Adjoint Matching through the Lens of the Stochastic Maximum Principle in Optimal Control

Anonymous authors
Paper under double-blind review

Abstract

Reward fine-tuning of diffusion and flow models and sampling from tilted or Boltzmann distributions can both be formulated as stochastic optimal control (SOC) problems, where learning an optimal generative dynamics corresponds to optimizing a control under SDE constraints. In this work, we revisit and generalize *Adjoint Matching*, a recently proposed SOC-based method for learning optimal controls, and place it on a rigorous footing by deriving it from the *Stochastic Maximum Principle* (SMP). We formulate a general Hamiltonian adjoint matching objective for SOC problems with control-dependent drift and diffusion and convex running costs, and show that its expected value has the same first variation as the original SOC objective. As a consequence, critical points satisfy the Hamilton–Jacobi–Bellman (HJB) stationarity conditions. In the important practical case of state- and control-independent diffusion, we recover the *lean* adjoint matching loss previously introduced in adjoint matching, which avoids second-order terms and whose critical points coincide with the optimal control under mild uniqueness assumptions. Finally, we show that adjoint matching can be precisely interpreted as a continuous-time method of successive approximations induced by the SMP, yielding a practical and implementable alternative to classical SMP-based algorithms, which are obstructed by intractable martingale terms in the stochastic setting. These results are also of independent interest to the stochastic control community, providing new implementable objectives and a viable pathway for SMP-based iterations in stochastic problems.

1 Introduction

Flow Matching (Lipman et al., 2023; Albergo & Vanden-Eijnden, 2023; Liu et al., 2023) and denoising diffusion models (Song & Ermon, 2019; Ho et al., 2020; Song et al., 2021; Kingma et al., 2021) are the state-of-the-art for many generative modeling applications, including text-to-image (Rombach et al., 2022; Esser et al., 2024), text-to-video (Singer et al., 2022), text-to-audio (Le et al., 2024; Vyas et al., 2023), and scientific tasks like protein structure prediction (Abramson et al., 2024). The generation process involves solving ordinary or stochastic differential equations (ODEs/SDEs) with vector fields pretrained on large-scale datasets. Oftentimes, the pretrained base model does not achieve the desired sample quality. If we have access to a reward model $r(x)$ that assesses the sample quality, a common remedy is to post-train, or *fine-tune*, the base model in order to sample from the *tilted distribution* with density:

$$p^*(x) \propto p_{\text{base}}(x) \exp(r(x)), \tag{1}$$

where p_{base} denotes the density of the distribution induced by the pretrained base model.

Flow matching and diffusion models are typically pre-trained using samples from the distribution that we want to model (images, audio, video, etc.). A related problem that is central in the computational sciences is sampling from the Boltzmann distribution, defined by an unnormalized energy function E, i.e.

$$p^*(x) \propto \exp(-\beta E(x)), \tag{2}$$

where $\beta > 0$ denotes the inverse temperature. This fundamental task has wide applications in molecular modeling and Bayesian inference. Markov Chain Monte Carlo (MCMC, Neal (2001; 2011)) and Sequential Monte Carlo (SMC, Del Moral et al. (2006)) are classical approaches that have been studied extensively, but often suffer from slow mixing and poor scalability to high-dimensional settings. Albergo et al. (2019); Arbel et al. (2021); Gabrié et al. (2022) augment Monte Carlo methods using normalizing flows (Rezende & Mohamed, 2015; Chen et al., 2018), and building on these ideas, a more recent line of work Tzen & Raginsky (2019); Zhang & Chen (2022); Berner et al. (2023); Bruna & Han (2024); Richter & Berner (2024); Havens et al. (2025); Ren et al. (2026) consider generation algorithms based on continuous normalizing flows, i.e. samples are generated by solving ODEs or SDEs with learned vector fields.

As we review in Section 2, both reward fine-tuning and Boltzmann sampling can be cast as stochastic optimal control (SOC) problems. In the high-dimensional settings typical of these applications, classical grid-based methods for SOC are intractable, and the control is usually parameterized by a neural network and optimized via stochastic gradient methods. Domingo-Enrich et al. (2025) proposed *adjoint matching*, an SOC method that frames learning the optimal control as an L^2 regression problem in which the target is computed by solving an adjoint ODE backward in time. This approach was applied successfully to reward fine-tuning of a text-to-image flow matching model, yielding empirical improvements over existing methods at the time.

Despite its empirical success, adjoint matching was originally proposed heuristically within the control-affine quadratic-cost setting, and its optimality was verified a posteriori rather than derived from first principles (see Section 2.3). This heuristic construction does not reveal the underlying control-theoretic structure that explains *why* the simplification preserves optimality, nor does it extend to more general SOC formulations.

In this paper, we revisit and generalize adjoint matching, and establish its connection to stochastic optimal control through the *stochastic maximum principle* (SMP). Beyond the generative modeling context, these results are also of independent interest to the stochastic control community, as they yield new implementable algorithms for SOC and a practical resolution to a long-standing obstacle in SMP-based methods. Our main contributions are as follows:

- **A general Hamiltonian BAM formulation for SOC (Theorem 1).** We formulate a *general Hamiltonian basic adjoint matching* (BAM) objective for SOC problems with general control-dependent drift and diffusion and convex running costs. This formulation relies on unbiased estimators of the cost-to-go gradient and Hessian via first- and second-order adjoint equations, and substantially extends the setting of Domingo-Enrich et al. (2025), which arises as a special case. We show that the expected BAM objective has the same first variation as the original SOC objective, and that its critical points satisfy the SMP/HJB stationarity conditions.
- **A lean adjoint matching loss for time-dependent diffusion (Theorem 2).** In the practically important case where the diffusion depends only on time, the BAM framework reduces to a *lean adjoint matching* (AM) loss that avoids second-order terms and recovers the original adjoint matching objective of Domingo-Enrich et al. (2025) in the control-affine quadratic-cost setting as a special case. Under standard uniqueness assumptions, the critical points of this lean AM loss recover the optimal control.
- **A precise SMP interpretation of adjoint matching (Theorem 4).** The preceding results characterize the critical points of the lean AM loss but do not address whether gradient-based optimization of that loss implements a principled improvement procedure. We further show that the gradient flow induced by the expected lean AM loss coincides with a continuous-time method of successive approximations (MSA) derived from the stochastic maximum principle (SMP) for optimal stochastic control, while bypassing the intractable martingale term that obstructs classical SMP-based schemes. This gradient-flow viewpoint is particularly relevant for parameterized controls (e.g. neural networks), where classical function-space optimality conditions do not directly translate into implementable parameter updates; the MSA interpretation certifies that each gradient step is a valid control-improvement step.

Roadmap. Section 2 reviews the SOC formulation and states the original adjoint matching objective. Section 3 introduces the general adjoint matching objective from a Hamiltonian perspective (Theorem 1),

then specializes to purely time-dependent diffusion to obtain a lean first-order objective whose critical points recover the optimal control (Theorem 2). Section 4 reviews the stochastic maximum principle and examines the structural differences that obstruct direct SMP-based algorithms. Section 5 shows how the lean adjoint matching of this paper is precisely the continuous-time MSA induced by the SMP (Theorem 4). Proofs and technical extensions are deferred to the appendix.

2 Background

We begin by reviewing the stochastic optimal control (SOC) formulation that underlies both reward fine-tuning and Boltzmann sampling, and then recall the adjoint matching objective proposed by Domingo-Enrich et al. (2025).

2.1 Stochastic Optimal Control Formulation

We start with the following generic SOC problem. Let $f : \mathbb{R}^d \times \mathbb{R}^k \times [0,T] \to \mathbb{R} \cup \{+\infty\}$ be proper, lower semicontinuous, and convex in its second argument, let $b : \mathbb{R}^d \times \mathbb{R}^k \times [0,T] \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \times \mathbb{R}^k \times [0,T] \to \mathbb{R}^{d\times m}$ be C^1 in x, and B_t denote m-dimensional Brownian motion. We aim to solve:

$$\min_{u\in\mathcal{U}} \ \mathbb{E}\Big[\int_0^T f(X_t^u, u(X_t^u,t),t)\,\mathrm{d}t + g(X_T^u)\Big], \tag{3}$$

$$\text{s.t.} \quad \mathrm{d}X_t^u = b(X_t^u, u(X_t^u,t),t)\,\mathrm{d}t + \sigma(X_t^u, u(X_t^u,t),t)\,\mathrm{d}B_t, \qquad X_0^u \sim P_0. \tag{4}$$

Here P_0 is the initial distribution, f is the running cost, g is the terminal cost, b is the drift, σ is the diffusion coefficient, and $\mathcal{U}$ is the space of Markov control functions $u : \mathbb{R}^d \times [0,T] \to \mathbb{R}^k$.

2.2 Equivalence to Reward Fine-Tuning and Sampling

Reward fine-tuning of diffusion and flow models (problem 1) and sampling with continuous normalizing flows (problem 2) are closely connected: if we let p_{base} be a Gaussian, given an energy E we can construct a reward function $r(x) = -\beta E(x) - \log p_{\text{base}}(x)$ such that problem 2 reduces to problem 1. Moreover, both problems are equivalent to SOC problems. Generically, the control-affine quadratic-cost SOC problem reads

$$\min_{u\in\mathcal{U}} \mathbb{E}\big[\int_0^T \tfrac{1}{2}\|u(X_t^u,t)\|^2 + f(X_t^u,t)\,\mathrm{d}t + g(X_T^u)\big], \tag{5}$$

$$\text{s.t.} \ \ \mathrm{d}X_t^u = (b(X_t^u,t) + \sigma(X_t^u,t)u(X_t^u,t))\,\mathrm{d}t + \sigma(X_t^u,t)\mathrm{d}B_t, \quad X_0^u \sim P_0. \tag{6}$$

Here the symbols are as in equation (3), with b now denoting the base (uncontrolled) drift and f the state-dependent running cost (as opposed to the joint running cost $f(x,u,t)$ above). According to the path integral formulation of SOC (Kappen (2005), see (Domingo-Enrich et al., 2024, Eq. 8, App. B) for a self-contained proof), the Radon-Nikodym derivative between the probability measures of the optimally controlled process $\mathbb{P}^\star$ and the uncontrolled process $\mathbb{P}^{\text{base}}$ ($u \equiv 0$) is equal to

$$\tfrac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{P}^{\text{base}}}(X) = \exp\big(-\int_0^T f(X_t,t)\,\mathrm{d}t - g(X_T) + V(X_0,0)\big), \tag{7}$$

with X denoting the uncontrolled process and V denoting the *value function* or *cost-to-go*, defined by

$$V(x,t) = -\log \mathbb{E}[\exp(-\int_t^T f(X_s,s)\mathrm{d}s - g(X_T)) \mid X_t = x].$$

Domingo-Enrich et al. (2025, Sec. 4.2, 4.3) show that when $f \equiv 0$ and the base process satisfies the *memorylessness* condition[1], equation (7) implies that

$$p^*(x) \propto p^{\text{base}}(x)\exp(-g(x)), \tag{8}$$

[1]The uncontrolled process X from $\mathrm{d}X_t = b(X_t,t)\mathrm{d}t + \sigma(X_t,t)\mathrm{d}B_t$ is memoryless when the random variables X_0 and X_T are independent. See (Domingo-Enrich et al., 2025, Sec. 4.3) for sufficient conditions arising from generative modeling that guarantee this.

where p^* and p^{base} are the marginals of $\mathbb{P}^\star$ and $\mathbb{P}^{\text{base}}$ at terminal time $t = T$, respectively. A comparison between equation (8) and equations (1)(2) shows that we can recast reward fine-tuning and sampling into the SOC problem equation (5) by setting (i) $f \equiv 0$, (ii) the distribution P_0 and diffusion coefficient σ such that X is memoryless, and (iii) the terminal cost as $g(x) = -r(x)$, $g(x) = \beta E(x) + \log p_{\text{base}}(x)$, respectively.

The SOC viewpoint is one of several frameworks for reward fine-tuning of diffusion models. An alternative line of work treats the generation process as a discrete-time Markov decision process and applies reinforcement learning techniques such as policy gradient algorithms (Fan et al., 2023; Black et al., 2024). Other methods directly backpropagate the reward gradient through the sampling chain (Clark et al., 2024; Prabhudesai et al., 2023). By contrast, adjoint matching and the methods studied in this paper operate in the continuous-time SOC framework equation (5) (Uehara et al., 2024; Tang & Zhou, 2024; Han et al., 2025; Gao et al., 2026), leveraging adjoint equations to compute the optimal control without backpropagating through the full sampling procedure.

2.3 The Original Adjoint Matching Objective

Domingo-Enrich et al. (2025) proposed *adjoint matching* for the control-affine quadratic-cost SOC problem equation (5) with purely time-dependent diffusion $\sigma(x, u, t) = \sigma(t)$. In this setting, they define the *lean adjoint* $\tilde{a}(t; X) \in \mathbb{R}^d$ as the solution of the backward ODE

$$\frac{\mathrm{d}\tilde{a}}{\mathrm{d}t}(t; X) = -\nabla_x b(X_t, t)^\top \tilde{a}(t; X) - \nabla_x f(X_t, t), \tag{9}$$

$$\tilde{a}(T; X) = \nabla_x g(X_T), \tag{10}$$

and the adjoint matching loss is

$$\mathcal{L}_{\text{AM}}(u; X^{\bar{u}}) = \int_0^T \tfrac{1}{2} \|u(X_t^{\bar{u}}, t) + \sigma(t)^\top \tilde{a}(t; X^{\bar{u}})\|^2 \, \mathrm{d}t, \qquad \bar{u} = \text{stopgrad}(u). \tag{11}$$

In words, $\mathcal{L}_{\text{AM}}$ is an L^2 regression loss that trains the control u to match the target $-\sigma(t)^\top \tilde{a}(t; X^{\bar{u}})$ along trajectories simulated under the current (stop-gradiented) control $\bar{u}$.

This objective was motivated by adjoint sensitivity analysis but derived heuristically: the lean adjoint $\tilde{a}$ is a simplified version of the full adjoint process, obtained by omitting terms whose expectation vanishes at the optimal control u^*. Domingo-Enrich et al. (2025) proved that u^* is the unique critical point of the resulting loss, but the derivation relies on verifying this property a posteriori rather than constructing the objective from first principles. In the following sections, we provide such a construction via the Hamiltonian formulation of SOC and generalize the objective to settings with arbitrary control-dependent drift and running costs (Section 3). We further show that the resulting lean adjoint matching admits an interpretation as a continuous-time MSA induced by the SMP (Section 5).

3 The General Adjoint Matching Loss Functions

This section derives the *adjoint matching* objectives studied in this paper from a Hamiltonian perspective. The starting point is the Hamiltonian formulation of stochastic optimal control: at an optimal control, the HJB stationarity conditions reduce to a pointwise minimization of the Hamiltonian, which involves the cost-to-go derivatives $\nabla_x J$ and $\nabla_x^2 J$. Adjoint matching turns this stationarity condition into a learning objective by replacing $\nabla_x J$ and $\nabla_x^2 J$ with pathwise adjoint estimators computed along simulated trajectories. Table 1 summarizes the Hamiltonians and adjoint processes used throughout.

We first treat the general SOC problem equation (3) where both drift and diffusion may depend on the control. This yields a *basic adjoint matching* (BAM) loss involving first- and second-order adjoint equations. We then specialize to the practically important setting $\sigma(x, u, t) = \sigma(t)$ (as in diffusion models), where the diffusion term does not affect the Hamiltonian minimizer; this leads to the *lean* adjoint matching (AM) loss that avoids second-order quantities, recovering the objective stated in Section 2.3 as a special case. Full regularity and well-posedness conditions are collected in Appendix A; whenever we identify a critical point with the optimal control, we additionally invoke the verification hypotheses in Lemma 9.

Table 1: Key Hamiltonians and adjoint processes used in this paper.

