CONDITIONING DEGENERATE DIFFUSION MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

Current conditioned generative models heavily rely on score functions for guidance during training. When the generative model is a diffusion process with a singular diffusion coefficient and the underlying (conditional) densities either do not exist or are not smooth, we use causal optimal transport to define *approximate* loss functions that identify a minimum-entropy control for guidance under minimal assumptions. Our approach relies on causal optimal transport and its characterization through the predictable representation property of (conditioned) diffusion processes whose associated martingale problem is well posed, à la Üstünel.

1 INTRODUCTION

Conditioned diffusion processes appear in physical chemistry Bolhuis et al. (2002), genetics Wang et al. (2011), economics Elerian et al. (2001), and guided generative models Zhang et al. (2023); Domingo i Enrich et al. (2025). In practice, a common feature of these problems is that the conditioned process must be constructed from data generated by an unconditioned process $X = (X_t)_{0\le t\le T}$ together with a terminal condition Baudoin (2002); Tzen & Raginsky (2019); Pidstrigach et al. (2025). Hence, at time t, the available data are the state history and the condition.

In particular, consider an n-dimensional stochastic differential equation (SDE) of the form

$$dX_t = b(t, X_t)\,dt + \sigma(t, X_t)\,dB_t, \qquad X_0 = \eta, \tag{1}$$

where B is a d-dimensional Brownian motion, and a terminal condition $Y = G(X_T)$, where $G : \mathbb{R}^n \to \mathbb{R}^k$ and η is potentially a random variable. Degeneracy refers to the case in which $\sigma(t, x) \in \mathbb{R}^{n\times d}$ does not have full rank (which can be also extended to the case where the corresponding Malliavin covariance operator is also singular). When the conditional likelihood is sufficiently smooth, the Doob h-transform identifies a controlled equation

$$dX^u_t = \big(b(t, X^u_t) + \sigma(t, X^u_t)u_t(X^u_t, y)\big)\,dt + \sigma(t, X^u_t)\,dB_t, \qquad X^u_0 \sim \mathcal{L}(\eta \mid Y = y), \tag{2}$$

whose state-path law agrees with $\mathcal{L}(X_{[0,T]} \mid Y = y)$ Pidstrigach et al. (2025); Tzen & Raginsky (2019). If $Z^y_t = h_t(X_t; y)$ is a smooth conditional likelihood, then the Brownian-space shift is $u_t = \sigma(t, X_t)^\top \nabla_x \log h_t(X_t; y)$, and the added state drift is $\sigma u = \sigma\sigma^\top \nabla_x \log h$. This shift is the Föllmer drift of the corresponding Brownian realization, and it is characterized by a minimum-entropy property Lassalle (2018), which also has connections with causal optimal transport.

When σ is not invertible, this problem is considerably more difficult than in the invertible case. First, the filtration $\mathcal{F}(X)$ need not coincide with that of the driving (ambient) Brownian motion B. As suggested by the notation, we require the *control* $(u_t(X^u_t, y))_t$ to be $\mathcal{F}(X^u)$-adapted, i.e. *a closed-loop control*. This is important because adaptiveness of u to the path of X^u is crucial for the controlled SDE equation 2 to remain *closed-loop*; otherwise, its implementation would require information external to $\mathcal{F}(X) \vee \sigma(Y)$ generated from equation 1. Furthermore, when G is the identity map, as is the case, for instance, when equation 1 is used as a bridge model, the corresponding conditional law at the terminal time is a Dirac measure and is therefore singular rather than smooth.

2 RELATED WORK AND CONTRIBUTIONS

2.1 RELATED WORK

In this work, we introduce loss functions that yield a control u such that X and X^u, introduced in equation 1 and equation 2, respectively, have the same law. This result relies on the adapted representation property of SDEs satisfying weak uniqueness Üstünel (2019), which was first established

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

in Stroock & Yor (1980); see also Elworthy et al. (2010) for a special case on Riemannian manifolds. Using this property, Üstünel (2019) also characterized the associated causal optimal transport problem. We use these results to derive a *closed-loop* variant of the so-called Föllmer drift $(u_t)_t$ under minimal conditions, which has minimum energy among the admissible shifts (that satisfy Girsanov's theorem). Moreover, through a localization argument, we relax the strict positivity condition imposed in Üstünel (2019) in their treatment of the causal optimal transport problem.

With the advancements of generative models, there has been recent interest on the behavior of degenerate diffusion models. The works Dockhorn et al. (2021); Blessing et al. (2025) study bridge models under degeneracy when the densities are sufficiently smooth. A theory for manifold-based generative models has been developed in De Bortoli (2022); De Bortoli et al. (2022), which are inherently degenerate and satisfy less restrictive smoothness conditions on the corresponding densities compared to their Euclidean variants.

In the presence of degeneracy or non-smooth densities, recently Malliavin calculus based approaches have been used Pidstrigach et al. (2025); Mirafzali et al. (2025). However, Malliavin calculus based approaches still require sufficient smoothness and regularity (as in non-degeneracy of the corresponding Malliavin covariance operator) of b and σ. Several different approaches have also appeared Conforti et al. (2025); Blessing et al. (2025). However, all these works either assume that $\mathcal{L}(X_t|Y=y)$ is continuously differentiable, or suppose that b and σ are sufficiently smooth/regular.

2.2 Contributions

Our contributions in this paper are summarized as follows.

1. We extend the predictable-representation result for degenerate diffusions in Üstünel (2019) to random initializations.
2. In Section 4, we introduce local regression losses that do not require a density or score function. Under local square integrability of the conditional Radon–Nikodym derivative and a state-feedback factorization, the corresponding minimizer is the unique minimum-energy Girsanov shift.
3. We present this loss function as a solution of a control problem (Subsection 5.1). Our loss functions depend on choosing sufficient increments of the data, say ε. To demonstrate the performance of our loss function under different ε, in Subsection 5.2 we provide a numerical experiment. We also provide an application to an image generation problem in Subsection 5.3.

