mathbf{S}_{n+1}}$ enforces orthonormality, while polar unitary alignment removes gauge rotation. The scheme eliminates inner SCF iterations and consists solely of dense linear-algebra operations and relies solely on dense matrix operations.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The main claims stated in the abstract and introduction are aligned with the theoretical formulation and experimental results presented in the paper.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [N/A]

   Justification: The paper discusses the practical scope, assumptions, and constraints of the proposed approach in the manuscript and appendix.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: The theoretical results state the required assumptions and are accompanied by derivations and proofs in the paper and appendix.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: The paper provides the information needed to reproduce the main empirical results, including the formulation, model design, and experimental protocol.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [N/A]

Justification: We provide detailed methodological and experimental descriptions in the submission, and we plan to release the code upon acceptance.

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

Answer: [Yes]

Justification: The manuscript and appendix describe the training, evaluation, and implementation settings used for the experiments.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [No]

Justification: The paper reports point estimates of trajectory-level metrics and runtime under a fixed evaluation protocol, but it does not report error bars, confidence intervals, or statistical significance tests for the main experimental claims.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: The paper provides the relevant runtime and implementation details needed to understand the computational setup of the experiments.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We have reviewed the NeurIPS Code of Ethics and believe that this research conforms to it.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: This work is a foundational methodological study on scientific simulation, and we do not identify a direct broader societal impact requiring a dedicated discussion.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The paper does not release high-risk models, datasets, or other assets that would require misuse-specific safeguards.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: Existing datasets, software tools, and prior methods used in this work are credited in the manuscript.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.

- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: This submission does not release a new public asset alongside the anonymized paper submission.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: This work does not involve crowdsourcing or research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: This work does not involve crowdsourcing or research with human subjects, so IRB review is not applicable.

    Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: The core method development in this research does not involve LLMs as an important, original, or non-standard component.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.