Symbol	Description	Reference
HAMILTONIANS		
$H(x,t;u,p,M)$	Full HJB Hamiltonian (general σ)	Eq. (19)
$\tilde{H}(x,t;u,p)$	Simplified Hamiltonian (σ=$\sigma(t)$; no second-order term)	Eq. (27)
$\mathcal{H}(x,t;u,p,q)$	SMP Hamiltonian (includes martingale integrand q)	Eq. (41)
ADJOINT PROCESSES		
a_t	First-order adjoint SDE; estimates $\nabla_x J$	Eq. (15)
$\tilde{a}_t$	Lean adjoint ODE (σ=$\sigma(t)$); estimates $\nabla_x J$	Eq. (25)
A_t	Second-order adjoint SDE; estimates $\nabla_x^2 J$	Eq. (17)
$(p_t^*,\, q_t^*)$	SMP adjoint BSDE pair; q_t^* is unknown integrand	Eq. (44)

Our construction starts with the full adjoint processes that provide (unbiased) access to the first-order and second-order derivatives of the cost-to-go under a fixed Markov control policy with respect to the current state. Concretely, for a fixed control $\bar{u}$, define the cost-to-go

$$J(\bar{u};x,t) := \mathbb{E}\Big[\int_t^T f(X_s^{\bar{u}}, \bar{u}(X_s^{\bar{u}}, s), s)\,\mathrm{d}s + g(X_T^{\bar{u}}) \;\Big|\; X_t^{\bar{u}} = x\Big]. \tag{12}$$

The adjoint processes a_t and A_t defined below will provide pathwise estimates for the gradient and Hessian of $J(\bar{u};\cdot,\cdot)$ along the controlled dynamics (for a primer on first-order adjoint sensitivity for SDEs, see Leburu et al. (2026); our constructions go further by incorporating second-order adjoint processes and general control-dependent diffusion):

$$\mathbb{E}\big[a(t;X^{\bar{u}}) \mid X_t^{\bar{u}} = x\big] = \nabla_x J(\bar{u};x,t), \tag{13}$$

$$\mathbb{E}\big[A(t;X^{\bar{u}}) \mid X_t^{\bar{u}} = x\big] = \nabla_x^2 J(\bar{u};x,t). \tag{14}$$

To put it formally with a convenient notation, we denote by $\sigma_{\cdot k}(x,u,t) \in \mathbb{R}^d$ the k-th column of $\sigma(x,u,t)$ for $k \in \{1,\dots,m\}$, and set the total derivative

$$G_k^u(x,t) := \nabla_y \sigma_{\cdot k}(y, u(y,t), t)|_{y=x} \in \mathbb{R}^{d\times d}.$$

Let $a_t := a(t;X) \in \mathbb{R}^d$ solve the first-order adjoint SDE backwardly in time:

$$\begin{aligned} \mathrm{d}a_t = \Big(&-\nabla_x b(x,u(x,t),t)\big|_{x=X_t}^\top a_t - \nabla_x f(x,u(x,t),t)\big|_{x=X_t} \\ &+ \textstyle\sum_{k=1}^m G_k^u(X_t,t)^\top G_k^u(X_t,t)^\top a_t\Big)\,\mathrm{d}t - \textstyle\sum_{k=1}^m G_k^u(X_t,t)^\top a_t\,\mathrm{d}B_t, \end{aligned} \tag{15}$$

$$a_T = \nabla_x g(X_T). \tag{16}$$

and let $A_t := A(t;X) \in \mathbb{R}^{d\times d}$ solve the second-order adjoint SDE backwardly in time:

$$\begin{aligned} \mathrm{d}A_t = -\Big(&\nabla_x b(x,u(x,t),t)\big|_{x=X_t}^\top A_t + A_t \nabla_x b(x,u(x,t),t)\big|_{x=X_t} + \textstyle\sum_{k=1}^m G_k^u(X_t,t)^\top A_t G_k^u(X_t,t) \\ &+ \nabla_x^2 f(x,u(x,t),t)\big|_{x=X_t} + \textstyle\sum_{i=1}^d a_t^{(i)}\, \nabla_x^2 b_i(x,u(x,t),t)\big|_{x=X_t} \\ &+ \tfrac{1}{2}\textstyle\sum_{k=1}^m \sum_{i=1}^d a_t^{(i)}\, \nabla_x^2\big(\sigma_{\cdot k}^{(i)}(x,u(x,t),t)\big)\big|_{x=X_t} G_k^u(X_t,t)\Big)\,\mathrm{d}t + \textstyle\sum_{k=1}^m U_{t,k}\,\mathrm{d}B_t^k, \\ A_T = \nabla_x^2 g(X_T), \end{aligned} \tag{17}$$

where B_t^k denotes the k-th component and the diffusion matrices $U_{t,k} \in \mathbb{R}^{d\times d}$ are given explicitly by

$$U_{t,k} = -\Big(A_t\, G_k^u(X_t,t) + G_k^u(X_t,t)^\top A_t + \textstyle\sum_{i=1}^d a_t^{(i)}\, \nabla_x^2\big(\sigma_{ik}(x,u(x,t),t)\big)|_{x=X_t}\Big). \tag{18}$$

We note that although equation (15) and equation (17) are posed backward in time, every diffusion coefficient ($G_k^\top a_t$ and $U_{t,k}$, respectively) is given as an explicit function of the current state. These equations can therefore be solved pathwise via time-reversal. This is in contrast to the *backward stochastic differential equation (BSDE)* for the SMP adjoint introduced in Section 4, where an additional unknown martingale integrand q_t must be determined jointly with the backward state.

With the above adjoint processes in hand, the following general basic matching loss function provides a reformulation of the original SOC problem. Its proof can be found in Appendix A.

Theorem 1 (General Hamiltonian basic adjoint matching)**.** *Consider the SOC problem (3)–(4). Define the Hamiltonian*

$$H(x,t;u,p,M) := f(x,u,t) + \langle b(x,u,t),p\rangle + \tfrac{1}{2}\operatorname{Tr}\big(\sigma(x,u,t)\sigma(x,u,t)^\top M\big), \tag{19}$$

$$\text{where } (x,t,u,p,M) \in \mathbb{R}^d \times [0,T] \times \mathbb{R}^k \times \mathbb{R}^d \times \mathbb{R}^{d\times d}, \tag{20}$$

and consider the basic adjoint matching loss

$$\mathcal{L}_{\mathrm{BAM}}(u;X^{\bar u}) := \int_0^T H\big(X_t^{\bar u},t;\, u(X_t^{\bar u},t),\, a(t;X^{\bar u}),\, A(t;X^{\bar u})\big)\,\mathrm{d}t, \qquad \bar u = \operatorname{stopgrad}(u). \tag{21}$$

Alternatively, $a(t;X^{\bar u})$ and $A(t;X^{\bar u})$ can be replaced by any unbiased estimates of $\nabla_x J(\bar u;X_t^{\bar u},t)$ and $\nabla_x^2 J(\bar u;X_t^{\bar u},t)$ (cf. (13) and (14)).

Then, the functional $u \mapsto \mathbb{E}[\mathcal{L}_{\mathrm{BAM}}(u;X^{\bar u})]$ has the same first variation as the original control objective (3). Moreover, any critical point[2] $\hat u$ of $u \mapsto \mathbb{E}[\mathcal{L}_{\mathrm{BAM}}(u;X^{\bar u})]$ satisfies the pointwise stationarity condition

$$\nabla_u H\Big(x,t;\hat u(x,t),\nabla_x J(\hat u;x,t),\nabla_x^2 J(\hat u;x,t)\Big) = 0. \tag{22}$$

If, moreover, $u \mapsto H(x,t;u,p,M)$ is convex, then stationarity is equivalent to pointwise minimization:

$$\hat u(x,t) \in \arg\min_{u\in\mathbb{R}^k} H\Big(x,t;u,\nabla_x J(\hat u;x,t),\nabla_x^2 J(\hat u;x,t)\Big). \tag{23}$$

In this case, $J(\hat u;\cdot,\cdot)$ satisfies the HJB equation

$$\partial_t J + \inf_{u\in\mathbb{R}^k} H\Big(x,t;u,\nabla_x J,\nabla_x^2 J\Big) = 0, \qquad J(\cdot,T) = g. \tag{24}$$

In particular, if the HJB solution is unique, then $\hat u = u^\star$ is the optimal control.

The BAM objective (21) in Theorem 1 provides a variationally correct Hamiltonian formulation for general SOC, but its implementation requires integrating the $d \times d$ matrix SDE equation (17) for the second-order adjoint A_t. In the practically important setting $\sigma(x,u,t) = \sigma(t)$, standard in diffusion models, two key simplifications occur: (i) $G_k^u \equiv 0$, so the first-order adjoint SDE equation (15) reduces to an ODE; and (ii) the trace term $\frac{1}{2}\operatorname{Tr}(\sigma(t)\sigma(t)^\top A_t)$ in the Hamiltonian equation (19) no longer depends on u, so A_t drops out of the stationarity condition, eliminating the need for the matrix SDE entirely. While this specializes the diffusion, the drift and running cost remain fully general, extending beyond the control-affine quadratic-cost setting of Domingo-Enrich et al. (2025); this yields a lean AM loss whose critical points, under a standard HJB uniqueness assumption, recover the optimal control.

Theorem 2 (General Hamiltonian (lean) adjoint matching)**.** *Consider the SOC problem and notation from Theorem 1, in the particular case that the diffusion coefficient depends only on time: $\sigma(x,u,t) := \sigma(t)$. Let $\nabla_1 f(x,u(x,t),t)$ and $\nabla_1 b(x,u(x,t),t)$ denote the partial derivatives with respect to the first argument x, as opposed to $\nabla_x f(y,u(y,t),t)\big|_{x=y}^\top$ and $\nabla_x b(y,u(y,t),t)\big|_{x=y}^\top$, which are the total derivatives. Let $\tilde a(t;X)$ solve the lean adjoint ODE:*

$$\tfrac{\mathrm{d}\tilde a}{\mathrm{d}t}(t;X) = -\nabla_1 b(X_t,u(X_t,t),t)^\top\,\tilde a(t;X) - \nabla_1 f(X_t,u(X_t,t),t), \tag{25}$$

$$\tilde a(T;X) = \nabla_x g(X_T). \tag{26}$$

[2]A critical point is a point with zero first-order variation.

Define the simplified Hamiltonian

$$\tilde{H}(x,t;u,p) = f(x,u,t) + \langle b(x,u,t), p\rangle, \tag{27}$$

and the (lean) adjoint matching loss

$$\mathcal{L}_{\mathrm{AM}}(u;X^{\bar{u}}) := \int_0^T \tilde{H}\big(X_t^{\bar{u}}, t;\, u(X_t^{\bar{u}}, t),\, \tilde{a}(t;X^{\bar{u}})\big)\,\mathrm{d}t, \qquad \bar{u} = \mathrm{stopgrad}(u). \tag{28}$$

Any critical point $\hat{u}$ of $u \mapsto \mathbb{E}[\mathcal{L}_{\mathrm{AM}}(u;X^{\bar{u}})]$ satisfies the pointwise stationarity condition

$$\nabla_u \tilde{H}\Big(x,t;\hat{u}(x,t), \nabla_x J(\hat{u};x,t)\Big) = 0. \tag{29}$$

If, moreover, $u \mapsto \tilde{H}(x,t;u,p)$ is convex, then stationarity is equivalent to pointwise minimization:

$$\hat{u}(x,t) \in \arg\min_{u\in\mathbb{R}^k} \tilde{H}\Big(x,t;u,\nabla_x J(\hat{u};x,t)\Big). \tag{30}$$

In particular, if the HJB solution is unique, then $\hat{u} = u^\star$ is the optimal control.

The proof is given in Appendix B. We now verify that the original adjoint matching objective of Domingo-Enrich et al. (2025) is recovered as a special case.

Remark 3 (Recovering the original adjoint matching loss)**.** *The control-affine quadratic-cost SOC problem equation (5) studied by Domingo-Enrich et al. (2025) (cf. Section 2.3) is a special case of Theorem 2. Using the notation of Section 2.2, the drift, running cost, and diffusion in the general formulation equation (3) specialize to*

$$b(x,u,t) = b(x,t) + \sigma(t)u, \qquad f(x,u,t) = f(x,t) + \tfrac{1}{2}\|u\|^2, \qquad \sigma(x,u,t) = \sigma(t), \tag{31}$$

where $b(x,t)$, $f(x,t)$, and $\sigma(t)$ on the right-hand side are the base drift, state cost, and diffusion coefficient from equation (5). The simplified Hamiltonian equation (27) then becomes

$$\tilde{H}(x,t;u,p) = f(x,t) + \tfrac{1}{2}\|u\|^2 + \langle b(x,t) + \sigma(t)u,\, p\rangle, \tag{32}$$

and the lean adjoint ODE equation (25) reduces to equation (9). Substituting into the lean adjoint matching loss equation (28) and dropping terms independent of u yields

$$\mathcal{L}_{\mathrm{AM}}(u;X^{\bar{u}}) = \int_0^T \tfrac{1}{2}\|u(X_t^{\bar{u}},t) + \sigma(t)^\top \tilde{a}(t;X^{\bar{u}})\|^2\,\mathrm{d}t, \qquad \bar{u} = \mathrm{stopgrad}(u), \tag{33}$$

which recovers exactly the adjoint matching objective in (Domingo-Enrich et al., 2025, Equations (37)–(39)) as stated in Section 2.3.

4 The Stochastic Maximum Principle

While Theorem 2 establishes that the lean adjoint loss has the correct stationarity properties, the variational proof does not make explicit the control-theoretic structure behind the result. To reveal this structure, we now analyze the problem through the lens of maximum principles. We begin with the deterministic Pontryagin maximum principle (PMP) and its associated method of successive approximations (MSA), which provide a clean algorithmic framework. We then present the stochastic maximum principle (SMP) (Yong & Zhou, 2012), which generalizes the PMP to stochastic settings, and explain why the presence of a martingale term in the SMP adjoint equation renders classical MSA-based algorithms impractical.

Pontryagin maximum principle and method of successive approximations. The Pontryagin maximum principle (PMP) provides necessary optimality conditions for deterministic control problems of the form

$$\min_{u\in\mathcal{U}} \ \mathbb{E}\Big[\int_0^T f(X_t^u, u(X_t^u,t),t)\,\mathrm{d}t + g(X_T^u)\Big], \tag{34}$$

$$\text{s.t.} \quad \mathrm{d}X_t^u = b(X_t^u, u(X_t^u,t),t)\,\mathrm{d}t, \qquad X_0^u \sim P_0. \tag{35}$$

Recalling the simplified Hamiltonian $\tilde{H}(x,t;u,p) = f(x,u,t) + \langle b(x,u,t), p\rangle$ from equation (27), the PMP takes the form

$$\begin{cases} \mathrm{d}X_t^* = b(X_t^*, u_t^*, t)\,\mathrm{d}t, & (36)\\ \mathrm{d}p_t^* = -\nabla_x \tilde{H}\big(X_t^*, t; u_t^*, p_t^*\big)\,\mathrm{d}t, & (37)\\ X_0^* \sim P_0, \qquad p_T^* = \nabla_x g(X_T^*), & (38)\\ \tilde{H}\big(X_t^*, t; u_t^*, p_t^*\big) = \min\limits_{u\in\mathbb{R}^k} \tilde{H}\big(X_t^*, t; u, p_t^*\big). & (39)\end{cases}$$

A crucial feature of the PMP is that all quantities are fully specified: the forward state ODE and the backward adjoint ODE have known coefficients, and the Hamiltonian minimization is a pointwise optimization. This gives rise to the method of successive approximations (MSA, Chernousko & Lyubushin (1982), (Li et al., 2018, Algorithm 1)), an iterative procedure repeating the following three steps:

1. Solve forwardly $\mathrm{d}X_t^k = b(X_t^k, u_t^k, t)\,\mathrm{d}t, \quad X_0^k \sim P_0$.
2. Solve backwardly $\mathrm{d}p_t^k = -\nabla_x \tilde{H}\big(X_t^k, t; u_t^k, p_t^k\big)\,\mathrm{d}t, \quad p_T^k = \nabla_x g(X_T^k)$.
3. Solve the Hamiltonian optimization condition:

$$u_t^{k+1} \in \arg\min\nolimits_{u\in\mathbb{R}^k}\{\tilde{H}\big(X_t^k, t; u, p_t^k\big)\}. \tag{40}$$

Each step of MSA is implementable, and the algorithm has well-studied convergence properties in the deterministic setting (Li et al., 2018).

Stochastic maximum principle. The SMP generalizes the PMP to stochastic control problems. In what follows, we restrict attention to the first-order version of SMP when the diffusion coefficient does not depend on the control, $\sigma(x,u,t) := \sigma(x,t)$, which captures the regime relevant to our analysis. The full second-order formulation of SMP for general control-dependent diffusion is summarized in Appendix C.2 for completeness.

To state the SMP, we introduce another Hamiltonian that couples the backward adjoint p with the martingale integrand q:

$$\mathcal{H}(x,t;u,p,q) = f(x,u,t) + \langle p, b(x,u,t)\rangle + \mathrm{tr}(\sigma(x,t)^\top q), \tag{41}$$

$$\text{where } (x,t,u,p,q) \in \mathbb{R}^d \times [0,T] \times \mathbb{R}^k \times \mathbb{R}^d \times \mathbb{R}^{d\times m}. \tag{42}$$

For the SOC problem (5), SMP states the following necessary condition for optimality. For consistency with Section 3, we use the *minimization* convention (i.e. the Hamiltonian is minimized over u): given an optimal control $u_t^* := u^*(X_t^*, t)$ and its corresponding state process X_t^*, there exist adapted processes $p_t^* \in \mathbb{R}^d$ and $q_t^* \in \mathbb{R}^{d\times m}$ such that

$$\begin{cases} \mathrm{d}X_t^* = b(X_t^*, u_t^*, t)\,\mathrm{d}t + \sigma(X_t^*, t)\,\mathrm{d}B_t, & (43)\\ dp_t^* = -\nabla_x \mathcal{H}\big(X_t^*, t; u_t^*, p_t^*, q_t^*\big)\,\mathrm{d}t + q_t^*\,\mathrm{d}B_t, & (44)\\ X_0^* \sim P_0, \qquad p_T^* = \nabla_x g(X_T^*), & (45)\\ \mathcal{H}\big(X_t^*, t; u_t^*, p_t^*, q_t^*\big) = \min\limits_{u\in\mathbb{R}^k} \mathcal{H}\big(X_t^*, t; u, p_t^*, q_t^*\big). & (46)\end{cases}$$

Here 'adapted' means p_t and q_t depend only on the Brownian history up to time t; more precisely, they are $\mathcal{F}_t$-measurable, where $(\mathcal{F}_t)$ is the filtration generated by B. Note that since $\sigma(X_t^*, t)$ does not depend on u, the trace term $\mathrm{tr}(\sigma(X_t^*, t)^\top q_t^*)$ drops out of the minimization and equation (46) is equivalently $\tilde{H}(X_t^*, t; u_t^*, p_t^*) = \min_u \tilde{H}(X_t^*, t; u, p_t^*)$, recovering the same pointwise condition (39) as in the PMP.