3 Notation

For a process X on a filtered probabiltiy space $(\Omega, \mathcal{F}, (\mathcal{G}_t)_t, \mathbb{P})$, we let $\mathcal{F}_t^0(X) := \sigma\{X_s : 0 \leq s \leq t\}$ and let $\mathcal{F}_t(X)$ denote its usual $\mathbb{P}$-augmentation. For another probability measure $\mathbb{Q}$, $\mathcal{F}_t^{\mathbb{Q}}(X)$ denotes the corresponding $\mathbb{Q}$-augmentation. The ambient filtration is denoted by $(\mathcal{G}_t)_{t\leq T}$. For a measure $\mathbb{Q}$ and a σ-field $\mathcal{H}$, $\mathbb{Q}|_{\mathcal{H}}$ denotes restriction to $\mathcal{H}$. We write $\mathcal{L}_{\mathbb{Q}}(U)$ for the law of U under $\mathbb{Q}$ and $D_{\mathrm{KL}}(\mathbb{Q}\|\mathbb{P})$ for relative entropy. The space $L_a^2(\mathcal{F}(X), \mathbb{Q}; \mathbb{R}^d)$ consists of $\mathcal{F}(X)$-predictable processes v such that $\mathbb{E}^{\mathbb{Q}} \int_0^T |v_t|^2 \, dt < \infty$. The notation $L_a^2([0,r], \mathcal{F}(X), \mathbb{Q}; \mathbb{R}^d)$ uses the same definition with the integral restricted to $[0,r]$, and $L_{a,\mathrm{loc}}^2([0,T), \mathcal{F}(X), \mathbb{Q}; \mathbb{R}^d)$ means that this condition holds on every interval $[0,r]$ with $r < T$. The symbol $A^\dagger$ denotes the Moore–Penrose pseudoinverse and $\|A\|_{\mathrm{F}}$ the Frobenius norm. Given a matrix valued curve $a : [0,T] \to \mathbb{R}^{n\times d}$, we use

$$\widetilde{\Pi}(t,a) := \Pi(a(t)) := a(t)^\top \big(a(t)a^\top(t)\big)^\dagger a(t), \qquad P_t(X) := \widetilde{\Pi}(t, \sigma(\cdot, X_\cdot)) = \Pi(\sigma(t, X_t)),$$

i.e., $\Pi(a(t))$ is the orthogonal projection onto $\mathrm{Ran}(a(t)^\top) = (\ker a(t))^\perp$ under Euclidean inner product; hence $P_t(X)$ is an orthogonal projection onto $\mathrm{Ran}(\sigma(t,X_t)^\top) = (\ker \sigma(t,X_t))^\perp$. All process equalities are understood up to the product measure explicitly specified in the statement.

4 Loss Functions for Degenerate Conditioned Diffusions

In this section, we introduce our main regression formulae and explain how they can be used as population losses. The predictable-representation and minimum-energy arguments are given in Section 6, and the detailed proofs are deferred to the appendix.

Assumption 1 (Regularity and well-posedness)**.** *On a complete filtered probability space $(\Omega, \mathcal{F}, (\mathcal{G}_t)_{t\le T}, \mathbb{P})$ satisfying the usual conditions, B is a d-dimensional $(\mathcal{G}_t)$-Brownian motion and X_0 is $\mathcal{G}_0$-measurable. Equation 1 has a pathwise-continuous strong solution that is unique up to indistinguishability. The maps $b : [0,T] \times \mathbb{R}^n \to \mathbb{R}^n$ and $\sigma : [0,T] \times \mathbb{R}^n \to \mathbb{R}^{n\times d}$ are bounded and Borel measurable. If $\mu_0 = \mathcal{L}_{\mathbb{P}}(X_0)$, then for μ_0-a.e. x the equation started from x has a weak solution that is unique in law. A regular conditional family $(\mathbb{P}_x)_{x\in\mathbb{R}^n}$ of $\mathbb{P}$ given X_0 exists.*

Global Lipschitz continuity in the state variable together with linear growth is a standard sufficient condition for the existence of a unique strong solution for equation 1. Boundedness is used below only for simple integrability estimates and can be replaced by corresponding localized moment assumptions. We only use pathwise uniqueness of equation 1 for the minimum entropy property of the Girsanov shift u that we use to construct equation 2.

Assumption 2 (Regularity of conditional density)**.** *Let $\mathbb{P}^y := \mathbb{P}(\cdot \mid Y = y)$ be a regular conditional probability. For $\mathcal{L}_{\mathbb{P}}(Y)$-a.e. y and every $t < T$,*

$$\mathbb{P}^y|_{\mathcal{F}_t(X)} \ll \mathbb{P}|_{\mathcal{F}_t(X)}, \qquad Z_t^y(X) := \frac{d(\mathbb{P}^y|_{\mathcal{F}_t(X)})}{d(\mathbb{P}|_{\mathcal{F}_t(X)})} \in L^2(\mathcal{F}_t(X), \mathbb{P}).$$

When y is fixed, we suppress the superscript and write $Z_t(X)$ for $Z_t^y(X)$. The only restrictve condition that is used in this work is the square integrability of the process $Z^y(X) = (Z_t^y(X))_t$, which is needed to apply the martingale representation theorem.