Comparing the SMP with the PMP reveals a fundamental structural difference. In the PMP, the backward adjoint equation equation (37) is an ODE with known coefficients. In the SMP, the backward adjoint equation equation (44) contains an additional stochastic term $q_t^*\,\mathrm{d}B_t$, where the diffusion coefficient q_t^* is not prescribed a priori but must be solved jointly with p^* so that the pair is adapted to the underlying filtration. Equations of this type are *backward stochastic differential equations* (BSDEs) Pardoux & Peng (2005), in which both the backward state and its martingale integrand are unknowns determined implicitly

by the terminal condition and the dynamics. As noted in Section 3, this stands in contrast to the adjoint SDEs equation (15)–equation (17), whose diffusion coefficients are explicit and which can therefore be solved pathwise.

In the context of SMP, as shown in Appendix C.1, the process (44) admits a concrete interpretation:

$$p_t^* = \nabla_x V(X_t^*, t), \qquad q_t^* = \nabla_x^2 V(X_t^*, t)\sigma(X_t^*, t), \tag{47}$$

where $V(x,t) := J(u^*; x, t)$ is the value function. In particular, q_t^* encodes second-order value function information, which is precisely the quantity that the lean adjoint matching avoids computing.

The obstacle to MSA in stochastic settings. The BSDE structure of the SMP adjoint has direct algorithmic consequences. Kerimkulov et al. (2021) studied a variant of the MSA for stochastic problems, but their algorithm (and any MSA algorithm based on the SMP) requires access to the matrix q_t^k in order to solve the BSDE $dp_t^k = -\nabla_x \mathcal{H}\big(X_t^k, t; u_t^k, p_t^k, q_t^k\big)\,\mathrm{d}t + q_t^k\,\mathrm{d}B_t$, which the SMP does not specify how to compute. This is required even when σ does not depend on u or x, as q_t^k always appears in the stochastic term $q_t^k\,\mathrm{d}B_t$. In principle, one could approximate q_t^k by solving the full high-dimensional BSDE using deep learning–based methods such as the Deep BSDE approach E et al. (2017); Han et al. (2018) (see Han et al. (2024) for a review), but this incurs substantial additional cost. Alternatively, using (47), one could construct $q_t^k = A(t, X^k)\sigma(X_t^k, u_t^k, t)$, where $A(t, X^k)$ is the unbiased estimate of $\nabla_x^2 J(u^k; X_t^k, t)$ that solves the matrix SDE equation (17); however, this has not been tried before and is significantly more expensive as it requires propagating a $d \times d$ matrix SDE. Hence, existing MSA approaches in the stochastic setting have not resulted in practical algorithms.

5 Connecting Adjoint Matching to the Stochastic Maximum Principle

Having established the adjoint matching losses as principled objectives in Section 3 and reviewed the stochastic maximum principle in Section 4, we now connect the two. Recall that the results of Section 3 are variational: they show that the lean AM loss has the correct critical points, so that any stationary control is optimal, but do not say whether the *optimization dynamics* used to reach those critical points are themselves principled. We now formulate the MSA induced by the SMP in the $\sigma = \sigma(t)$ setting, and then show that the gradient flow of the lean AM loss coincides with a continuous-time version of this procedure.

Consider the SMP in the case where $\sigma(x, u, t) = \sigma(t)$, which simplifies to

$$\begin{cases} \mathrm{d}X_t^* = b(X_t^*, u_t^*, t)\mathrm{d}t + \sigma(t)\mathrm{d}B_t, & (48)\\ dp_t^* = -\big(\nabla_1 f(X_t^*, u_t^*, t) + \nabla_1 b(X_t^*, u_t^*, t)^\top p_t^*\big)\,\mathrm{d}t + q_t^*\,\mathrm{d}B_t, & (49)\\ X_0^* \sim P_0, \qquad p_T^* = \nabla_x g(X_T^*), & (50)\\ 0 = \nabla_2 f(X_t^*, u_t^*, t) + \nabla_2 b(X_t^*, u_t^*, t)^\top p_t^*. & (51)\end{cases}$$

The resulting MSA reads:

1. Solve forwardly $\mathrm{d}X_t^k = b(X_t^k, u_t^k, t)\mathrm{d}t + \sigma(t)\mathrm{d}B_t, \quad X_0^k \sim P_0$.
2. Solve backwardly $dp_t^k = -\big(\nabla_1 f(X_t^k, u_t^k, t) + \nabla_1 b(X_t^k, u_t^k, t)^\top p_t^k\big)\,\mathrm{d}t + q_t^k\,\mathrm{d}B_t, \quad p_T^k = \nabla_x g(X_T^k)$.
3. Solve the Hamiltonian optimization condition:

$$u_t^{k+1} \in \arg\min_{u \in \mathbb{R}^k} \{\tilde{H}\big(X_t^k, t; u, p_t^k\big)\}, \qquad \text{where } \tilde{H} \text{ is the simplified Hamiltonian in (27).} \tag{52}$$

Observe that this algorithm faces two obstacles in high dimensions. First, Step 2 is not fully specified: computing q^k requires solving a high-dimensional BSDE, which the SMP does not prescribe how to do (cf. the discussion at the end of Section 4). Second, while the Hamiltonian minimization in Step 3 is well defined and may even admit an explicit solution in function space, this does not directly translate into an implementable update once the control is represented by a neural network; one must instead rely on an

optimization procedure in parameter space. The following theorem resolves both issues at once: the gradient flow of the lean AM loss coincides with a continuous-time MSA that directly optimizes the Hamiltonian of Step 3, while avoiding any explicit computation of the adapted process q_t required in Step 2.

Theorem 4 (Adjoint matching as the continuous-time MSA)**.** *Consider the following continuous-time variant of the MSA where we parameterize the control with a vector field: $u_t = u(X_t^u, t)$, and we consider the following dynamics for the vector field u indexed by a continuous variable τ instead of k:*

$$\begin{cases} \mathrm{d}X_t^\tau = b(X_t^\tau, u^\tau(X_t^\tau, t), t)\mathrm{d}t + \sigma(t)\mathrm{d}B_t, \qquad X_0^\tau \sim P_0, & (53)\\ \mathrm{d}p_t^\tau = -\big(\nabla_1 f(X_t^\tau, u^\tau(X_t^\tau, t), t) + \nabla_1 b(X_t^\tau, u^\tau(X_t^\tau, t), t)^\top p_t^\tau\big)\,\mathrm{d}t + q_t^\tau\,\mathrm{d}B_t, \qquad p_T^\tau = \nabla_x g(X_T^\tau), & (54)\\ \frac{\mathrm{d}}{\mathrm{d}\tau}u^\tau = -\frac{\mathrm{d}}{\mathrm{d}u}\mathbb{E}\big[\int_0^T \tilde{H}(X_t^\tau, t; u(X_t^\tau, t), p_t^\tau)\,\mathrm{d}t\big]|_{u=u^\tau}, & (55)\end{cases}$$

where the last line is understood in the formal functional-derivative sense on the control space. With this interpretation, the variational update induced by the expected lean adjoint matching loss coincides with that induced by the continuous-time MSA objective:

$$\frac{\mathrm{d}}{\mathrm{d}\tau}u^\tau = -\frac{\mathrm{d}}{\mathrm{d}u}\mathbb{E}\big[\mathcal{L}_{\mathrm{AM}}(u; X^{\bar{u}})\big]|_{u=u^\tau}. \tag{56}$$

In other words, each gradient step of the lean AM loss can be viewed as an implementable surrogate for one idealized MSA improvement step: it simulates the forward SDE, uses the corresponding adjoint BSDE p_t^τ as the conceptual SMP reference object, and improves the control by descending the simplified Hamiltonian $\tilde{H}$ rather than by carrying out the full control-space minimization in the original MSA. The key observation, detailed in the proof, is that although p_t^τ in equation (54) is coupled to the unknown q_t^τ through the adjoint BSDE, the lean adjoint $\tilde{a}$ still yields an unbiased estimator of the Hamiltonian gradient with respect to the control, and the gradient flow therefore recovers the MSA control-improvement direction without ever requiring access to q_t, resolving the obstacle identified at the end of Section 4.

This result can be viewed as the stochastic counterpart of an observation by Li et al. (2018) in the deterministic setting, where they show that gradient descent with backpropagation is equivalent to a discrete-time MSA with the Hamiltonian maximization replaced by a single gradient step. Theorem 4 extends this correspondence to the stochastic regime, where the lean adjoint bypasses the martingale term q_t in the SMP adjoint BSDE that would otherwise need to be solved explicitly. More broadly, this suggests that the lean adjoint construction may serve as a practical alternative to full BSDE solvers, opening a route to MSA-type algorithms for stochastic control beyond generative modeling.

Remark 5 (Trust-region perspective)**.** *Interpreting the lean adjoint matching as a continuous-time version of MSA is useful beyond the proof itself. It places the lean adjoint matching within a well-studied control-theoretic framework and suggests concrete directions for more stable algorithm design. In the deterministic setting, Li et al. (2018) show that the basic MSA can diverge when the Hamiltonian maximization step induces large "feasibility errors" in the forward–backward dynamics, and propose an extended MSA that regularizes the control update to prevent this. This analysis carries a clear message for the lean adjoint matching in the stochastic setting: updates should be moderated when the iterate is still far from a self-consistent forward–backward pair. This viewpoint naturally motivates trust-region style variants of the lean adjoint matching, where each step is restricted to stay within a neighborhood of the current iterate, improving robustness and preventing divergence, as recently explored in Blessing et al. (2025). Developing analogous convergence guarantees for the lean adjoint matching, informed by the error estimates of Li et al. (2018), is an interesting direction for future work.*

6 Conclusion

In this work, we revisited adjoint matching from the perspective of stochastic optimal control and clarified its structure at three levels. At the most general level, we introduced a Hamiltonian BAM objective whose expected first variation matches that of the original SOC problem. In the practically important case of purely time-dependent diffusion, this reduces to a lean AM objective that avoids second-order quantities while retaining the correct critical points under the corresponding verification assumptions. Finally, through the

stochastic maximum principle, we showed how the lean adjoint matching can be interpreted as a principled continuous-time MSA-type update that avoids explicit computation of the martingale term in the SMP adjoint BSDE. This perspective provides a principled foundation for the adjoint matching framework and suggests a systematic route for designing scalable SOC-inspired algorithms in generative modeling. From a stochastic control standpoint, our results contribute new implementable objectives for general SOC whose critical points satisfy the HJB stationarity conditions, together with a practical pathway for MSA-type iterations that avoids solving BSDEs. We hope this encourages further exchange between the stochastic control and generative modeling communities.

References

Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J Ballard, Joshua Bambrick, et al. Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature*, 630(8016):493–500, 2024.

Michael S Albergo, Gurtej Kanwar, and Phiala E Shanahan. Flow-based generative models for Markov chain Monte Carlo in lattice field theory. *Physical Review D*, 100(3):034515, 2019.

Michael Samuel Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. In *The Eleventh International Conference on Learning Representations*, 2023.

Michael Arbel, Alex Matthews, and Arnaud Doucet. Annealed flow transport Monte Carlo. In *International Conference on Machine Learning*, pp. 318–330. PMLR, 2021.

Julius Berner, Lorenz Richter, and Karen Ullrich. An optimal control perspective on diffusion-based generative modeling. *arXiv preprint arXiv:2211.01364*, 2023.

Kevin Black, Michael Janner, Yilun Du, Ilya Kostrikov, and Sergey Levine. Training diffusion models with reinforcement learning. In *The Twelfth International Conference on Learning Representations*, 2024.

Denis Blessing, Julius Berner, Lorenz Richter, Carles Domingo-Enrich, Yuanqi Du, Arash Vahdat, and Gerhard Neumann. Trust region constrained measure transport in path space for stochastic optimal control and inference. *arXiv preprint arXiv:2508.12511*, 2025.

Joan Bruna and Jiequn Han. Provable posterior sampling with denoising oracles via tilted transport. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 37, pp. 82863–82894, 2024.

Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.

Felix L Chernousko and AA Lyubushin. Method of successive approximations for solution of optimal control problems. *Optimal Control Applications and Methods*, 3(2):101–114, 1982.

Kevin Clark, Paul Vicol, Kevin Swersky, and David J. Fleet. Directly fine-tuning diffusion models on differentiable rewards. In *The Twelfth International Conference on Learning Representations*, 2024.

Pierre Del Moral, Arnaud Doucet, and Ajay Jasra. Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 68(3):411–436, 2006.

Carles Domingo-Enrich, Jiequn Han, Brandon Amos, Joan Bruna, and Ricky T. Q. Chen. Stochastic optimal control matching. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.

Carles Domingo-Enrich, Michal Drozdzal, Brian Karrer, and Ricky T. Q. Chen. Adjoint matching: Fine-tuning flow and diffusion generative models with memoryless stochastic optimal control. In *The Thirteenth International Conference on Learning Representations*, 2025.

Weinan E, Jiequn Han, and Arnulf Jentzen. Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations. *Communications in mathematics and statistics*, 5(4):349–380, 2017.

Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Müller, Harry Saini, Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, et al. Scaling rectified flow transformers for high-resolution image synthesis. In *Forty-first International Conference on Machine Learning*, 2024.

Ying Fan, Olivia Watkins, Yuqing Du, Hao Liu, Moonkyung Ryu, Craig Boutilier, Pieter Abbeel, Mohammad Ghavamzadeh, Kangwook Lee, and Kimin Lee. DPOK: Reinforcement learning for fine-tuning text-to-image diffusion models. *Advances in Neural Information Processing Systems*, 36:79858–79885, 2023.

Marylou Gabrié , Grant M. Rotskoff, and Eric Vanden-Eijnden. Adaptive Monte Carlo augmented with normalizing flows. *Proceedings of the National Academy of Sciences*, 119(10), mar 2022.

Weiguo Gao, Ming Li, and Qianxiao Li. Terminally constrained flow-based generative models from an optimal control perspective. *arXiv preprint arXiv:2601.09474*, 2026.

Jiequn Han, Arnulf Jentzen, and Weinan E. Solving high-dimensional partial differential equations using deep learning. *Proceedings of the National Academy of Sciences*, 115(34):8505–8510, 2018.

Jiequn Han, Arnulf Jentzen, and Weinan E. A brief review of the Deep BSDE method for solving high-dimensional partial differential equations. In *Proceedings of the International Congress of Basic Science (ICBS)*, 2024.

Yinbin Han, Meisam Razaviyayn, and Renyuan Xu. Stochastic control for fine-tuning diffusion models: Optimality, regularity, and convergence. In *Forty-second International Conference on Machine Learning*, 2025.

Aaron J Havens, Benjamin Kurt Miller, Bing Yan, Carles Domingo-Enrich, Anuroop Sriram, Daniel S. Levine, Brandon M Wood, Bin Hu, Brandon Amos, Brian Karrer, Xiang Fu, Guan-Horng Liu, and Ricky T. Q. Chen. Adjoint sampling: Highly scalable diffusion samplers via adjoint matching. In *Forty-second International Conference on Machine Learning*, 2025.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33. Curran Associates, Inc., 2020.

H J Kappen. Path integrals and symmetry breaking for optimal control theory. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(11), nov 2005.

Bekzhan Kerimkulov, David Šiška, and Lukasz Szpruch. A modified msa for stochastic control problems. *Applied Mathematics & Optimization*, 84(3):3417–3436, 2021.

Patrick Kidger, James Foster, Xuechen Li, Harald Oberhauser, and Terry Lyons. Neural sdes as infinite-dimensional gans. In *International Conference on Machine Learning*, 2021.

Diederik P Kingma, Tim Salimans, Ben Poole, and Jonathan Ho. On density estimation with diffusion models. In A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, 2021.

Matthew Le, Apoorv Vyas, Bowen Shi, Brian Karrer, Leda Sari, Rashel Moritz, Mary Williamson, Vimal Manohar, Yossi Adi, Jay Mahadeokar, et al. Voicebox: Text-guided multilingual universal speech generation at scale. *Advances in neural information processing systems*, 36, 2024.

Rishi Leburu, Levon Nurbekyan, and Lars Ruthotto. Differentiating through stochastic differential equations: A primer. *arXiv preprint arXiv:2601.08594*, 2026.

Qianxiao Li, Long Chen, Cheng Tai, and Weinan E. Maximum principle based algorithms for deep learning. *Journal of Machine Learning Research*, 18(165):1–29, 2018.

Xuechen Li, Ting-Kam Leonard Wong, Ricky TQ Chen, and David Duvenaud. Scalable gradients for stochastic differential equations. In *International Conference on Artificial Intelligence and Statistics*, pp. 3870–3882. PMLR, 2020.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations*, 2023.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *The Eleventh International Conference on Learning Representations*, 2023.

Radford M Neal. Annealed importance sampling. *Statistics and computing*, 11:125–139, 2001.

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

Etienne Pardoux and Shige Peng. Backward stochastic differential equations and quasilinear parabolic partial differential equations. In *Stochastic Partial Differential Equations and Their Applications: Proceedings of IFIP WG 7/1 International Conference University of North Carolina at Charlotte, NC June 6–8, 1991*, pp. 200–217. Springer, 2005.

Mihir Prabhudesai, Anirudh Goyal, Deepak Pathak, and Katerina Fragkiadaki. Aligning text-to-image diffusion models with reward backpropagation (2023). *arXiv preprint arXiv:2310.03739*, 2023.

Yinuo Ren, Wenhao Gao, Lexing Ying, Grant M. Rotskoff, and Jiequn Han. Driftlite: Lightweight drift control for inference-time scaling of diffusion models. In *International Conference on Learning Representations (ICLR)*, 2026.

Danilo Rezende and Shakir Mohamed. Variational inference with normalizing flows. In *Proceedings of the 32nd International Conference on Machine Learning*, 2015.

Lorenz Richter and Julius Berner. Improved sampling via learned diffusions. In *The Twelfth International Conference on Learning Representations*, 2024.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10684–10695, 2022.

Uriel Singer, Adam Polyak, Thomas Hayes, Xi Yin, Jie An, Songyang Zhang, Qiyuan Hu, Harry Yang, Oron Ashual, Oran Gafni, et al. Make-a-video: Text-to-video generation without text-video data. *arXiv preprint arXiv:2209.14792*, 2022.

Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. *arXiv preprint arXiv:1907.05600*, 2019.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations (ICLR 2021)*, 2021.

Wenpin Tang and Fuzhong Zhou. Fine-tuning of diffusion models via stochastic control: entropy regularization and beyond. *arXiv preprint arXiv:2403.06279*, 2024.

Belinda Tzen and Maxim Raginsky. Theoretical guarantees for sampling and inference in generative models with latent diffusions. *arXiv:1903.01608*, 2019.

Masatoshi Uehara, Yulai Zhao, Kevin Black, Ehsan Hajiramezanali, Gabriele Scalia, Nathaniel Lee Diamant, Alex M Tseng, Tommaso Biancalani, and Sergey Levine. Fine-tuning of continuous-time diffusion models as entropy-regularized control. *arXiv preprint arXiv:2402.15194*, 2024.

Apoorv Vyas, Bowen Shi, Matthew Le, Andros Tjandra, Yi-Chiao Wu, Baishan Guo, Jiemin Zhang, Xinyue Zhang, Robert Adkins, William Ngan, et al. Audiobox: Unified audio generation with natural language prompts. *arXiv preprint arXiv:2312.15821*, 2023.

Jiongmin Yong and Xun Yu Zhou. *Stochastic controls: Hamiltonian systems and HJB equations*, volume 43. Springer Science & Business Media, 2012.

Qinsheng Zhang and Yongxin Chen. Path integral sampler: A stochastic control approach for sampling. In *International Conference on Learning Representations*, 2022.

A Proof of Theorem 1: Basic Adjoint Matching for General Hamiltonians

Throughout the proofs, we assume b, σ, f are C^1 in (x, u) and g is C^1, all with derivatives of at most polynomial growth; that the controlled SDE equation (4) is well-posed for every admissible control and sufficiently small perturbations thereof, with finite moments of all relevant orders; and that these bounds justify differentiation under the expectation wherever invoked below. When individual results require additional regularity (e.g. C^2 for the second-order adjoint, or $J(u;\cdot,\cdot) \in C^{1,2}$ for the verification step), this is stated explicitly.

We write the original control objective as

$$\mathcal{J}(u) := \mathbb{E}\Big[\int_0^T f(X_t^u, u(X_t^u, t), t)\,\mathrm{d}t + g(X_T^u)\Big], \qquad \mathrm{d}X_t^u = b(X_t^u, u(X_t^u, t), t)\,\mathrm{d}t + \sigma(X_t^u, u(X_t^u, t), t)\,\mathrm{d}B_t. \tag{57}$$

Fix a Markov control u and a perturbation v (same dimension as u), and consider $u^\epsilon := u + \epsilon v$ with ϵ small. By Theorem 6, the standard first-variation identity reads

$$\tfrac{\mathrm{d}}{\mathrm{d}\epsilon}\mathcal{J}(u^\epsilon)\Big|_{\epsilon=0} = \mathbb{E}\Big[\int_0^T \Big\langle v(X_t^u, t),\, \nabla_u H\big(X_t^u, t;\, u(X_t^u, t),\, p_t^u,\, M_t^u\big)\Big\rangle\,\mathrm{d}t\Big], \tag{58}$$

where the Hamiltonian H is defined in equation (19), and where

$$p_t^u = \nabla_x J(u; X_t^u, t), \qquad M_t^u = \nabla_x^2 J(u; X_t^u, t).$$

Equivalently, one may use any unbiased representation of p_t^u and M_t^u coming from an adjoint/second-adjoint construction, provided it produces the contractions $\langle b, p\rangle$ and $\mathrm{Tr}(\sigma\sigma^\top M)$ inside H.

Now we consider the basic adjoint matching loss (21). Then, for the same perturbation v, we obtain immediately that

$$\tfrac{\mathrm{d}}{\mathrm{d}\epsilon}\,\mathbb{E}[\mathcal{L}_{\mathrm{BAM}}(u^\epsilon; X^{\bar u})]\Big|_{\epsilon=0} = \mathbb{E}\Big[\int_0^T \Big\langle v(X_t^{\bar u}, t),\, \nabla_u H\big(X_t^{\bar u}, t;\, u(X_t^{\bar u}, t),\, a(t, X^{\bar u}), A(t, X^{\bar u})\big)\Big\rangle\,\mathrm{d}t\Big]. \tag{59}$$

Observe that by the tower property of expectation,

$$\begin{aligned}
&\mathbb{E}\Big[\int_0^T \Big\langle v(X_t^{\bar u}, t),\, \nabla_u H\big(X_t^{\bar u}, t;\, u(X_t^{\bar u}, t),\, a(t, X^{\bar u}),\, A(t, X^{\bar u})\big)\Big\rangle\,\mathrm{d}t\Big] \\
&= \mathbb{E}\Big[\int_0^T \Big\langle v(X_t^{\bar u}, t),\, \nabla_2 f(X_t^{\bar u}, u(X_t^{\bar u}, t), t) + \nabla_2 b(X_t^{\bar u}, u(X_t^{\bar u}, t), t)^\top a(t, X^{\bar u}) \\
&\qquad\qquad + \tfrac{1}{2}\nabla_u \mathrm{Tr}\big(\sigma(X_t^{\bar u}, u, t)\sigma(X_t^{\bar u}, u, t)^\top A(t, X^{\bar u})\big)\big|_{u=u(X_t^{\bar u}, t)}\Big\rangle\,\mathrm{d}t\Big] \\
&= \mathbb{E}\Big[\int_0^T \Big\langle v(X_t^{\bar u}, t),\, \nabla_2 f(X_t^{\bar u}, u(X_t^{\bar u}, t), t) + \nabla_2 b(X_t^{\bar u}, u(X_t^{\bar u}, t), t)^\top\, \mathbb{E}[a(t, X^{\bar u})|X_t^{\bar u}] \\
&\qquad\qquad + \tfrac{1}{2}\nabla_u \mathrm{Tr}\big(\sigma(X_t^{\bar u}, u, t)\sigma(X_t^{\bar u}, u, t)^\top\, \mathbb{E}[A(t, X^{\bar u})|X_t^{\bar u}]\big)\big|_{u=u(X_t^{\bar u}, t)}\Big\rangle\,\mathrm{d}t\Big] \\
&= \mathbb{E}\Big[\int_0^T \Big\langle v(X_t^{\bar u}, t),\, \nabla_2 f(X_t^{\bar u}, u(X_t^{\bar u}, t), t) + \nabla_2 b(X_t^{\bar u}, u(X_t^{\bar u}, t), t)^\top \nabla_x J(\bar u; X_t^{\bar u}, t) \\
&\qquad\qquad + \tfrac{1}{2}\nabla_u \mathrm{Tr}\big(\sigma(X_t^{\bar u}, u, t)\sigma(X_t^{\bar u}, u, t)^\top \nabla_x^2 J(\bar u; X_t^{\bar u}, t)\big)\big|_{u=u(X_t^{\bar u}, t)}\Big\rangle\,\mathrm{d}t\Big] \\
&= \mathbb{E}\Big[\int_0^T \Big\langle v(X_t^{\bar u}, t),\, \nabla_u H\big(X_t^{\bar u}, t;\, u(X_t^{\bar u}, t),\, p_t^{\bar u},\, M_t^{\bar u}\big)\Big\rangle\,\mathrm{d}t\Big]
\end{aligned} \tag{60}$$

Here, the third equality follows from the fact that $\nabla_x J(\bar u; x, t) = \mathbb{E}[a(t, X^{\bar u})|X_t^{\bar u} = x]$ and $\nabla_x^2 J(\bar u; x, t) = \mathbb{E}[A(t, X^{\bar u})|X_t^{\bar u} = x]$ by Lemma 7 and Lemma 8, respectively.

Finally, evaluate the above at $\bar{u} = u$ (the intended usage): since the state process and adjoint objects coincide,

$$(X^{\bar{u}}, p^{\bar{u}}, M^{\bar{u}}) = (X^u, p^u, M^u),$$

we obtain the *same* first variation as in equation (58):

$$\tfrac{\mathrm{d}}{\mathrm{d}\epsilon}\,\mathbb{E}[\mathcal{L}_{\mathrm{BAM}}(u^\epsilon; X^{\bar{u}})]\Big|_{\epsilon=0,\ \bar{u}=\mathrm{stopgrad}(u)} = \tfrac{\mathrm{d}}{\mathrm{d}\epsilon}\mathcal{J}(u^\epsilon)\Big|_{\epsilon=0}.$$

This proves the claim that $u \mapsto \mathbb{E}[\mathcal{L}_{\mathrm{BAM}}(u; X^{\bar{u}})]$ has the same first variation as the original objective.

Now let $\hat{u}$ be a critical point of $u \mapsto \mathbb{E}[\mathcal{L}(u; X^{\bar{u}})]$ (with $\bar{u} = \mathrm{stopgrad}(u)$). Then the first variation above vanishes for all perturbations v, hence the integrand must vanish (in the usual weak/pointwise sense), yielding the stationarity condition

$$\nabla_u H\big(x, t; \hat{u}(x,t), \nabla_x J(\hat{u}; x, t), \nabla_x^2 J(\hat{u}; x, t)\big) = 0 \quad \text{for a.e. } (x, t), \tag{61}$$

which is equation (22). If, moreover, $u \mapsto H(x, t; u, p, M)$ is convex, this stationarity is equivalent to the pointwise minimizer condition

$$\hat{u}(x,t) \in \arg\min_u H\big(x, t; u, \nabla_x J(\hat{u}; x, t), \nabla_x^2 J(\hat{u}; x, t)\big). \tag{62}$$

By Lemma 9 (the general f verification step), this implies that $J(\hat{u}; \cdot, \cdot)$ satisfies the HJB equation

$$\partial_t J + \inf_u H(x, t; u, \nabla_x J, \nabla_x^2 J) = 0, \qquad J(\cdot, T) = g. \tag{63}$$

If the HJB solution is unique, then $J(\hat{u}; \cdot, \cdot) = V$ and $\hat{u} = u^\star$ is the optimal control.

Lemma 6 (First variation of $\mathcal{J}$ via adjoints (controlled diffusion))**.** *Fix a Markov control u and a perturbation v of the same dimension, and set $u^\epsilon := u + \epsilon v$. Consider the controlled diffusion*

$$\mathrm{d}X_t^\epsilon = b(X_t^\epsilon, u^\epsilon(X_t^\epsilon, t), t)\,\mathrm{d}t + \sigma(X_t^\epsilon, u^\epsilon(X_t^\epsilon, t), t)\,\mathrm{d}B_t, \qquad X_0^\epsilon \sim P_0,$$

and the objective

$$\mathcal{J}(u^\epsilon) = \mathbb{E}\left[\int_0^T f(X_t^\epsilon, u^\epsilon(X_t^\epsilon, t), t)\,\mathrm{d}t + g(X_T^\epsilon)\right].$$

Assume b, σ, f are C^1 in (x, u) with derivatives of at most polynomial growth, g is C^1 with at most polynomial growth, and that the SDE is well-posed for all sufficiently small ϵ with moments ensuring the differentiations below are justified. Assume in addition that, for the frozen control u, the associated cost-to-go $J(u; \cdot, \cdot)$ is a classical $C^{1,2}$ solution of the backward equation used in the proof below. Then the directional derivative exists and is given by

$$\tfrac{\mathrm{d}}{\mathrm{d}\epsilon}\mathcal{J}(u^\epsilon)\Big|_{\epsilon=0} = \mathbb{E}\left[\int_0^T \Big\langle v(X_t^u, t),\ \nabla_u H\big(X_t^u, t;\ u(X_t^u, t),\ p_t,\ M_t\big)\Big\rangle\,\mathrm{d}t\right], \tag{64}$$

where H is the Hamiltonian

$$H(x, t; u, p, M) = f(x, u, t) + \langle b(x, u, t), p\rangle + \tfrac{1}{2}\,\mathrm{Tr}\big(\sigma(x, u, t)\sigma(x, u, t)^\top M\big),$$

and where (p_t, M_t) are the first- and second-order sensitivity objects

$$p_t = \nabla_x J(u; X_t^u, t), \qquad M_t = \nabla_x^2 J(u; X_t^u, t).$$

Equivalently, one may replace (p_t, M_t) by any unbiased adjoint representation that yields the contractions $\langle b, p\rangle$ and $\mathrm{Tr}(\sigma\sigma^\top M)$ appearing in H.

Lemma 7 (Adjoint method for SDEs with state-dependent diffusion (Domingo-Enrich et al., 2024; Li et al., 2020; Kidger et al., 2021))**.** *Let $(B_t)_{t\in[0,T]}$ be an m-dimensional Brownian motion and let $X : \Omega\times[0,T] \to \mathbb{R}^d$ solve the uncontrolled SDE*

$$\mathrm{d}X_t = b(X_t,t)\,\mathrm{d}t + \sigma(X_t,t)\,\mathrm{d}B_t, \qquad X_0 = x, \tag{65}$$

where $\sigma(x,t) \in \mathbb{R}^{d\times m}$. Let $f : \mathbb{R}^d \times [0,T] \to \mathbb{R}$, $h : \mathbb{R}^d \times [0,T] \to \mathbb{R}^m$, and $g : \mathbb{R}^d \to \mathbb{R}$ be differentiable in x. For $k \in \{1,\dots,m\}$, denote by $\sigma_{\cdot k}(x,t) \in \mathbb{R}^d$ the k-th column of $\sigma(x,t)$, and let

$$G_k(x,t) \;:=\; \nabla_x\sigma_{\cdot k}(x,t) \in \mathbb{R}^{d\times d}, \qquad \nabla_x h_k(x,t) \in \mathbb{R}^d. \tag{66}$$

Define the adapted process $a : \Omega \times [0,T] \to \mathbb{R}^d$ as the solution of the backward SDE

$$\begin{aligned}\mathrm{d}a_t = &\Big(-\nabla_x b(X_t,t)^\top a_t - \nabla_x f(X_t,t) + \textstyle\sum_{k=1}^m G_k(X_t,t)^\top\big(G_k(X_t,t)^\top a_t + \nabla_x h_k(X_t,t)\big)\Big)\,\mathrm{d}t \\ &- \textstyle\sum_{k=1}^m \big(G_k(X_t,t)^\top a_t + \nabla_x h_k(X_t,t)\big)\,\mathrm{d}B_t^k, \qquad a_T = \nabla_x g(X_T).\end{aligned} \tag{67}$$

Then

$$\nabla_{X_0}\big(\textstyle\int_0^T f(X_t,t)\,\mathrm{d}t + \int_0^T \langle h(X_t,t),\mathrm{d}B_t\rangle + g(X_T)\big) = a_0, \tag{68}$$

$$\nabla_x\,\mathbb{E}\Big[\textstyle\int_0^T f(X_t,t)\,\mathrm{d}t + \int_0^T \langle h(X_t,t),\mathrm{d}B_t\rangle + g(X_T)\,\Big|\,X_0 = x\Big] = \mathbb{E}[a_0], \tag{69}$$

and

$$\begin{aligned}&\nabla_x\,\mathbb{E}\Big[\exp\Big(-\textstyle\int_0^T f(X_t,t)\,\mathrm{d}t - \int_0^T \langle h(X_t,t),\mathrm{d}B_t\rangle - g(X_T)\Big)\,\Big|\,X_0 = x\Big] \\ &= -\,\mathbb{E}\Big[a_0\exp\Big(-\textstyle\int_0^T f(X_t,t)\,\mathrm{d}t - \int_0^T \langle h(X_t,t),\mathrm{d}B_t\rangle - g(X_T)\Big)\Big].\end{aligned} \tag{70}$$

Moreover, if $b(x,t) = b_\theta(x,t)$ is differentiable with respect to the parameter θ, and we write X^θ for the solution,

$$\nabla_\theta\big(\textstyle\int_0^T f(X_t,t)\,\mathrm{d}t + \int_0^T \langle h(X_t,t),\mathrm{d}B_t\rangle + g(X_T)\big) = \int_0^T \big(\nabla_\theta b_\theta(X_t^\theta,t)\big)^\top a_t\,\mathrm{d}t, \tag{71}$$

$$\nabla_\theta\,\mathbb{E}\Big[\textstyle\int_0^T f(X_t,t)\,\mathrm{d}t + \int_0^T \langle h(X_t,t),\mathrm{d}B_t\rangle + g(X_T)\,\Big|\,X_0 = x\Big] = \mathbb{E}\Big[\int_0^T \big(\nabla_\theta b_\theta(X_t^\theta,t)\big)^\top a_t\,\mathrm{d}t\Big]. \tag{72}$$

Lemma 8 (Second-order adjoint / Hessian identity for state-dependent diffusion)**.** *Let X solve*

$$\mathrm{d}X_t = b(X_t,t)\,\mathrm{d}t + \sigma(X_t,t)\,\mathrm{d}B_t, \qquad X_0 = x,$$

with $\sigma(x,t) \in \mathbb{R}^{d\times m}$, and define, for $k \in \{1,\dots,m\}$,

$$\sigma_{\cdot k}(x,t) \in \mathbb{R}^d, \qquad G_k(x,t) := \nabla_x\sigma_{\cdot k}(x,t) \in \mathbb{R}^{d\times d}.$$

Assume b,σ,f,h,g are C^2 in x (with polynomial growth bounds and moments ensuring the differentiations and Itô steps below are justified).