Remark 1 (Square integrability and entropy)**.** *The $L^2(\mathbb{P})$ condition is used to invoke the predictable-representation theorem, which is the primary additional restriction in our work. Since for $x > 0$ it holds that $x \log x \le C(1 + x^2)$, Assumption 2 also implies that we have finite entropy: $\mathbb{E}^{\mathbb{P}}[Z_t^y(X) \log Z_t^y(X)] < \infty$ for every $t < T$.*

Assumption 3. *There exists a Borel measurable function G such that $Y = G(X_T)$ and solutions of equation 1 are Markov processes.*

Remark 2. *We need Assumpton 3 to have loss functions that rely on (X_t, Y) at time t rather than $(X_{[0,t]}, Y)$, where $X_{[0,t]}$ denotes the whole trajectory of the process X up to time t.*

For the rest of this section, Assumptions 1, 2, and 3 are in force.

Theorem 1 (Degenerate Tweedie formula-1)**.** *For $\mathcal{L}_{\mathbb{P}}(Y)$-a.e. y and every $r < T$, there is a unique realized process $u_\cdot(X_\cdot, y) \in L^2_{a,\mathrm{loc}}([0,T), \mathcal{F}(X), \mathbb{P}^y; \mathbb{R}^d)$ satisfying*

$$u_t(X_t, y) = \lim_{\varepsilon\downarrow 0} \mathbb{E}^{\mathbb{P}^y}\left[\frac{1}{\varepsilon}\int_t^{t+\varepsilon} P_s(X)\, dB_s \,\middle|\, X_t\right] = \lim_{\varepsilon\downarrow 0} \mathbb{E}^{\mathbb{P}}\left[\frac{1}{\varepsilon}\int_t^{t+\varepsilon} P_s(X)\, dB_s \,\middle|\, X_t, Y = y\right] \tag{3}$$

where both limits are in $L^2([0,r]\times\Omega, dt \otimes d\mathbb{P}^y)$. Moreover, $P_t(X)u_t(X_t, y) = u_t(X_t, y)\ dt \otimes d\mathbb{P}^y$-a.e. There is a corresponding weak solution of equation 2 whose state-path law on $[0,r]$ is $\mathcal{L}_{\mathbb{P}}(X_{[0,r]} \mid Y = y)$, which is given by the solution of equation 1 under the measure $Z_t^y(X)d\mathbb{P}$ with the driving Brownian motion $B_t - \int_0^t u_s(X_s, y)ds$. Among the admissible Girsanov shifts v, u uniquely minimizes $\mathbb{E}^{\mathbb{P}^y}\int_0^r |v_t|^2\, dt$.

Remark 3 (Independent simulation)**.** *The proof relies on Girsanov theorem, which proves existence of a weak controlled solution with the desired law. If the controlled equation 2 is weakly unique, then every weak solution of that equation has the same law. Weak uniqueness of the original equation alone does not imply weak uniqueness after adding a merely measurable control.*

Corollary 1 (Degenerate Tweedie formula-2)**.** *Equation 3 can also be written as*

$$u_t(X_t, y) = \lim_{\varepsilon\downarrow 0} \mathbb{E}^{\mathbb{P}^y}\left[\frac{B_{t+\varepsilon} - B_t}{\varepsilon} \,\middle|\, X_t\right] = \lim_{\varepsilon\downarrow 0} \mathbb{E}^{\mathbb{P}}\left[\frac{B_{t+\varepsilon} - B_t}{\varepsilon} \,\middle|\, X_t, Y = y\right]. \tag{4}$$

A Girsanov shift u changes the state drift only through σu. A component of u in $\ker \sigma$ would therefore leave the state law unchanged while increasing its energy. The minimum-energy control must consequently satisfy $P_t(X)u_t = u_t$ a.e. Corollary 1 is useful in simulation because the Brownian increments are already available and no numerical projection is required.

If only the sample trajectories of the states X and the condition Y are observed, one can instead learn the induced shift σu on the drift of equation 1 instead of u.

Theorem 2 (Degenerate Tweedie formula-3)**.** *Let $u_t(X_t, y)$ be the control obtained from equation 3. Then, for every $r < T$,*

$$\sigma(t, X_t)u_t(X_t, y) = \lim_{\varepsilon \downarrow 0} \mathbb{E}^{\mathbb{P}^y}\left[\frac{X_{t+\varepsilon} - X_t}{\varepsilon} - b(t, X_t) \middle| X_t \right] \tag{5}$$

exists in $L^2\big([0, r] \times \Omega, dt \otimes d\mathbb{P}^y\big)$.

Under the conditioned measure $\mathbb{P}^y$, enlarging the information by Y changes the semimartingale decomposition of X. The predictable-representation result below (Theorem 4) identifies the new finite-variation term as σu and rules out an additional orthogonal martingale contribution. Although σu does not identify an arbitrary Brownian control modulo $\ker \sigma$, it does identify the minimum-energy control. Indeed,

$$\sigma(t, X_t)^\dagger \sigma(t, X_t)u_t = P_t(X)u_t = u_t,$$

where $\sigma^\dagger = \sigma^\top(\sigma\sigma^\top)^\dagger$. Hence, equation 5 recovers the control u through the Moore–Penrose inverse even when σ is noninjective, which then can be used as a Girsanov shift as described in Theorem 1.

The next corollary recovers the classical Tweedie formula when it is available. Notice that the preceding recovery of the control does not recover the entire Brownian path. In a degenerate model, B cannot generally be reconstructed from $\int_0^\cdot \sigma(s, X_s)\, dB_s$. The missing kernel component is irrelevant to the minimum-energy control and to the controlled state equation 2.

Corollary 2. *Suppose that $Z_t^y(X) = h(t, X_t; y)$ with $h > 0$, and that h admits the Itô differential*

$$dh(t, X_t; y) = \big\langle \sigma(t, X_t)^\top \nabla_x h(t, X_t; y), dB_t \big\rangle;$$

for example, this holds for a $C^{1,2}$ solution of the corresponding backward equation. Then,

$$u_t(X_t, y) = \sigma(t, X_t)^\top \nabla_x \log h(t, X_t; y) \quad dt \otimes d\mathbb{P}^y\text{-a.e.}$$

The desired control u can be approximated by the following local regression problem.

Proposition 1. *Fix $r < T$ and $0 < \varepsilon < T - r$. The problem*

$$\inf_{v \in L_a^2(\mathcal{F}(X), dt \otimes d\mathbb{P}^y; \mathbb{R}^d)} \mathbb{E}^{\mathbb{P}^y} \int_0^r \left| v_t - \frac{1}{\varepsilon} \int_t^{t+\varepsilon} P_s(X) dB_s \right|^2 dt. \tag{6}$$

has the unique minimizer given by the product-space conditional expectation that satisfies

$$u_\varepsilon(t, X_t, y) = \mathbb{E}^{\mathbb{P}^y}\left[\frac{1}{\varepsilon} \int_t^{t+\varepsilon} P_s(X)\, dB_s \middle| X_t \right] \quad dt \otimes d\mathbb{P}^y\text{-a.e.}$$

Moreover, $u_\varepsilon \to u_\cdot(X_\cdot, y)$ in $L^2([0, r] \times \Omega, dt \otimes d\mathbb{P}^y)$.

Proof. Conditional expectation is the orthogonal projection onto the closed product-space subspace $L_a^2(\mathcal{F}(X), dt \otimes d\mathbb{P}^y; \mathbb{R}^d)$. The convergence is from Theorem 1. □

Remark 4. *Analogous projection statements hold for Corollary 1 and Theorem 2. Replacing the targets $S_{t_j}^i$ in Algorithm 1 leads to the corresponding algorithms.*

In Algorithm 1 we adapt (Pidstrigach et al., 2025, Algorithm 1) to the degenerate setting under our loss functions.

For example, for an overparameterized ReLU parametrization with finite-dimensional separated inputs, vanishing finite-sample optimization error can be justified under the random-initialization,

Algorithm 1 Training with equation 6

Require: SDE equation 1, batch size N, current control approximation u^θ, uniform grid $t_k = kh$, window $\varepsilon = qh$, and $r < T - \varepsilon$.

1: **for** $i = 1$ to N **do**

2: Sample a path X^i and its corresponding Brownian path B^i from equation 1.

3: Compute the corresponding condition $Y^i = G(X^i_T)$.

4: For every $t_j \le r$, compute the discrete projected-noise target

$$S^i_{t_j} = \frac{1}{qh} \sum_{k=j}^{j+q-1} P_{t_k}(X^i)\big(B^i_{t_{k+1}} - B^i_{t_k}\big). \tag{7}$$

5: Calculate the single-path loss

$$\ell_i(\theta) = h \sum_{\{j:\, t_j \le r\}} \left\| u^\theta_{t_j}(X^i_{t_j}, Y^i) - S^i_{t_j} \right\|^2. \tag{8}$$

6: **end for**

7: Calculate the Monte Carlo full-batch loss

$$\mathcal{L}^N(\theta) = \frac{1}{N} \sum_{i=1}^{N} \ell_i(\theta).$$

8: Take a gradient step on $\mathcal{L}^N(\theta)$.

width, data-separation, and learning-rate conditions of Allen-Zhu et al. (2019), which are sufficient to establish convergence of the algorithm considered in Tzen & Raginsky (2019). In particular, Allen-Zhu et al. (2019) establish linear convergence of gradient descent and stochastic gradient descent for ℓ_2 regression. When these specific conditions are not imposed, we retain vanishing optimization error as an explicit assumption. We discuss the corresponding implementation details in Appendix G. We emphasize, however, that the regression step equation 8 must in practice be approximated using a neural network.

5 Applications

5.1 A Stochastic Control Problem

The variational formula of Boué & Dupuis (1998) is interpreted in Tzen & Raginsky (2019) as a stochastic control problem with a quadratic running cost. This control problem approach has been used in the generative model for conditioning in recent works, for instance, see Domingo i Enrich et al. (2025). We next extend this interpretation to the degenerate setting. The statement is informal; Theorem 10 in Appendix D gives the precise version for an arbitrary target endpoint law.

Theorem 3 (Informal)**.** *Suppose that $X_0 = x_0$ is deterministic, and let*

$$z^y_r := \frac{d\mathcal{L}_{\mathbb{P}}(X_r \mid Y = y)}{d\mathcal{L}_{\mathbb{P}}(X_r)}.$$

Define the endpoint density process and the control by

$$Z^{y,r}_t(X) := \mathbb{E}^{\mathbb{P}}[z^y_r \mid \mathcal{F}_t(X)] = 1 + \int_0^t \langle \xi^{y,r}_s, dB_s \rangle, \qquad u^{y,r}_t := \frac{\xi^{y,r}_t}{Z^{y,r}_t(X)} \quad \text{on } \{Z^{y,r}_t(X) > 0\}.$$

Consider

$$V(r, y) := \inf_{(\mathbb{Q}, v)} \mathbb{E}^{\mathbb{Q}} \left[\frac{1}{2} \int_0^r |v_t|^2 dt - \log z^y_r \right], \tag{9}$$

where the infimum is taken over admissible Girsanov shifts.