Define the pathwise functional

$$F(x) := \textstyle\int_0^T f(X_t,t)\,\mathrm{d}t + \int_0^T \langle h(X_t,t),\mathrm{d}B_t\rangle + g(X_T), \qquad X_0 = x, \tag{73}$$

and its value $J(x) := \mathbb{E}[F(x) \mid X_0 = x]$.

Let $(a_t)_{t\in[0,T]}$ be the first-order adjoint from Lemma 7, i.e. the adapted solution of

$$\begin{aligned}\mathrm{d}a_t = &\Big(-\nabla_x b(X_t,t)^\top a_t - \nabla_x f(X_t,t) + \textstyle\sum_{k=1}^m G_k(X_t,t)^\top\big(G_k(X_t,t)^\top a_t + \nabla_x h_k(X_t,t)\big)\Big)\,\mathrm{d}t \\ &- \textstyle\sum_{k=1}^m \big(G_k(X_t,t)^\top a_t + \nabla_x h_k(X_t,t)\big)\,\mathrm{d}B_t^k, \qquad a_T = \nabla_x g(X_T).\end{aligned} \tag{74}$$

Define also the adapted matrix-valued *process* $A_t \in \mathbb{R}^{d\times d}$ *as the solution of the BSDE*

$$\begin{aligned}\mathrm{d}A_t = -\Big(&\nabla_x b(X_t,t)^\top A_t + A_t \nabla_x b(X_t,t) + \textstyle\sum_{k=1}^m G_k(X_t,t)^\top A_t G_k(X_t,t) + \nabla_x^2 f(X_t,t)\\ &+ \textstyle\sum_{k=1}^m \nabla_x^2 h_k(X_t,t)\, G_k(X_t,t) + \sum_{i=1}^d a_t^{(i)}\, \nabla_x^2 b_i(X_t,t) + \frac{1}{2}\sum_{k=1}^m \sum_{i=1}^d a_t^{(i)}\, \nabla_x^2\big(\sigma_{\cdot k}^{(i)}(X_t,t)\big)\, G_k(X_t,t)\Big)\,\mathrm{d}t\\ &+ \textstyle\sum_{k=1}^m U_{t,k}\,\mathrm{d}B_t^k, \qquad A_T = \nabla_x^2 g(X_T),\end{aligned} \tag{75}$$

where the diffusion matrices $U_{t,k} \in \mathbb{R}^{d\times d}$ *are given explicitly by*

$$U_{t,k} = -\Big(A_t\, G_k(X_t,t) + G_k(X_t,t)^\top A_t + \nabla_x^2 h_k(X_t,t) + \textstyle\sum_{i=1}^d a_t^{(i)}\, \nabla_x^2 \sigma_{ik}(X_t,t)\Big). \tag{76}$$

Then, for all directions $v, w \in \mathbb{R}^d$*, the second directional derivative exists and satisfies the* pathwise Hessian-bilinear identity

$$D_{v,w}^2 F(x) \;=\; \langle v,\; A_0\, w\rangle, \tag{77}$$

and consequently,

$$\langle v,\; (\nabla_x^2 J)(x)\, w\rangle \;=\; \mathbb{E}\left[\langle v,\; A_0\, w\rangle\right]. \tag{78}$$

Equivalently, $\nabla_x^2 J(x) = \mathbb{E}[A_0]$ *(entrywise) under the same differentiation-under-expectation conditions.*

Lemma 9 (General-f HJB verification step)**.** *Let* u *be a Markov control and define the cost-to-go*

$$J(u;x,t) = \mathbb{E}\left[\int_t^T f\big(X_s^u, u(X_s^u,s), s\big)\,\mathrm{d}s + g(X_T^u)\,\middle|\, X_t^u = x\right], \tag{79}$$

where X^u *solves*

$$\mathrm{d}X_s^u = b\big(X_s^u, u(X_s^u,s), s\big)\,\mathrm{d}s + \sigma\big(X_s^u, u(X_s^u,s), s\big)\,\mathrm{d}B_s, \qquad X_t^u = x. \tag{80}$$

Assume $J(u;\cdot,\cdot) \in C^{1,2}$ *and that, for each* (x,t,p,M)*, the map* $\bar u \mapsto H(x,t;\bar u,p,M)$ *attains its minimum, where the Hamiltonian is*

$$H(x,t;\bar u,p,M) := f(x,\bar u,t) + \langle b(x,\bar u,t), p\rangle + \tfrac{1}{2}\,\mathrm{Tr}\big(\sigma(x,\bar u,t)\sigma(x,\bar u,t)^\top M\big). \tag{81}$$

If, for all (x,t)*,*

$$u(x,t) \in \arg\min_{\bar u\in\mathbb{R}^k} H\big(x,t;\bar u, \nabla_x J(u;x,t), \nabla_x^2 J(u;x,t)\big), \tag{82}$$

then $J(u;\cdot,\cdot)$ *satisfies the HJB equation*

$$0 = \partial_t J + \inf_{\bar u\in\mathbb{R}^k}\left\{f(x,\bar u,t) + \langle b(x,\bar u,t), \nabla_x J\rangle + \tfrac{1}{2}\,\mathrm{Tr}\big(\sigma(x,\bar u,t)\sigma(x,\bar u,t)^\top \nabla_x^2 J\big)\right\}, \qquad J(\cdot,T) = g. \tag{83}$$

In particular, if the HJB solution is unique, then $J(u;x,t) = V(x,t)$ *and* u *is optimal.*

A.1 Proof of Lemma 6

Fix a Markov control u and a perturbation v, and set $u^\varepsilon := u + \varepsilon v$. Let X^ε be the controlled diffusion

$$\mathrm{d}X_t^\varepsilon = b\big(X_t^\varepsilon, u^\varepsilon(X_t^\varepsilon,t), t\big)\,\mathrm{d}t + \sigma\big(X_t^\varepsilon, u^\varepsilon(X_t^\varepsilon,t), t\big)\,\mathrm{d}B_t, \qquad X_0^\varepsilon \sim P_0. \tag{84}$$

Step 1: Value function and martingale identity. Define the value function associated with the *fixed* control u by

$$J(x,t) := \mathbb{E}\Big[\int_t^T f\big(X_s^{u;t,x}, u(X_s^{u;t,x}, s), s\big)\,\mathrm{d}s + g(X_T^{u;t,x})\Big],$$

where $X^{u;t,x}$ denotes the solution started from x at time t and driven by u. Under the stated regularity assumptions, $J \in C^{1,2}$ and satisfies the backward PDE

$$\partial_t J(x,t) + H\big(x,t;\, u(x,t), \nabla_x J(x,t), \nabla_x^2 J(x,t)\big) = 0, \qquad J(x,T) = g(x),$$

with Hamiltonian

$$H(x,t;u,p,M) = f(x,u,t) + \langle b(x,u,t), p\rangle + \tfrac{1}{2}\operatorname{Tr}\big(\sigma(x,u,t)\sigma(x,u,t)^\top M\big).$$

Apply Itô's formula to the process $t \mapsto J(X_t^\varepsilon, t)$ under the dynamics of X^ε:

$$\begin{aligned} dJ(X_t^\varepsilon,t) = \Big(&\partial_t J + \langle \nabla_x J, b(X_t^\varepsilon, u^\varepsilon(X_t^\varepsilon,t),t)\rangle \\ &\qquad + \tfrac{1}{2}\operatorname{Tr}\big(\sigma(X_t^\varepsilon, u^\varepsilon(X_t^\varepsilon,t),t)\sigma(X_t^\varepsilon, u^\varepsilon(X_t^\varepsilon,t),t)^\top \nabla_x^2 J\big)\Big)\,\mathrm{d}t \\ &+ \nabla_x J(X_t^\varepsilon,t)\,\sigma(X_t^\varepsilon, u^\varepsilon(X_t^\varepsilon,t),t)\,\mathrm{d}B_t. \end{aligned}$$

Adding $f(X_t^\varepsilon, u^\varepsilon(X_t^\varepsilon,t),t)\,\mathrm{d}t$ and integrating from 0 to T yields

$$g(X_T^\varepsilon) - J(X_0^\varepsilon, 0) + \int_0^T f(X_t^\varepsilon, u^\varepsilon(X_t^\varepsilon,t),t)\,\mathrm{d}t = \int_0^T G_t^\varepsilon\,\mathrm{d}t + \int_0^T Z_t^\varepsilon\,\mathrm{d}B_t, \tag{85}$$

where

$$\begin{aligned} G_t^\varepsilon &:= \partial_t J(X_t^\varepsilon,t) + f(X_t^\varepsilon, u^\varepsilon(X_t^\varepsilon,t),t) + \langle \nabla_x J(X_t^\varepsilon,t), b(X_t^\varepsilon, u^\varepsilon(X_t^\varepsilon,t),t)\rangle \\ &\qquad + \tfrac{1}{2}\operatorname{Tr}\big(\sigma(X_t^\varepsilon, u^\varepsilon(X_t^\varepsilon,t),t)\sigma(X_t^\varepsilon, u^\varepsilon(X_t^\varepsilon,t),t)^\top \nabla_x^2 J(X_t^\varepsilon,t)\big), \\ Z_t^\varepsilon &:= \nabla_x J(X_t^\varepsilon,t)\,\sigma(X_t^\varepsilon, u^\varepsilon(X_t^\varepsilon,t),t). \end{aligned}$$

By the PDE satisfied by J, the drift term can be written as the difference of Hamiltonians:

$$G_t^\varepsilon = H\big(X_t^\varepsilon, t;\, u^\varepsilon(X_t^\varepsilon,t), p(X_t^\varepsilon,t), M(X_t^\varepsilon,t)\big) - H\big(X_t^\varepsilon, t;\, u(X_t^\varepsilon,t), p(X_t^\varepsilon,t), M(X_t^\varepsilon,t)\big),$$

where $p = \nabla_x J$ and $M = \nabla_x^2 J$.

Taking expectations in equation (85) and using that the stochastic integral has zero mean, together with the fact that $\mathbb{E}[J(X_0^\varepsilon, 0)]$ does not depend on ε, we obtain

$$\mathcal{J}(u^\varepsilon) - \mathcal{J}(u) = \mathbb{E}\int_0^T \Big(H(X_t^\varepsilon,t;u^\varepsilon(X_t^\varepsilon,t),p(X_t^\varepsilon,t),M(X_t^\varepsilon,t)) - H(X_t^\varepsilon,t;u(X_t^\varepsilon,t),p(X_t^\varepsilon,t),M(X_t^\varepsilon,t))\Big)\,\mathrm{d}t. \tag{86}$$

Step 2: Differentiation at $\varepsilon = 0$. Divide equation (86) by ε and let $\varepsilon \to 0$. By the C^1 regularity of H in u, polynomial growth bounds, and the convergence $X_t^\varepsilon \to X_t^u$ in L^p, dominated convergence yields

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\mathcal{J}(u^\varepsilon)\Big|_{\varepsilon=0} = \mathbb{E}\int_0^T \Big\langle v(X_t^u,t),\, \nabla_u H\big(X_t^u, t;\, u(X_t^u,t), p_t, M_t\big)\Big\rangle\,\mathrm{d}t,$$

where

$$p_t := \nabla_x J(X_t^u,t), \qquad M_t := \nabla_x^2 J(X_t^u,t).$$

This proves the claim.

A.2 Proof of Lemma 7

We prove the first identity (68); the identity (69) follows by exchanging the gradient and the expectation, and the exponential identity (70) follows by the chain rule.

Step 1: directional derivative and tangent process. Fix a direction $v \in \mathbb{R}^d$ and let

$$\xi_t \;:=\; \nabla_x X_t \, v \in \mathbb{R}^d \tag{87}$$

denote the directional derivative of X_t with respect to the initial condition. Differentiating equation (65) in the direction v yields the (linear) tangent SDE

$$\mathrm{d}\xi_t = \nabla_x b(X_t,t)\,\xi_t\,\mathrm{d}t + \textstyle\sum_{k=1}^m G_k(X_t,t)\,\xi_t\,\mathrm{d}B_t^k, \qquad \xi_0 = v. \tag{88}$$

Step 2: directional derivative of the functional. Define the pathwise functional

$$F(x) \;:=\; \textstyle\int_0^T f(X_t,t)\,\mathrm{d}t + \int_0^T \langle h(X_t,t), \mathrm{d}B_t\rangle + g(X_T), \qquad X_0 = x. \tag{89}$$

By the chain rule, its directional derivative along v is

$$D_v F(x) = \textstyle\int_0^T \langle \nabla_x f(X_t,t), \xi_t\rangle\,\mathrm{d}t + \sum_{k=1}^m \int_0^T \langle \nabla_x h_k(X_t,t), \xi_t\rangle\,\mathrm{d}B_t^k + \langle \nabla_x g(X_T), \xi_T\rangle. \tag{90}$$

Step 3: choose an adjoint so that an Itô product becomes exact. Let a_t be an adapted process to be chosen later and consider the scalar product $\langle a_t, \xi_t\rangle$. Apply Itô's product rule using equation (88):

$$\mathrm{d}\langle a_t, \xi_t\rangle = \langle \mathrm{d}a_t, \xi_t\rangle + \langle a_t, \mathrm{d}\xi_t\rangle + \mathrm{d}\langle a, \xi\rangle_t, \tag{91}$$

where the quadratic covariation term is

$$\mathrm{d}\langle a, \xi\rangle_t = \textstyle\sum_{k=1}^m \Big\langle (\text{diffusion coeff. of } a_t \text{ in } dB^k), (\text{diffusion coeff. of } \xi_t \text{ in } dB^k)\Big\rangle \mathrm{d}t. \tag{92}$$

Write the (unknown) diffusion of a_t as

$$\mathrm{d}a_t = \alpha_t\,\mathrm{d}t + \textstyle\sum_{k=1}^m \beta_{t,k}\,\mathrm{d}B_t^k, \tag{93}$$

with $\alpha_t \in \mathbb{R}^d$, $\beta_{t,k} \in \mathbb{R}^d$ adapted. Then, from equation (88),

$$\langle a_t, \mathrm{d}\xi_t\rangle = \langle a_t, \nabla_x b(X_t,t)\xi_t\rangle \mathrm{d}t + \textstyle\sum_{k=1}^m \langle a_t, G_k(X_t,t)\xi_t\rangle \mathrm{d}B_t^k, \tag{94}$$

and

$$\mathrm{d}\langle a, \xi\rangle_t = \textstyle\sum_{k=1}^m \langle \beta_{t,k},\, G_k(X_t,t)\xi_t\rangle\,\mathrm{d}t = \Big\langle \sum_{k=1}^m G_k(X_t,t)^\top \beta_{t,k},\, \xi_t\Big\rangle \mathrm{d}t. \tag{95}$$

Substituting into equation (91) gives

$$\begin{aligned} \mathrm{d}\langle a_t, \xi_t\rangle = &\Big\langle \alpha_t + \nabla_x b(X_t,t)^\top a_t + \textstyle\sum_{k=1}^m G_k(X_t,t)^\top \beta_{t,k},\, \xi_t\Big\rangle \mathrm{d}t \\ &+ \textstyle\sum_{k=1}^m \Big\langle \beta_{t,k} + G_k(X_t,t)^\top a_t,\, \xi_t\Big\rangle \mathrm{d}B_t^k. \end{aligned} \tag{96}$$

Step 4: match the desired $\mathrm{d}t$ and $\mathrm{d}B_t$ terms. We would like $\mathrm{d}\langle a_t, \xi_t\rangle$ to reproduce (minus) the integrands in equation (90), i.e.

$$\mathrm{d}\langle a_t, \xi_t\rangle = -\langle \nabla_x f(X_t,t), \xi_t\rangle\,\mathrm{d}t - \textstyle\sum_{k=1}^m \langle \nabla_x h_k(X_t,t), \xi_t\rangle\,\mathrm{d}B_t^k. \tag{97}$$

Comparing equation (96) and equation (97), it suffices to choose $\beta_{t,k}$ and α_t so that, for all k,

$$\beta_{t,k} + G_k(X_t,t)^\top a_t = -\nabla_x h_k(X_t,t), \tag{98}$$

$$\alpha_t + \nabla_x b(X_t,t)^\top a_t + \textstyle\sum_{k=1}^m G_k(X_t,t)^\top \beta_{t,k} = -\nabla_x f(X_t,t). \tag{99}$$

From equation (98), $\beta_{t,k} = -G_k(X_t,t)^\top a_t - \nabla_x h_k(X_t,t)$. Substituting into equation (99) yields

$$\alpha_t = -\nabla_x b(X_t,t)^\top a_t - \nabla_x f(X_t,t) + \textstyle\sum_{k=1}^m G_k(X_t,t)^\top \big(G_k(X_t,t)^\top a_t + \nabla_x h_k(X_t,t)\big). \tag{100}$$

Therefore the choice equation (67) makes equation (97) hold.