Then, $V(r, y) = 0$. The endpoint control $u^{y,r}$ is the unique minimizer in a class of admissible one-sided Girsanov shifts, and the measure $\mathbb{Q}^{y,r}$ satisfies

$$\mathcal{L}_{\mathbb{Q}^{y,r}}(X_r) = \mathcal{L}_{\mathbb{P}}(X_r \mid Y = y).$$

Among all admissible Girsanov shifts of equation 1 that satisfies this conditional terminal law, $u^{y,r}$ is the unique Girsanov shift with the minimum-entropy property:

$$\frac{1}{2}\mathbb{E}^{\mathbb{Q}^{y,r}}\int_0^r |u_t^{y,r}|^2 dt = \inf_{\substack{(\mathbb{Q},v):\\ \mathcal{L}_{\mathbb{Q}}(X_r)=\mathcal{L}_{\mathbb{P}}(X_r|Y=y)}} \frac{1}{2}\mathbb{E}^{\mathbb{Q}}\int_0^r |v_t|^2 dt.$$

5.2 Degenerate Nonlinear Diffusion

We consider the nonlinear diffusion

$$dX_t = b(X_t)\,dt + \sigma\,dB_t, \qquad X_0 = (1, 0, 0.4)^\top, \qquad Y = X_T^{(1)}, \tag{10}$$

where B_t is a standard three-dimensional Brownian motion and

$$b(x) = \begin{pmatrix} 20x_1(1-x_1^2) \\ -x_2 + 0.35x_1 \\ -0.8x_3 \end{pmatrix}, \qquad \sigma = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 2 & 4 \end{pmatrix}. \tag{11}$$

Since $\mathrm{rank}(\sigma) = 2 < 3$, the instantaneous covariance $\sigma\sigma^\top = \begin{pmatrix} 3 & 3 & 6 \\ 3 & 5 & 10 \\ 6 & 10 & 20 \end{pmatrix}$ is singular. The corresponding projection to $\mathrm{Ran}(\sigma^\top)$ is $P = \sigma^\dagger\sigma = \begin{pmatrix} 5/6 & 1/3 & -1/6 \\ 1/3 & 1/3 & 1/3 \\ -1/6 & 1/3 & 5/6 \end{pmatrix}$.

For each $\varepsilon \in \{0.005, 0.01, 0.1\}$, we train a separate control using the fixed-ε loss

$$\mathcal{L}_\varepsilon(\theta) = \mathbb{E}\left[\left\| Pu_\theta(t, X_t, Y) - P\frac{B_{t+\varepsilon} - B_t}{\varepsilon}\right\|^2\right]. \tag{12}$$

The training pool contains $80{,}000$ independent reference trajectories. Each model is trained for $20{,}000$ Adam updates with batch size 2048.

At evaluation time, we impose the terminal condition $Y = -1$ and simulate

$$d\widehat{X}_t = \Big[b(\widehat{X}_t) + \sigma P u_\theta(t, \widehat{X}_t, -1)\Big]\,dt + \sigma\,dB_t, \tag{13}$$

where Pu_θ is obtained from equation 12 through a neural network Tzen & Raginsky (2019). Because $X_t^{(1)}$ is autonomous, the ground-truth can be computed from the one-dimensional backward Kolmogorov equation. Writing

$$h(t, x_1) = p\Big(X_T^{(1)} = -1 \,\Big|\, X_t^{(1)} = x_1\Big), \tag{14}$$

the corresponding backward Kolmogorov equation has diffusion term $\frac{3}{2}\partial_{x_1x_1}h$. The corresponding exact Girsanov shift is

$$u^\star(t, x) = \sigma^\top \nabla_x \log h(t, x_1) = \begin{pmatrix} \partial_{x_1}\log h(t, x_1) \\ \partial_{x_1}\log h(t, x_1) \\ \partial_{x_1}\log h(t, x_1) \end{pmatrix}, \tag{15}$$

and this induces the following shift on the drift term of equation 1

$$\sigma u^\star(t, x) = (3, 3, 6)^\top \partial_{x_1}\log h(t, x_1), \tag{16}$$

which we use to compare our empirical results against the ground-truth. We evaluate each method using $15{,}000$ independently simulated trajectories. Sample trajectories obtained from this experiment can be seen on Figure 1.

5.3 Conditioned Image Generation

We next give a proof-of-concept application to image inpainting. In contrast to Algorithm 1, the terminal samples in this experiment are drawn from an empirical data law rather than from the

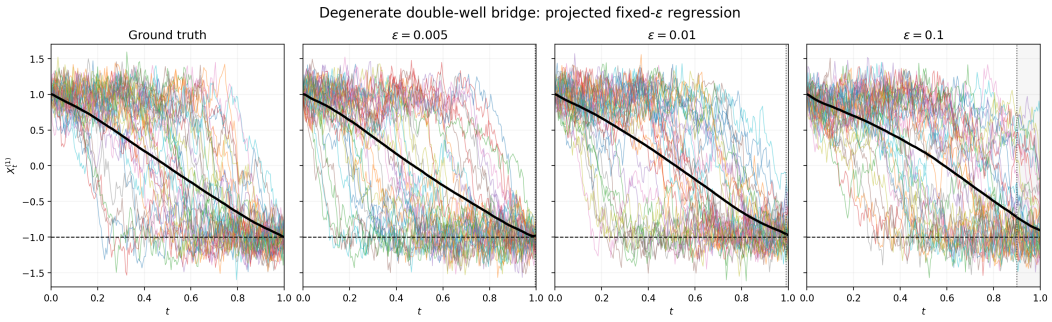


Figure 1: Sample paths from the numerical ground-truth Doob bridge and from the learned controls with $\varepsilon = 0.005$, $\varepsilon = 0.01$, and $\varepsilon = 0.1$. Thin curves represent individual trajectories, the bold black curve is the empirical mean, and the horizontal dashed line denotes the terminal target $X_T^{(1)} = -1$.

terminal law of the reference SDE. Our implementation of this experiment closely follows the one in Baker et al. (2024).

Degenerate models are natural when the variability relevant to the conditioning task is intrinsically lower-dimensional Li & He (2026). For example, in an inpainting problem in which only a localized feature is missing, one can use the accessible directions of the diffusion to represent the variability of the missing region while conditioning on the observed context.

In our numerical illustration, we train on the Shirt class of Fashion-MNIST Xiao et al. (2017), condition on the left half of a 28×28 image, and generate its missing right half. The forward model is

$$dX_t = AS\, dB_t, \qquad X_0 = x_0,$$

where the deterministic x_0 is the empirical training mean, A embeds the accessible latent directions into the image space, and S rescales these directions according to their empirical standard deviations. The latent coordinates comprise the 392 whitened visible pixels and 256 whitened principal components of the hidden half. Before training, each hidden half is projected onto these retained principal components, so the reconstructed terminal target lies in the affine support of the model. Hence, the accessible dimension is $648 < 784$, and thus every generated path remains on a fixed affine subspace and no image-space density or score function is required.

Let ν_T be the empirical law of the projected terminal images. We first sample a target $E \sim \nu_T$ and then sample an additive Gaussian bridge from x_0 to E. This defines a bridge-mixture law $\mathbb{R}^{\nu_T}$; it is not the reference law of $dX_t = AS\, dB_t$. Although ν_T is atomic and singular at T, for every observed condition y and every $r < T$, the conditional bridge-mixture marginal $\nu_r^y = \mathcal{L}_{\mathbb{R}^{\nu_T}}(X_r \mid Y = y)$ is absolutely continuous with respect to the reference marginal μ_r.

The regression expectation in this experiment is taken under $\mathbb{R}^{\nu_T}$, not under the reference path law $\mathbb{P}$. We use the finite target $\frac{B_{t+\varepsilon} - B_t}{\varepsilon}$, . This choice does not send ε to zero.

Since the diffusion is supported on a 648-dimensional affine subspace of the 784-dimensional image space, X_t is singular with respect to the Lebesgue measure and there is no available score function $\nabla_x \log p_t(x)$. Our construction requires neither such a score function nor a smooth extension of the conditional likelihood. Since the left half of the terminal image is observed, its corresponding control is available analytically. A small U-Net therefore learns only the unknown hidden component. We train its parameters using AdamW. Its inputs are the current decoded image, the observed left half, the inpainting mask, and time, and its image-shaped output is mapped onto the retained PCA coordinates of the hidden half.

At generation time, an exponential moving average of the trained network repeatedly predicts the hidden control. Each Euler–Maruyama step combines the predicted control with a new Gaussian increment in the accessible coordinates and decodes the updated coordinates back into image space. The time grid is refined near the terminal time, producing multiple right-half completions consistent with the supplied left half. Results are shown in Figure 2.

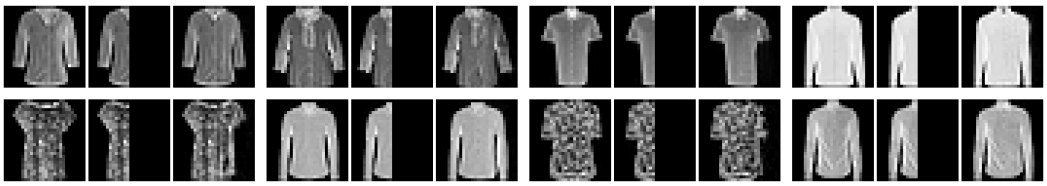


Figure 2: Conditioned image-generation results for eight held-out examples. First column is the ground truth, the second column is the given condition, and the last column is the generated image. The SDE we used for conditioned model is degenerate, cannot be "time-reverted" via the methods used in the existing literature and $\nabla_x \log p(Y = y|X_t = x)$, and σ^{-1} do not exist.

6 Predictable Representation Property and Causal Optimal Transport

The loss functions we derived rely on the predictable representation property of diffusion processes when the corresponding martingale problem is well-posed. To derive those loss functions, we extend the martingale representation result of Üstünel (2019) to a random initialization in the conditioned setting. Related constant-rank results for manifold-valued diffusions appear in Elworthy et al. (2010). However, when the rank is not constant, the associated geometric distributions need not retain the same smoothness; see (Lee, 2013, Theorem 10.34) and (Elworthy et al., 2010, Subsection 9.2).

Theorem 4. *Let (X, B) be a weak solution of equation 1 on a complete filtered probability space and b, and σ have linear growth, where B is a $(\mathcal{G}_t)$-Brownian motion and X_0 is $\mathcal{G}_0$-measurable. Let $\mu_0 = \mathcal{L}_{\mathbb{P}}(X_0)$, and suppose that for μ_0-a.e. x equation 1 with $X_0 = x$ has a weak solution that is unique in law. Then, every $F \in L^2(\mathcal{F}_T(X), \mathbb{P}; \mathbb{R})$ has a representation*

$$F = \mathbb{E}^{\mathbb{P}}[F \mid X_0] + \int_0^T \langle H_s, dB_s \rangle, \tag{17}$$

for a unique $H \in L^2_a(\mathcal{F}(X), \mathbb{P}; \mathbb{R}^d)$ satisfying $P_t(X)H_t = H_t\, dt \otimes d\mathbb{P}$-a.e. Conversely, if $g(X_0) \in L^2(\sigma(X_0), \mathbb{P})$ and H has these properties, then $g(X_0) + \int_0^T \langle H_s, dB_s \rangle$ belongs to $L^2(\mathcal{F}_T(X), \mathbb{P}; \mathbb{R})$.

Proof. See Appendix C for a complete proof. □

In particular, predictability and square integrability of the integrand is not sufficient to ensure that F is $\mathcal{F}_T(X)$ measurable. In our next result, we essentially show that the corresponding minimum entropy Girsanov shifts therefore need to be of the form σu as in equation 2, which generalizes (Üstünel, 2019, Theorem 3) when the diffusion has random initialization.