Step 5: conclude. Integrating equation (97) from 0 to T gives

$$\langle a_T, \xi_T\rangle - \langle a_0, \xi_0\rangle = -\int_0^T \langle \nabla_x f(X_t,t), \xi_t\rangle \,\mathrm{d}t - \sum_{k=1}^m \int_0^T \langle \nabla_x h_k(X_t,t), \xi_t\rangle \,\mathrm{d}B_t^k. \tag{101}$$

Using $\xi_0 = v$ and setting the terminal condition $a_T = \nabla_x g(X_T)$, we obtain

$$\langle a_0, v\rangle = \int_0^T \langle \nabla_x f(X_t,t), \xi_t\rangle \,\mathrm{d}t + \sum_{k=1}^m \int_0^T \langle \nabla_x h_k(X_t,t), \xi_t\rangle \,\mathrm{d}B_t^k + \langle \nabla_x g(X_T), \xi_T\rangle = D_v F(x), \tag{102}$$

where the last equality is equation (90). Since this holds for all directions v, it implies the pathwise identity $\nabla_x F(x) = a_0$.

Step 6: exchange gradient and expectation. Under standard regularity and integrability assumptions ensuring that $\nabla_x F(x)$ is integrable and that differentiation under the expectation is valid, taking expectations in equation (120) yields

$$\nabla_x \,\mathbb{E}\big[F(x) \,\big|\, X_0 = x\big] = \mathbb{E}[\nabla_x F(x)] = \mathbb{E}[a_0]\,, \tag{103}$$

which is the desired identity (69).

Next, we prove the parameter-gradient identities (71) and (72) following an analogous structure. Throughout, fix θ and write $X_t \equiv X_t^\theta$ for the solution of

$$\mathrm{d}X_t = b_\theta(X_t,t)\,\mathrm{d}t + \sigma(X_t,t)\,\mathrm{d}B_t, \qquad X_0 = x, \tag{104}$$

with σ independent of θ. Let a_t be the adjoint process defined in (67) along the trajectory X^θ.

Step 1': directional derivative in parameter space and parameter tangent process. Let $\vartheta \in \mathbb{R}^p$ be an arbitrary direction in parameter space and define

$$\eta_t \;\coloneqq\; D_\vartheta X_t^\theta \in \mathbb{R}^d, \tag{105}$$

the directional derivative of X_t^θ with respect to θ along ϑ. Differentiating equation (104) in direction ϑ yields the (linear) parameter-tangent SDE

$$\mathrm{d}\eta_t = \Big(\nabla_x b_\theta(X_t,t)\,\eta_t + \nabla_\theta b_\theta(X_t,t)\,\vartheta\Big)\,\mathrm{d}t + \sum_{k=1}^m G_k(X_t,t)\,\eta_t\,\mathrm{d}B_t^k, \qquad \eta_0 = 0, \tag{106}$$

where $\nabla_\theta b_\theta(x,t) \in \mathbb{R}^{d\times p}$ and $\nabla_\theta b_\theta(x,t)\,\vartheta \in \mathbb{R}^d$. The initial condition is $\eta_0 = 0$ because $X_0 = x$ does not depend on θ.

Step 2': directional derivative of the functional. Define the pathwise functional

$$F(\theta) \;\coloneqq\; \int_0^T f(X_t,t)\,\mathrm{d}t + \int_0^T \langle h(X_t,t), \mathrm{d}B_t\rangle + g(X_T). \tag{107}$$

By the chain rule, its directional derivative along ϑ is

$$D_\vartheta F(\theta) = \int_0^T \langle \nabla_x f(X_t,t), \eta_t\rangle \,\mathrm{d}t + \sum_{k=1}^m \int_0^T \langle \nabla_x h_k(X_t,t), \eta_t\rangle \,\mathrm{d}B_t^k + \langle \nabla_x g(X_T), \eta_T\rangle. \tag{108}$$

Step 3': Itô product with the same adjoint. Consider the scalar product $\langle a_t, \eta_t\rangle$. Applying Itô's product rule gives

$$\mathrm{d}\langle a_t, \eta_t\rangle = \langle \mathrm{d}a_t, \eta_t\rangle + \langle a_t, \mathrm{d}\eta_t\rangle + \mathrm{d}\langle a, \eta\rangle_t. \tag{109}$$

Write the diffusion of a_t (from (67)) as

$$\mathrm{d}a_t = \alpha_t\,\mathrm{d}t + \sum_{k=1}^m \beta_{t,k}\,\mathrm{d}B_t^k, \tag{110}$$

where

$$\beta_{t,k} = -\big(G_k(X_t,t)^\top a_t + \nabla_x h_k(X_t,t)\big), \tag{111}$$

$$\alpha_t = -\nabla_x b_\theta(X_t,t)^\top a_t - \nabla_x f(X_t,t) + \textstyle\sum_{k=1}^m G_k(X_t,t)^\top \big(G_k(X_t,t)^\top a_t + \nabla_x h_k(X_t,t)\big). \tag{112}$$

From equation (106),

$$\langle a_t, \mathrm{d}\eta_t\rangle = \Big\langle a_t, \nabla_x b_\theta(X_t,t)\eta_t + \nabla_\theta b_\theta(X_t,t)\,\vartheta\Big\rangle \mathrm{d}t + \textstyle\sum_{k=1}^m \langle a_t, G_k(X_t,t)\eta_t\rangle \mathrm{d}B_t^k. \tag{113}$$

Moreover, the quadratic covariation term is

$$\mathrm{d}\langle a,\eta\rangle_t = \textstyle\sum_{k=1}^m \langle \beta_{t,k},\, G_k(X_t,t)\eta_t\rangle\,\mathrm{d}t = \Big\langle \sum_{k=1}^m G_k(X_t,t)^\top \beta_{t,k},\, \eta_t\Big\rangle \mathrm{d}t. \tag{114}$$

Substituting equation (110)–equation (114) into equation (109) yields

$$\begin{aligned}\mathrm{d}\langle a_t,\eta_t\rangle = &\Big\langle \alpha_t + \nabla_x b_\theta(X_t,t)^\top a_t + \textstyle\sum_{k=1}^m G_k(X_t,t)^\top \beta_{t,k},\, \eta_t\Big\rangle \mathrm{d}t + \Big\langle a_t, \nabla_\theta b_\theta(X_t,t)\,\vartheta\Big\rangle \mathrm{d}t \\ &+ \textstyle\sum_{k=1}^m \Big\langle \beta_{t,k} + G_k(X_t,t)^\top a_t,\, \eta_t\Big\rangle \mathrm{d}B_t^k.\end{aligned} \tag{115}$$

Step 4': cancel the $\mathrm{d}t$ and $\mathrm{d}B_t$ terms and keep the forcing term. By the specific choice equation (67) (equivalently equation (111)), we have for all k,

$$\beta_{t,k} + G_k(X_t,t)^\top a_t = -\nabla_x h_k(X_t,t), \qquad \alpha_t + \nabla_x b_\theta(X_t,t)^\top a_t + \textstyle\sum_{k=1}^m G_k(X_t,t)^\top \beta_{t,k} = -\nabla_x f(X_t,t). \tag{116}$$

Substituting these identities into equation (115) gives

$$\mathrm{d}\langle a_t,\eta_t\rangle = -\langle \nabla_x f(X_t,t), \eta_t\rangle\,\mathrm{d}t - \textstyle\sum_{k=1}^m \langle \nabla_x h_k(X_t,t), \eta_t\rangle\,\mathrm{d}B_t^k + \Big\langle a_t, \nabla_\theta b_\theta(X_t,t)\,\vartheta\Big\rangle \mathrm{d}t. \tag{117}$$

Step 5': integrate and conclude the pathwise gradient formula. Integrating equation (117) from 0 to T yields

$$\begin{aligned}\langle a_T,\eta_T\rangle - \langle a_0,\eta_0\rangle = &-\textstyle\int_0^T \langle \nabla_x f(X_t,t), \eta_t\rangle\,\mathrm{d}t - \sum_{k=1}^m \int_0^T \langle \nabla_x h_k(X_t,t), \eta_t\rangle\,\mathrm{d}B_t^k \\ &+ \textstyle\int_0^T \Big\langle a_t, \nabla_\theta b_\theta(X_t,t)\,\vartheta\Big\rangle \mathrm{d}t.\end{aligned} \tag{118}$$

Using $\eta_0 = 0$ and the terminal condition $a_T = \nabla_x g(X_T)$, we rearrange equation (118) to obtain

$$\begin{aligned}\textstyle\int_0^T \Big\langle a_t, \nabla_\theta b_\theta(X_t,t)\,\vartheta\Big\rangle \mathrm{d}t &= \textstyle\int_0^T \langle \nabla_x f(X_t,t), \eta_t\rangle\,\mathrm{d}t + \sum_{k=1}^m \int_0^T \langle \nabla_x h_k(X_t,t), \eta_t\rangle\,\mathrm{d}B_t^k + \langle \nabla_x g(X_T), \eta_T\rangle \\ &= D_\vartheta F(\theta),\end{aligned} \tag{119}$$

where the last equality is equation (108). Since equation (119) holds for all directions $\vartheta \in \mathbb{R}^p$, it implies the vector identity

$$\nabla_\theta F(\theta) = \textstyle\int_0^T \big(\nabla_\theta b_\theta(X_t,t)\big)^\top a_t\,\mathrm{d}t. \tag{120}$$

This proves the first desired identity.

Step 6': exchange gradient and expectation. Under standard regularity and integrability assumptions ensuring that $\nabla_\theta F(\theta)$ is integrable and that differentiation under the expectation is valid, taking expectations in equation (120) yields

$$\nabla_\theta\, \mathbb{E}\big[F(\theta) \,\big|\, X_0 = x\big] = \mathbb{E}[\nabla_\theta F(\theta)] = \mathbb{E}\Big[\textstyle\int_0^T \big(\nabla_\theta b_\theta(X_t,t)\big)^\top a_t\,\mathrm{d}t\Big], \tag{121}$$

which is the second desired identity.

A.3 Proof of Lemma 8

Proof (sketch). We prove equation (77); equation (78) follows by exchanging second differentiation and expectation under the assumed integrability bounds.

Step 1: first and second tangent processes. Fix $v, w \in \mathbb{R}^d$. Let $\xi_t^v := (\nabla_x X_t)\, v$ and $\xi_t^w := (\nabla_x X_t)\, w$. Differentiating the SDE gives the first-order tangent equations

$$\mathrm{d}\xi_t^\bullet = \nabla_x b(X_t, t)\, \xi_t^\bullet \,\mathrm{d}t + \textstyle\sum_{k=1}^m G_k(X_t, t)\, \xi_t^\bullet \,\mathrm{d}B_t^k, \qquad \xi_0^\bullet = \bullet, \quad \bullet \in \{v, w\}.$$

Define the mixed second derivative (second tangent) $\zeta_t^{v,w} := D^2_{v,w} X_t \in \mathbb{R}^d$. A standard differentiation of the tangent SDE yields

$$\begin{aligned} \mathrm{d}\zeta_t^{v,w} &= \Big(\nabla_x b(X_t, t)\, \zeta_t^{v,w} + \nabla_x^2 b(X_t, t)[\xi_t^v, \xi_t^w]\Big)\,\mathrm{d}t \\ &\quad + \textstyle\sum_{k=1}^m \Big(G_k(X_t, t)\, \zeta_t^{v,w} + \nabla_x G_k(X_t, t)[\xi_t^v, \xi_t^w]\Big)\,\mathrm{d}B_t^k, \qquad \zeta_0^{v,w} = 0, \end{aligned} \tag{122}$$

where $\nabla_x^2 b[\cdot,\cdot] \in \mathbb{R}^d$ denotes the bilinear action of the Hessian tensor of b and $\nabla_x G_k[\cdot,\cdot] \in \mathbb{R}^d$ denotes the bilinear action of the derivative of G_k.

Step 2: second directional derivative of the functional. Differentiate $F(x)$ twice (chain rule) to obtain

$$\begin{aligned} D^2_{v,w} F(x) &= \textstyle\int_0^T \Big(\langle \nabla_x^2 f(X_t, t)\, \xi_t^v, \xi_t^w\rangle + \langle \nabla_x f(X_t, t), \zeta_t^{v,w}\rangle\Big)\,\mathrm{d}t \\ &\quad + \textstyle\sum_{k=1}^m \int_0^T \Big(\langle \nabla_x^2 h_k(X_t, t)\, \xi_t^v, \xi_t^w\rangle + \langle \nabla_x h_k(X_t, t), \zeta_t^{v,w}\rangle\Big)\,\mathrm{d}B_t^k \\ &\quad + \langle \nabla_x^2 g(X_T)\, \xi_T^v, \xi_T^w\rangle + \langle \nabla_x g(X_T), \zeta_T^{v,w}\rangle. \end{aligned} \tag{123}$$

Step 3: eliminate the ζ-terms using the first adjoint. Apply Itô's product rule to $\langle a_t, \zeta_t^{v,w}\rangle$ using equation (74) and equation (122). Choose the diffusion of a_t as in equation (74) so that the resulting $\mathrm{d}t$ and $\mathrm{d}B_t$ coefficients match $-\langle \nabla_x f, \zeta\rangle$ and $-\langle \nabla_x h_k, \zeta\rangle$. After integrating from 0 to T and using $\zeta_0^{v,w} = 0$ and $a_T = \nabla_x g(X_T)$, one obtains

$$\begin{aligned} &\langle \nabla_x g(X_T), \zeta_T^{v,w}\rangle + \textstyle\int_0^T \langle \nabla_x f(X_t, t), \zeta_t^{v,w}\rangle\,\mathrm{d}t + \sum_{k=1}^m \int_0^T \langle \nabla_x h_k(X_t, t), \zeta_t^{v,w}\rangle\,\mathrm{d}B_t^k \\ &\quad = \textstyle\int_0^T \Big\langle a_t, \nabla_x^2 b(X_t, t)[\xi_t^v, \xi_t^w]\Big\rangle\,\mathrm{d}t + \sum_{k=1}^m \int_0^T \Big\langle a_t, \nabla_x G_k(X_t, t)[\xi_t^v, \xi_t^w]\Big\rangle\,\mathrm{d}B_t^k \\ &\qquad\qquad + (\text{terms depending only on } \xi^v, \xi^w), \end{aligned} \tag{124}$$

i.e. all ζ-dependence has been traded for terms involving only (a, ξ^v, ξ^w).

Step 4: matrix adjoint and identification of $U_{t,k}$. Consider the scalar process $\langle \xi_t^v, A_t \xi_t^w\rangle$ where

$$\mathrm{d}A_t = \Gamma_t\,\mathrm{d}t + \textstyle\sum_{k=1}^m U_{t,k}\,\mathrm{d}B_t^k.$$

Using the first-order tangent SDEs and Itô's formula, the $\mathrm{d}B_t^k$-coefficient of $\mathrm{d}\langle \xi_t^v, A_t \xi_t^w\rangle$ is

$$\big\langle \xi_t^v,\, \big(U_{t,k} + A_t G_k(X_t, t) + G_k(X_t, t)^\top A_t\big)\xi_t^w\big\rangle.$$

To cancel the remaining stochastic integrands in $D^2_{v,w} F(x)$ for all v, w, we require

$$U_{t,k} + A_t G_k + G_k^\top A_t = -\Big(\nabla_x^2 h_k + \textstyle\sum_{i=1}^d a_t^{(i)}\, \nabla_x^2 \sigma_{ik}\Big),$$

which yields equation (76). With this choice, all dB-terms cancel identically.

Matching the dt-terms then uniquely determines the drift Γ_t, resulting in the BSDE equation (75). Integrating from 0 to T, using $\xi_0^v = v$, $\xi_0^w = w$, and $A_T = \nabla_x^2 g(X_T)$, yields

$$D^2_{v,w} F(x) = \langle v, A_0 w \rangle,$$

which proves equation (77).

The algebraic details of Steps 3 and 4 (Itô product rule computations and coefficient matching) are standard but lengthy; since the present paper only uses this construction to motivate the BAM objective rather than to develop a new second-order SMP theory, we record the structure here and refer to (Yong & Zhou, 2012, Chapter 3) for the full derivation in the general diffusion setting. □

A.4 Proof of Lemma 9

By the dynamic programming principle, for $\Delta t > 0$,

$$J(u; x, t) = \mathbb{E}[J(u; X^u_{t+\Delta t}, t + \Delta t) \mid X^u_t = x] + \mathbb{E}\left[\int_t^{t+\Delta t} f(X^u_s, u(X^u_s, s), s)\, \mathrm{d}s \;\middle|\; X^u_t = x \right].$$

Dividing by Δt and sending $\Delta t \to 0$ gives

$$0 = \mathcal{T}^u J(u; x, t) + f(x, u(x, t), t),$$

where the controlled generator is

$$\mathcal{T}^u \phi = \partial_t \phi + \langle \nabla_x \phi,\, b(x, u(x, t), t) \rangle + \tfrac{1}{2} \operatorname{Tr}\big(\sigma(x, u(x, t), t) \sigma(x, u(x, t), t)^\top \nabla_x^2 \phi \big).$$

Therefore,

$$0 = \partial_t J + f(x, u(x, t), t) + \langle b(x, u(x, t), t), \nabla_x J \rangle + \tfrac{1}{2} \operatorname{Tr}\big(\sigma(x, u(x, t), t) \sigma(x, u(x, t), t)^\top \nabla_x^2 J \big).$$

If $u(x, t)$ achieves the pointwise minimum in equation (82), then the above identity is equivalent to the HJB equation (83).

B Proof of Theorem 2: Lean Adjoint Matching for General Hamiltonians

We follow the same blueprint as (Domingo-Enrich et al., 2025, Prop. 7), but now for a general Hamiltonian and state-dependent diffusion. Let u be an arbitrary control. If $\tilde{a}(t; X^u)$ is the solution of the Lean Adjoint ODE equation (25), it satisfies the integral equation

$$\tilde{a}(t; X^u) = \int_t^T \big(\nabla_1 b(X^u_s, u(X^u_s, s), s)^\top \tilde{a}(s; X^u) + \nabla_1 f(X^u_s, u(X^u_s, s), s) \big)\, \mathrm{d}s + \nabla g(X^u_T). \tag{125}$$

Hence, using the tower property of conditional expectation in the second equality:

$$\mathbb{E}\big[\tilde{a}(t; X^u) \big| X^u_t\big] = \mathbb{E}\Big[\int_t^T \big(\nabla_1 b(X^u_s, u(X^u_s, s), s)^\top \mathbb{E}\big[\tilde{a}(s; X^u) \big| X^u_s\big] + \nabla_1 f(X^u_s, u(X^u_s, s), s) \big)\, \mathrm{d}s + \nabla g(X^u_T) \Big| X^u_t \Big]. \tag{126}$$

Analogously to equations (59), if we consider the perturbation $u^\epsilon(x, t) = u(x, t) + \epsilon v(x, t)$, and let $\bar{u} = \operatorname{stopgrad}(u)$, we have that

$$\begin{aligned} \tfrac{\mathrm{d}}{\mathrm{d}\epsilon} \mathbb{E}[\mathcal{L}_{\mathrm{AM}}(u^\epsilon; X^{\bar{u}})]\Big|_{\epsilon=0} &= \mathbb{E}\left[\int_0^T \Big\langle v(X^{\bar{u}}_t, t),\, \nabla_u \tilde{H}\big(X^{\bar{u}}_t, t;\, \bar{u}(X^{\bar{u}}_t, t),\, \tilde{a}(t, X^{\bar{u}})\big) \Big\rangle \, \mathrm{d}t \right] \\ &= \mathbb{E}\left[\int_0^T \Big\langle v(X^{\bar{u}}_t, t),\, \nabla_u \tilde{H}\big(X^{\bar{u}}_t, t;\, \bar{u}(X^{\bar{u}}_t, t),\, \mathbb{E}[\tilde{a}(t, X^{\bar{u}}) | X^{\bar{u}}_t]\big) \Big\rangle \, \mathrm{d}t \right] \end{aligned} \tag{127}$$

Hence, if $\hat{u}$ is such that the first variation of $u \mapsto \mathbb{E}[\mathcal{L}_{\mathrm{AM}}(u; X^{\bar{u}})]$ at $\hat{u}$ is zero, then

$$\begin{aligned} 0 &= \nabla_u \tilde{H}\big(X^{\hat{u}}_t, t;\, \hat{u}(X^{\hat{u}}_t, t),\, \mathbb{E}[\tilde{a}(t, X^{\hat{u}}) | X^{\hat{u}}_t]\big) \\ &= \nabla_2 f(X^{\hat{u}}_t, \hat{u}(X^{\hat{u}}_t, t), t) + \nabla_2 b(X^{\hat{u}}_t, \hat{u}(X^{\hat{u}}_t, t), t)^\top \mathbb{E}[\tilde{a}(t, X^{\hat{u}}) | X^{\hat{u}}_t]. \end{aligned} \tag{128}$$

Hence, we have

$$\begin{aligned}&\nabla_x \hat{u}(X_t^{\hat{u}},t)^\top \big(\nabla_2 f(X_t^{\hat{u}},\hat{u}(X_t^{\hat{u}},t),t) + \nabla_2 b(X_t^{\hat{u}},\hat{u}(X_t^{\hat{u}},t),t)^\top \,\mathbb{E}[\tilde{a}(t,X^{\hat{u}})|X_t^{\hat{u}}]\big) = 0,\\ &\implies \mathbb{E}\big[\textstyle\int_t^T \nabla_x \hat{u}(X_s^{\hat{u}},s)^\top \big(\nabla_2 f(X_s^{\hat{u}},\hat{u}(X_s^{\hat{u}},s),s) + \nabla_2 b(X_s^{\hat{u}},\hat{u}(X_s^{\hat{u}},s),s)^\top \,\mathbb{E}[\tilde{a}(s,X^{\hat{u}})|X_s^{\hat{u}}]\big)\,\mathrm{d}s\big|X_t^{\hat{u}}\big] = 0.\end{aligned} \tag{129}$$

If we add equation (129) to the right-hand side of equation (126), we obtain that $\mathbb{E}[\tilde{a}(t,X^{\hat{u}})|X_t^{\hat{u}}]$ also solves the following integral equation:

$$\begin{aligned}&\mathbb{E}\big[\tilde{a}(t;X^{\hat{u}})\big|X_t^{\hat{u}}\big]\\ &= \mathbb{E}\Big[\textstyle\int_t^T \big((\nabla_1 b(X_s^{\hat{u}},\hat{u}(X_s^{\hat{u}},s),s) + \nabla_2 b(X_s^{\hat{u}},\hat{u}(X_s^{\hat{u}},s),s)\nabla_x \hat{u}(X_s^{\hat{u}},s))^\top \,\mathbb{E}\big[\tilde{a}(s;X^{\hat{u}})\big|X_s^{\hat{u}}\big]\\ &\qquad + \nabla_1 f(X_s^{\hat{u}},\hat{u}(X_s^{\hat{u}},s),s) + \nabla_x \hat{u}(X_s^{\hat{u}},s)^\top \nabla_2 f(X_s^{\hat{u}},\hat{u}(X_s^{\hat{u}},s),s)\big)\,\mathrm{d}s + \nabla g(X_T^{\hat{u}})\Big|X_t^{\hat{u}}\Big]\\ &= \mathbb{E}\Big[\textstyle\int_t^T \big(\nabla_x b(x,\hat{u}(x,s),s)|_{x=X_s^{\hat{u}}}^\top \,\mathbb{E}\big[\tilde{a}(s;X^{\hat{u}})\big|X_s^{\hat{u}}\big] + \nabla_x f(x,\hat{u}(x,s),s)|_{x=X_s^{\hat{u}}}\big)\,\mathrm{d}s + \nabla g(X_T^{\hat{u}})\Big|X_t^{\hat{u}}\Big].\end{aligned} \tag{130}$$

Now, observe that when $\sigma(x,u,t) := \sigma(t)$, the adjoint SDE (15)- (16) simplifies to the lean adjoint ODE:

$$\frac{\mathrm{d}a_t}{\mathrm{d}t} = -\nabla_x b(x,u(x,t),t)\big|_{x=X_t}^\top a_t - \nabla_x f(x,u(x,t),t)\big|_{x=X_t}, \tag{131}$$

$$a_T = \nabla_x g(X_T), \tag{132}$$

which, setting $u = \hat{u}$, means that $\mathbb{E}\big[a(t;X^{\hat{u}})\big|X_t^{\hat{u}}\big]$ satisfies the following integral equation:

$$\begin{aligned}&\mathbb{E}\big[a(t;X^{\hat{u}})\big|X_t^{\hat{u}}\big]\\ &= \mathbb{E}\Big[\textstyle\int_t^T \big(\nabla_x b(x,\hat{u}(x,s),s)|_{x=X_s^{\hat{u}}}^\top \,\mathbb{E}\big[a(s;X^{\hat{u}})\big|X_s^{\hat{u}}\big] + \nabla_x f(x,\hat{u}(x,s),s)|_{x=X_s^{\hat{u}}}\big)\,\mathrm{d}s + \nabla g(X_T^{\hat{u}})\Big|X_t^{\hat{u}}\Big],\end{aligned} \tag{133}$$

Comparing equation (130) with equation (133), we see that $\mathbb{E}\big[\tilde{a}(t;X^{\hat{u}})\big|X_t^{\hat{u}}\big]$ and $\mathbb{E}\big[a(t;X^{\hat{u}})\big|X_t^{\hat{u}}\big]$ satisfy the same integral equation. Applying Lemma 10, we obtain that

$$\mathbb{E}\big[\tilde{a}(t;X^{\hat{u}})\big|X_t^{\hat{u}}\big] = \mathbb{E}\big[a(t;X^{\hat{u}})\big|X_t^{\hat{u}}\big]. \tag{134}$$

We now evaluate both the BAM and AM first variations at the critical point $u = \hat{u}$ (so that $\bar{u} = \hat{u}$ and all trajectory superscripts coincide). By equations (59) and (60),

$$\begin{aligned}&\tfrac{\mathrm{d}}{\mathrm{d}\epsilon}\,\mathbb{E}[\mathcal{L}_{\mathrm{BAM}}(\hat{u}+\epsilon v;X^{\hat{u}})]\Big|_{\epsilon=0}\\ &= \mathbb{E}\Big[\textstyle\int_0^T \Big\langle v(X_t^{\hat{u}},t),\, \nabla_u H\big(X_t^{\hat{u}},t;\, \hat{u}(X_t^{\hat{u}},t),\, \mathbb{E}\big[a(t;X^{\hat{u}})\big|X_t^{\hat{u}}\big],\, \mathbb{E}\big[A(t;X^{\hat{u}})\big|X_t^{\hat{u}}\big]\big)\Big\rangle\,\mathrm{d}t\Big]\\ &= \mathbb{E}\Big[\textstyle\int_0^T \Big\langle v(X_t^{\hat{u}},t),\, \nabla_u H\big(X_t^{\hat{u}},t;\, \hat{u}(X_t^{\hat{u}},t),\, \mathbb{E}\big[\tilde{a}(t;X^{\hat{u}})\big|X_t^{\hat{u}}\big]\big)\Big\rangle\,\mathrm{d}t\Big] = \tfrac{\mathrm{d}}{\mathrm{d}\epsilon}\,\mathbb{E}[\mathcal{L}_{\mathrm{AM}}(\hat{u}+\epsilon v;X^{\hat{u}})]\Big|_{\epsilon=0}.\end{aligned} \tag{135}$$

Here, the second equality uses that since $\sigma(x,u,t) = \sigma(t)$, the second-order adjoint term drops out of $\nabla_u H$:

$$\begin{aligned}&\nabla_u H\big(X_t^{\hat{u}},t;\, \hat{u}(X_t^{\hat{u}},t),\, \mathbb{E}\big[a(t;X^{\hat{u}})\big|X_t^{\hat{u}}\big],\, \mathbb{E}\big[A(t;X^{\hat{u}})\big|X_t^{\hat{u}}\big]\big) = \nabla_u H\big(X_t^{\hat{u}},t;\, \hat{u}(X_t^{\hat{u}},t),\, \mathbb{E}\big[a(t;X^{\hat{u}})\big|X_t^{\hat{u}}\big]\big)\\ &= \nabla_u H\big(X_t^{\hat{u}},t;\, \hat{u}(X_t^{\hat{u}},t),\, \mathbb{E}\big[\tilde{a}(t;X^{\hat{u}})\big|X_t^{\hat{u}}\big]\big),\end{aligned} \tag{136}$$

where the last step uses equation (134). The rest of the proof is the same as the end of the proof of Theorem 1.

Lemma 10. *Let X^u be a solution of the SDE (4), and let $\mathcal{A} : \mathbb{R}^d \times [0,T] \to \mathbb{R}^{d\times d}$ satisfy $\sup_{t\in[0,T]} \mathbb{E}[\|\mathcal{A}(X_t^u,t)\|^2] < +\infty$. Consider the integral equation*

$$Y_t = \mathbb{E}\Big[\textstyle\int_t^T \big(\mathcal{A}(X_s^u,s)Y_s + c(X_s^u,s)\big)\mathrm{d}s + \psi(X_T^u)\Big|X_t^u\Big], \tag{137}$$

where $c : \mathbb{R}^d \times [0,T] \to \mathbb{R}^d$ and $\psi : \mathbb{R}^d \to \mathbb{R}^d$ are measurable with appropriate integrability, and Y_t is understood as a function of X_t^u. Then this equation has a unique solution.

Proof. Let Y^1, Y^2 be two solutions. Their difference satisfies

$$Y_t^1 - Y_t^2 = \mathbb{E}\Big[\int_t^T \mathcal{A}(X_s^u, s)\big(Y_s^1 - Y_s^2\big)\mathrm{d}s\Big|X_t^u\Big]. \tag{138}$$

Define $\phi(t) := \mathbb{E}\big[\|Y_t^1 - Y_t^2\|\big]$. Taking norms inside the conditional expectation and then full expectations:

$$\begin{aligned}\phi(t) &\le \mathbb{E}\Big[\int_t^T \|\mathcal{A}(X_s^u, s)\| \cdot \|Y_s^1 - Y_s^2\|\,\mathrm{d}s\Big] = \int_t^T \mathbb{E}\big[\|\mathcal{A}(X_s^u, s)\| \cdot \|Y_s^1 - Y_s^2\|\big]\,\mathrm{d}s \\ &\le \int_t^T \big(\mathbb{E}\big[\|\mathcal{A}(X_s^u, s)\|^2\big]\big)^{1/2}\phi(s)\,\mathrm{d}s \le C\int_t^T \phi(s)\,\mathrm{d}s,\end{aligned} \tag{139}$$

where $C := \sup_{s\in[0,T]}\big(\mathbb{E}[\|\mathcal{A}(X_s^u, s)\|^2]\big)^{1/2} < +\infty$. By the backward Grönwall inequality, $\phi(t) = 0$ for all $t \in [0, T]$. Hence $Y_t^1 = Y_t^2$ a.s. for every t. □

C Stochastic Maximum Principle

C.1 Value function representation of the SMP adjoint processes

We first examine the SMP adjoint processes (43)–(46) when the diffusion is control-independent, and prove the identities stated in equation (47)

$$p_t^* = \nabla_x V(X_t^*, t), \qquad q_t^* = \nabla_x^2 V(X_t^*, t)\,\sigma(X_t^*, t),$$

where $V(x, t) := J(u^*; x, t)$ is the value function under the optimal control u^*. We assume V is sufficiently smooth so that the computations below are justified. The argument proceeds in three steps: (1) apply Itô's formula to $\nabla_x V(X_t^*, t)$ to identify the diffusion coefficient, (2) use the HJB equation to simplify the drift, and (3) invoke BSDE uniqueness.

Throughout, we write $b^* := b(X_t^*, u_t^*, t)$, $\sigma^* := \sigma(X_t^*, t)$, and $\Sigma^* := \sigma^*(\sigma^*)^\top$ for brevity.