Theorem 5. *Under the hypotheses of Theorem 4, let $\mathcal{H}_t := \sigma(X_0) \vee \mathcal{F}_t(B)$, with the usual augmentation, and suppose $F \in L^2(\mathcal{H}_T, \mathbb{P}; \mathbb{R})$. Write*

$$F = \mathbb{E}^{\mathbb{P}}[F \mid X_0] + \int_0^T \langle \xi_s, dB_s \rangle, \tag{18}$$

where $\xi \in L^2_a((\mathcal{H}_t)_{t \le T}, \mathbb{P}; \mathbb{R}^d)$. Then,

$$\mathbb{E}^{\mathbb{P}}[F \mid \mathcal{F}_T(X)] = \mathbb{E}^{\mathbb{P}}[F \mid X_0] + \int_0^T \langle P_s(X)\mathbb{E}^{\mathbb{P}}[\xi_s | \mathcal{F}_s(X)], dB_s \rangle.$$

Proof. See Appendix C for a complete proof. □

The representation in equation 18 is satisfied for any such F. Using this representation, we derive a nonsmooth analogue of the Doob h-transform and connect it to causal optimal transport Lassalle (2012; 2018). In particular, using Theorem 5, we present a method to approximate the corresponding minimum entropy shift over all Girsanov shifts. As shown in Corollary 2, this naturally generalizes the Doob h-transform while preserving the desired properties.

Theorem 6. *Suppose that Assumptions 1–2 hold. Let the locally square-integrable integrand ξ be defined by*

$$Z_t(X) = Z_0(X) + \int_0^t \langle \xi_s, dB_s \rangle, \qquad P_s(X)\xi_s = \xi_s,$$

which exists and is unique by Assumption 2 and Theorem 4. Set $u_t = \xi_t / Z_t(X)$ on $\{Z_t(X) > 0\}$ and define it arbitrarily on $\{Z_t(X) = 0\}$. Then, for every $r < T$,

$$\lim_{\varepsilon \downarrow 0} \mathbb{E}^{\mathbb{P}^y} \left[\frac{1}{\varepsilon} \int_t^{t+\varepsilon} P_s(X)\, dB_s \middle| X_t \right] = u_t(X_t, y) \tag{19}$$

in $L^2([0,r] \times \Omega, dt \otimes d\mathbb{P}^y)$.

Proof. See Appendix C for a complete proof. □

Our next result verifies the minimum-energy property of u without requiring Z to be strictly positive under $dt \otimes d\mathbb{P}$, which extends (Üstünel, 2019, Theorem 7) to our setting. In particular, we verify that the limit equation 19 has the minimum entropy property, as desired.

Theorem 7. *Suppose that Assumptions 1–2 hold, fix $r < T$, and let ξ be the local integrand from Lemma 2. Set $u_t = \xi_t / Z_t(X)$ on $\{Z_t(X) > 0\}$ and $u_t = 0$ otherwise. Define $H_0^y := D_{\mathrm{KL}}\big(\mathbb{P}^y|_{\mathcal{F}_0(X)} \| \mathbb{P}|_{\mathcal{F}_0(X)}\big)$. Let $\mathcal{A}_r^y$ be the class of pairs $(\mathbb{Q}, v)$ such that $\mathbb{Q} \ll \mathbb{P}$ on $\mathcal{G}_r$,*

$$\frac{d(\mathbb{Q}|_{\mathcal{G}_0})}{d(\mathbb{P}|_{\mathcal{G}_0})} = Z_0(X),$$

v is $(\mathcal{G}_t)$-predictable, $B_t^{\mathbb{Q}} := B_t - \int_0^t v_s\, ds$ is a $(\mathcal{G}_t, \mathbb{Q})$-Brownian motion, $\mathbb{E}^{\mathbb{Q}} \int_0^r |v_t|^2 dt < \infty$, and

$$D_{\mathrm{KL}}(\mathbb{Q}\|\mathbb{P}) = H_0^y + \frac{1}{2}\mathbb{E}^{\mathbb{Q}} \int_0^r |v_t|^2 dt.$$

For every $(\mathbb{Q}, v) \in \mathcal{A}_r^y$ such that v is $\mathcal{F}(X)$ adapted and the equation

$$dX_t = \big(b(t, X_t) + \sigma(t, X_t)v_t\big)dt + \sigma(t, X_t)dB_t^{\mathbb{Q}} \tag{20}$$

has the law $\mathcal{L}_{\mathbb{P}}(X_{[0,r]}|Y = y)$,

$$\mathbb{E}^{\mathbb{Q}} \int_0^r |v_t|^2 dt \geq \mathbb{E}^{\mathbb{P}^y} \int_0^r |u_t|^2 dt. \tag{21}$$

The pair $(\mathbb{Q}^{y,r}, u)$, defined by $d(\mathbb{Q}^{y,r}|_{\mathcal{G}_r})/d(\mathbb{P}|_{\mathcal{G}_r}) = Z_r(X)$, belongs to $\mathcal{A}_r^y$ and attains equality. Equality holds for equation 21 if, and only if, $\mathbb{Q} = \mathbb{Q}^{y,r}$ and $(v_t)_t = (u_t)_t$ $dt \otimes d\mathbb{Q}^{y,r}$-a.e.