Step 1: Itô's formula. Define $P_t := \nabla_x V(X_t^*, t) \in \mathbb{R}^d$. Applying Itô's formula componentwise to $\partial_{x_i} V(X_t^*, t)$ gives

$$\mathrm{d}P_t = \Big[\partial_t \nabla_x V + \nabla_x^2 V\, b^* + \tfrac{1}{2}\,\mathcal{T}\Big]\mathrm{d}t + \nabla_x^2 V\,\sigma^*\,\mathrm{d}B_t, \tag{140}$$

where $\mathcal{T} \in \mathbb{R}^d$ has components $\mathcal{T}_i := \sum_{k,\ell} \partial_{x_k x_\ell x_i} V(X_t^*, t)\,\Sigma_{k\ell}^*$, and all functions are evaluated at (X_t^*, t). Setting $Q_t := \nabla_x^2 V(X_t^*, t)\,\sigma(X_t^*, t)$, we read off the diffusion coefficient directly:

$$Q_t = \nabla_x^2 V(X_t^*, t)\,\sigma(X_t^*, t). \tag{141}$$

Step 2: Simplifying the drift via the HJB equation. The value function satisfies the HJB equation

$$0 = \partial_t V + \inf_u \Big\{f(x, u, t) + \langle \nabla_x V,\, b(x, u, t)\rangle + \tfrac{1}{2}\,\mathrm{Tr}\big(\Sigma(x, t)\,\nabla_x^2 V\big)\Big\}, \tag{142}$$

where $\Sigma(x, t) := \sigma(x, t)\sigma(x, t)^\top$. Evaluating at the optimal control u^* and differentiating with respect to x, we obtain

$$0 = \partial_t \nabla_x V + \nabla_1 f^* + (\nabla_1 b^*)^\top \nabla_x V + \nabla_x^2 V\, b^* + \tfrac{1}{2}\nabla_x\,\mathrm{Tr}\big(\Sigma^* \nabla_x^2 V\big), \tag{143}$$

where $\nabla_1 f^* := \nabla_1 f(X_t^*, u_t^*, t)$ and $\nabla_1 b^* := \nabla_1 b(X_t^*, u_t^*, t)$ denote partial derivatives with respect to the first argument x. Expanding the last term componentwise:

$$\Big[\nabla_x\,\mathrm{Tr}\big(\Sigma\nabla_x^2 V\big)\Big]_i = \sum_{k,\ell}(\partial_{x_i}\Sigma_{k\ell})\,\partial_{x_k x_\ell} V + \underbrace{\sum_{k,\ell}\Sigma_{k\ell}\,\partial_{x_k x_\ell x_i} V}_{\mathcal{T}_i}. \tag{144}$$

Solving equation (143) for $\partial_t \nabla_x V$ and substituting into the Itô drift equation (140), the $\nabla_x^2 V\, b^*$ terms cancel and the $\mathcal{T}$ terms cancel, leaving

$$\mathrm{d}P_t = -\Big[\nabla_1 f^* + (\nabla_1 b^*)^\top P_t + \tfrac{1}{2}\sum_{k,\ell}(\nabla_x \Sigma_{k\ell}^*)\,\partial_{x_k x_\ell} V\Big]\mathrm{d}t + Q_t\,\mathrm{d}B_t. \tag{145}$$

Step 3: Matching with the SMP adjoint BSDE. It remains to show that the drift in equation (145) equals $-\nabla_x \mathcal{H}(X_t^*, t; u_t^*, P_t, Q_t)$, where

$$\nabla_x \mathcal{H}(x, t; u, p, q) = \nabla_1 f(x, u, t) + \nabla_1 b(x, u, t)^\top p + \nabla_x \operatorname{Tr}\big(\sigma(x, t)^\top q\big), \tag{146}$$

and the last term differentiates $\sigma(x, t)$ with respect to x while holding q fixed. Evaluating at $q = Q_t = \nabla_x^2 V\, \sigma$:

$$\Big[\nabla_x \operatorname{Tr}\big(\sigma^\top Q_t\big)\Big]_i = \textstyle\sum_{j,k} (\partial_{x_i} \sigma_{jk})\, (Q_t)_{jk} = \sum_{j,k} (\partial_{x_i} \sigma_{jk}) \sum_\ell (\partial_{x_j x_\ell} V)\, \sigma_{\ell k}. \tag{147}$$

Using the identity $\partial_{x_i} \Sigma_{j\ell} = \sum_k [(\partial_{x_i} \sigma_{jk})\, \sigma_{\ell k} + \sigma_{jk}\, (\partial_{x_i} \sigma_{\ell k})]$ together with the symmetry $\partial_{x_j x_\ell} V = \partial_{x_\ell x_j} V$, one verifies that

$$\nabla_x \operatorname{Tr}\big(\sigma^\top Q_t\big) = \tfrac{1}{2} \textstyle\sum_{k,\ell} (\nabla_x \Sigma_{k\ell})\, \partial_{x_k x_\ell} V. \tag{148}$$

Substituting equation (148) into equation (145) shows that P_t satisfies the BSDE

$$\mathrm{d}P_t = -\nabla_x \mathcal{H}\big(X_t^*, t; u_t^*, P_t, Q_t\big)\, \mathrm{d}t + Q_t\, \mathrm{d}B_t, \qquad P_T = \nabla_x g(X_T^*), \tag{149}$$

where the terminal condition follows from $V(x, T) = g(x)$. This is exactly the SMP adjoint BSDE equation (44). By uniqueness of the adapted solution, we conclude $p_t^* = P_t = \nabla_x V(X_t^*, t)$ and $q_t^* = Q_t = \nabla_x^2 V(X_t^*, t)\, \sigma(X_t^*, t)$.

C.2 Second-order statement for control-dependent diffusions

To state the stochastic maximum principle (SMP) in full generality, we first recall the *first-order Hamiltonian*

$$\mathcal{H}(x, t; u, p, q) = f(x, u, t) + \langle p, b(x, u, t)\rangle + \operatorname{Tr}\big(\sigma(x, u, t)^\top q\big), \text{ for } (x, t, u, p, q) \in \mathbb{R}^d \times [0, T] \times U \times \mathbb{R}^d \times \mathbb{R}^{d\times m}. \tag{150}$$

In this case, $U \subseteq \mathbb{R}^k$ is the control domain. Let $u_t^* := u^*(X_t^*, t)$ be an optimal control and X_t^* its associated state process. The *first-order adjoint processes* (p_t^*, q_t^*) are defined as the solution of the backward stochastic differential equation

$$\begin{cases} \mathrm{d}p_t^* = -\nabla_x \mathcal{H}\big(X_t^*, t; u_t^*, p_t^*, q_t^*\big)\, \mathrm{d}t + q_t^*\, \mathrm{d}B_t, & (151) \\ p_T^* = \nabla_x g(X_T^*). & (152) \end{cases}$$

This backward equation remains the same as equation (44) in Section 4 when σ does not depend on the control u.

Second-order adjoint. When the control domain U is nonconvex *and* the diffusion coefficient depends on the control, the first-order adjoint alone is not sufficient to characterize optimality. In this case, the SMP introduces a *second-order adjoint* (P_t^*, Q_t^*), where $P_t^* \in \mathbb{S}^d$ is symmetric and $Q_t^* = (Q_t^{*,1}, \ldots, Q_t^{*,m})$ with $Q_t^{*,k} \in \mathbb{R}^{d\times d}$. The second-order adjoint satisfies the matrix-valued BSDE

$$\begin{aligned} \mathrm{d}P_t^* = -\Big(&\nabla_1 b(X_t^*, u(X_t^*, t), t)^\top P_t^* + P_t^* \nabla_1 b(X_t^*, u(X_t^*, t), t) \\ &+ \textstyle\sum_{k=1}^m \nabla_1 \sigma_{\cdot k}(X_t^*, u(X_t^*, t), t)^\top P_t^* \nabla_1 \sigma_{\cdot k}(X_t^*, u(X_t^*, t), t) \\ &+ \textstyle\sum_{k=1}^m \big(\nabla_1 \sigma_{\cdot k}(X_t^*, u(X_t^*, t), t)^\top Q_t^{*,k} + Q_t^{*,k} \nabla_1 \sigma_{\cdot k}(X_t^*, u(X_t^*, t), t)\big) \\ &+ \nabla_x^2 \mathcal{H}(X_t^*, t; u_t^*, p_t^*, q_t^*)\Big)\mathrm{d}t \end{aligned} \tag{153}$$

$$P_T^* = \nabla_x^2 g(X_T^*). \tag{154}$$

where all derivatives are evaluated along (X_t^*, u_t^*).

Generalized Hamiltonian and optimality condition. Using the second-order adjoint, define the *generalized Hamiltonian*

$$\mathcal{H}_{\mathrm{SMP}}(x,t;u) = \mathcal{H}(x,t;u,p_t^*,q_t^*) + \tfrac{1}{2}\operatorname{Tr}\Big(\big(\sigma(x,u,t)-\sigma(x,u_t^*,t)\big)^\top P_t^*\big(\sigma(x,u,t)-\sigma(x,u_t^*,t)\big)\Big). \tag{155}$$

The stochastic maximum principle (Yong & Zhou, 2012, Theorem 3.2) states that, for almost every t and almost surely with respect to X^*,

$$\mathcal{H}_{\mathrm{SMP}}(X_t^*,t;u_t^*) = \min_{u\in\mathbb{R}^k}\mathcal{H}_{\mathrm{SMP}}(X_t^*,t;u). \tag{156}$$

First-order SMP as a special case. If the diffusion coefficient σ does not depend on the control, *or* if the control domain $U\subseteq\mathbb{R}^d$ is convex so that second-order variations vanish[3], the quadratic correction term in equation (155) disappears. In this case, the SMP reduces to the first-order condition

$$\mathcal{H}\big(X_t^*,t;u_t^*,p_t^*,q_t^*\big) = \min_{u\in\mathbb{R}^k}\mathcal{H}\big(X_t^*,t;u,p_t^*,q_t^*\big), \tag{157}$$

together with the state equation and the first-order adjoint BSDE equation (151). This is the setting considered in the main text of the paper.

D Proof of Theorem 4

To show that $\frac{\mathrm{d}}{\mathrm{d}\tau}u^\tau = -\frac{\mathrm{d}}{\mathrm{d}u}\mathbb{E}\big[\int_0^T \tilde{H}(X_t^\tau,t;u(X_t^\tau,t),p_t^\tau)\,\mathrm{d}t\big]|_{u=u^\tau}$ and $\frac{\mathrm{d}}{\mathrm{d}\tau}u^\tau = -\frac{\mathrm{d}}{\mathrm{d}u}\mathbb{E}\big[\mathcal{L}_{\mathrm{AM}}(u;X^{\bar{u}})\big]|_{u=u^\tau}$ are the same equation, it suffices to show that $\mathbb{E}\big[\int_0^T \tilde{H}(X_t^\tau,t;u(X_t^\tau,t),p_t^\tau)\,\mathrm{d}t\big] = \mathbb{E}\big[\mathcal{L}_{\mathrm{AM}}(u;X^{u^\tau})\big]$ for all functions $u:\mathbb{R}^d\times[0,T]\to\mathbb{R}$. By the tower property, we have that

$$\mathbb{E}\big[\int_0^T \tilde{H}(X_t^\tau,t;u(X_t^\tau,t),p_t^\tau)\,\mathrm{d}t\big] = \mathbb{E}\big[\int_0^T \tilde{H}(X_t^\tau,t;u(X_t^\tau,t),\mathbb{E}\big[p_t^\tau|X_t^\tau\big])\,\mathrm{d}t\big] \tag{158}$$

And by the definition of $\mathcal{L}_{\mathrm{AM}}(u;X^{u^\tau})$, the identity $X^\tau = X^{u^\tau}$, and the tower property, we obtain the following:

$$\begin{aligned}\mathbb{E}\big[\mathcal{L}_{\mathrm{AM}}(u;X^{u^\tau})\big] &= \mathbb{E}\big[\int_0^T \tilde{H}\big(X_t^{u^\tau},t;\,u(X_t^{u^\tau},t),\,\tilde{a}(t;X^{u^\tau})\big)\,\mathrm{d}t\big]\\ &= \mathbb{E}\big[\int_0^T \tilde{H}\big(X_t^\tau,t;\,u(X_t^\tau,t),\,\tilde{a}(t;X^\tau)\big)\,\mathrm{d}t\big] = \mathbb{E}\big[\int_0^T \tilde{H}\big(X_t^\tau,t;\,u(X_t^\tau,t),\,\mathbb{E}[\tilde{a}(t;X^\tau)|X_t^\tau]\big)\,\mathrm{d}t\big]\end{aligned} \tag{159}$$

Hence, it suffices to show the following equality:

$$\mathbb{E}[p_t^\tau|X_t^\tau] = \mathbb{E}[\tilde{a}(t;X^\tau)|X_t^\tau]. \tag{160}$$

Using the argument in equation (126), $\mathbb{E}\big[\tilde{a}(t;X^\tau)\big|X_t^\tau\big]$ satisfies the integral equation

$$\mathbb{E}\big[\tilde{a}(t;X^\tau)\big|X_t^\tau\big] = \mathbb{E}\Big[\int_t^T \big(\nabla_1 b(X_s^\tau,u^\tau(X_s^\tau,s),s)^\top\,\mathbb{E}\big[\tilde{a}(s;X^\tau)\big|X_s^\tau\big] + \nabla_1 f(X_s^\tau,u^\tau(X_s^\tau,s),s)\big)\,\mathrm{d}s + \nabla g(X_T^\tau)\Big|X_t^\tau\Big]. \tag{161}$$

Similarly, we can write an integral equation for the evolution of p_t^τ:

$$p_t^\tau = \int_t^T \big((p_s^\tau)^\top\nabla_1 b(X_s^\tau,u^\tau(X_s^\tau,s),s) + \nabla_1 f(X_s^\tau,u^\tau(X_s^\tau,s),s)\big)\mathrm{d}s - \int_t^T q_s^\tau\mathrm{d}B_s + \nabla g(X_T^\tau). \tag{162}$$

[3]When the control domain U is convex, admissible perturbations of an optimal control can be taken as convex combinations rather than spike variations. In this case, even if the diffusion coefficient depends on the control, first-order variations suffice and the stochastic maximum principle yields a first-order variational inequality $\langle\nabla_u\mathcal{H}(X_t^*,t;u_t^*,p_t^*,q_t^*),u-u_t^*\rangle\geq 0$ for all $u\in U$, which is equivalent (under standard convexity assumptions on $\mathcal{H}$ in u) to pointwise minimization of the Hamiltonian over U. By contrast, when U is nonconvex and the diffusion depends on the control, spike variations produce $O(\varepsilon)$ quadratic diffusion terms, necessitating a second-order adjoint and a generalized Hamiltonian.

Taking conditional expectations of the integral form of the BSDE defining p_t^τ with respect to X_t^τ, and using the tower property together with the martingale property of the stochastic integral, we obtain

$$\mathbb{E}[p_t^\tau|X_t^\tau] = \mathbb{E}\Big[\int_t^T \Big(\mathbb{E}[p_s^\tau|X_s^\tau]^\top \nabla_1 b(X_s^\tau, u^\tau(X_s^\tau, s), s) + \nabla_1 f(X_s^\tau, u^\tau(X_s^\tau, s), s)\Big)\,\mathrm{d}s + \nabla g(X_T^\tau)\,\Big|\, X_t^\tau\Big]. \quad (163)$$

That is, $\mathbb{E}[p_t^\tau|X_t^\tau]$ satisfies the same state-conditioned integral equation above as the lean adjoint conditional expectation $\mathbb{E}[\tilde{a}(t; X^\tau)|X_t^\tau]$ in equation (161) with $u = u^\tau$. Thus, by Lemma 10 we obtain that

$$\mathbb{E}[p_t^\tau|X_t^\tau] = \mathbb{E}[\tilde{a}(t; X^\tau)|X_t^\tau], \quad (164)$$

which concludes the proof.

D.1 Second proof approach in the path integral setting: using the Feynman-Kac formula

In path integral SOC, or quadratic-cost control-affine SOC, we have that $b(x, u, t) = \bar{b}(x, t) + \sigma(x, t)u$ and $f(x, u, t) = \bar{f}(x, t) + \frac{1}{2}\|u(x, t)\|^2$. We give another explanation to equation (160) from the viewpoint of the Feynman-Kac formula (Theorem 11). For the backward stochastic equation for p_t^τ, the solution p_t^τ can be written as

$$p_t^\tau = \mathbb{E}\left[\Phi_{t,T}\,\nabla_x g(X_T^\tau) + \int_t^T \Phi_{t,s}\,\nabla_x \bar{f}(X_s^\tau, s)\mathrm{d}s \,\Big|\, \mathcal{F}_t\right], \quad (165)$$

where $\Phi_{t,s}$ is the fundamental matrix of $\mathrm{d}\Phi_{t,s}/\mathrm{d}s = \nabla_x \bar{b}(X_s^\tau, s)\,\Phi_{t,s}$, $\Phi_{t,t} = I$.

Therefore, given a trajectory $\{X_t^\tau\}_{t=0}^T$ satisfying (48),

$$\Phi_{t,T}\,\nabla_x g(X_T^\tau) + \int_t^T \Phi_{t,s}\,\nabla_x \bar{f}(X_s^\tau, s)\mathrm{d}s$$

is an *unbiased estimator* of $\mathbb{E}[p_t^\tau|X_t^\tau]$. Meanwhile, it is straightforward to verify that

$$\tilde{a}(t; X^\tau) = \Phi_{t,T}\,\nabla_x g(X_T^\tau) + \int_t^T \Phi_{t,s}\,\nabla_x \bar{f}(X_s^\tau, s)\mathrm{d}s. \quad (166)$$

Therefore we have

$$\mathbb{E}[p_t^\tau|X_t^\tau] = \mathbb{E}[\tilde{a}(t; X^\tau)|X_t^\tau]. \quad (167)$$

From this perspective, having a good surrogate for the diffusion term q_t^τ may lead to a better unbiased estimator with reduced variance.

Theorem 11 (Feynman–Kac representation for a linear BSDE)**.** *Let $(X_t)_{t\in[0,T]}$ solve the forward SDE*

$$\mathrm{d}X_t = b(X_t, t)\,\mathrm{d}t + \sigma(X_t, t)\,\mathrm{d}B_t. \quad (168)$$

Assume that $A(\cdot,\cdot)$ and $c(\cdot,\cdot)$ are progressively measurable and bounded, and that $\psi(X_T) \in L^2(\Omega)$. Then there exists a unique adapted solution $(Y, Z) \in \mathcal{S}^2 \times \mathcal{H}^2$ to the linear BSDE

$$\begin{cases} \mathrm{d}Y_t = -\big(A(X_t, t)^\top Y_t + c(X_t, t)\big)\,\mathrm{d}t + Z_t\,\mathrm{d}B_t, \\ Y_T = \psi(X_T). \end{cases} \quad (169)$$

Moreover, the process Y_t admits the representation

$$Y_t = \mathbb{E}\Big[\Phi_{t,T}\,\psi(X_T) + \int_t^T \Phi_{t,s}\,c(X_s, s)\,\mathrm{d}s \,\Big|\, \mathcal{F}_t\Big], \quad (170)$$

where $\Phi_{t,s}$ is the fundamental matrix *(state-transition matrix) solving*

$$\tfrac{\mathrm{d}}{\mathrm{d}s}\Phi_{t,s} = A(X_s, s)\,\Phi_{t,s}, \qquad \Phi_{t,t} = I. \quad (171)$$

When $d = 1$, this reduces to the scalar exponential $\Phi_{t,s} = \exp\big(\int_t^s A(X_\tau, \tau)\,\mathrm{d}\tau\big)$. The process Z_t is uniquely determined by Y_t through the martingale representation theorem.