Proof. See Appendix C for a complete proof. □

Remark 5. *The preceding theorem compares Girsanov shifts on a fixed probability space and therefore does not require uniqueness of the controlled equation. Weak uniqueness of equation 2 is needed only to transfer the law $\mathcal{L}_{\mathbb{Q}^{y,r}}(X)$ to every independently constructed controlled solution through Girsanov's theorem; see Remark 3. Assumption 3 is not needed to compare the relative entropy of these transformations; it is imposed in the regression results only to express the Girsanov shift as a function of (t, X_t, y) under the probability measure $\mathbb{P}^y$.*

7 Conclusion

In this work, we have introduced regression losses for computing minimum-entropy adapted shifts when the diffusion is degenerate and the conditional law may be nonsmooth or singular. The construction uses the predictable representation property of the diffusion and does not require a score function or an invertible diffusion coefficient. Under mild conditions, we have demonstrated that the obtained controls satisfy the minimum-entropy in a class of "admissible controls".

Extensions to manifold-valued and infinite-dimensional diffusions are natural directions for future work.

REFERENCES

Zeyuan Allen-Zhu, Yuanzhi Li, and Zhao Song. A convergence theory for deep learning via over-parameterization. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 242–252, 2019.

Jürgen Amendinger. Martingale representation theorems for initially enlarged filtrations. *Stochastic Processes and Their Applications*, 89(1):101–116, 2000.

Elizabeth L Baker, Gefan Yang, Michael L Severinsen, Christy A Hipsley, and Stefan Sommer. Conditioning non-linear and infinite-dimensional diffusion processes. In *Advances in Neural Information Processing Systems*, volume 37, pp. 10801–10826, 2024.

Fabrice Baudoin. Conditioned stochastic differential equations: theory, examples and application to finance. *Stochastic Process. Appl.*, 100:109–145, 2002.

Denis Blessing, Julius Berner, Lorenz Richter, and Gerhard Neumann. Underdamped diffusion bridges with applications to sampling. In Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu (eds.), *International Conference on Learning Representations*, volume 2025, pp. 2970–3002, 2025.

Peter G Bolhuis, David Chandler, Christoph Dellago, and Phillip L Geissler. Transition path sampling: Throwing ropes over rough mountain passes, in the dark. *Annual Review of Physical Chemistry*, 53(1):291–318, 2002.

Michelle Boué and Paul Dupuis. A variational representation for certain functionals of Brownian motion. *The Annals of Probability*, 26(4):1641–1659, 1998.

Giovanni Conforti, Alain Durmus, and Marta Gentiloni Silveri. KL convergence guarantees for score diffusion models under minimal data assumptions. *SIAM Journal on Mathematics of Data Science*, 7(1):86–109, 2025.

Valentin De Bortoli. Convergence of denoising diffusion models under the manifold hypothesis. *Transactions on Machine Learning Research*, 2022. ISSN 2835-8856.

Valentin De Bortoli, Emile Mathieu, Michael Hutchinson, James Thornton, Yee Whye Teh, and Arnaud Doucet. Riemannian score-based generative modelling. In *Advances in Neural Information Processing Systems*, volume 35, pp. 2406–2422, 2022.

Tim Dockhorn, Arash Vahdat, and Karsten Kreis. Score-based generative modeling with critically-damped Langevin diffusion. *arXiv preprint arXiv:2112.07068*, 2021.

Carles Domingo i Enrich, Michal Drozdzal, Brian Karrer, and Ricky TQ Chen. Adjoint matching: Fine-tuning flow and diffusion generative models with memoryless stochastic optimal control. In *International Conference on Learning Representations*, volume 2025, pp. 53791–53846, 2025.

Ola Elerian, Siddhartha Chib, and Neil Shephard. Likelihood inference for discretely observed nonlinear diffusions. *Econometrica*, 69(4):959–993, 2001.

K David Elworthy, Yves Le Jan, and Xue-Mei Li. *The Geometry of Filtering*. Springer Science & Business Media, 2010.

Martin Hairer. Advanced stochastic analysis. *Lecture Notes, https://www. hairer. org*, 2026.

Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, second edition, 2013.

Nobuyuki Ikeda and Shinzo Watanabe. *Stochastic Differential Equations and Diffusion Processes*, volume 24. Elsevier, 2014.

Peter E. Kloeden and Eckhard Platen. *Numerical Solution of Stochastic Differential Equations*, volume 23 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1992.

Hui-Hsiung Kuo. *Gaussian Measures in Banach Spaces*. Springer, 2006.

Thomas Kurtz. The Yamada-Watanabe-Engelbert theorem for general stochastic equations and inequalities. *Electron. J. Probab.*, 12, 2007.

Rémi Lassalle. Invertibility of adapted perturbations of the identity on abstract Wiener space. *Journal of Functional Analysis*, 262(6):2734–2776, 2012.

Rémi Lassalle. Causal transport plans and their Monge–Kantorovich problems. *Stochastic Analysis and Applications*, 36(3):452–484, 2018.

John M Lee. *Introduction to Smooth Manifolds*. Springer, 2013.

Tianhong Li and Kaiming He. Back to basics: Let denoising generative models denoise. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 36115–36125, 2026.

Ehsan Mirafzali, Utkarsh Gupta, Patrick Wyrod, Frank Proske, Daniele Venturi, and Razvan Marinescu. Malliavin calculus for score-based diffusion models. *arXiv preprint arXiv:2503.16917*, 2025.

Ashkan Nikeghbali. An essay on the general theory of stochastic processes. *Probab. Surv.*, 3: 345–412, 2006.

Jakiw Pidstrigach, Elizabeth Louise Baker, Carles Domingo-Enrich, George Deligiannidis, and Nikolas Nüsken. Conditioning diffusions using Malliavin calculus. In *Forty-second International Conference on Machine Learning*, 2025.

Daniel Revuz and Marc Yor. *Continuous Martingales and Brownian Motion*. Springer Science & Business Media, 2013.

Daniel W Stroock and Marc Yor. On extremal solutions of martingale problems. *Annales Scientifiques de l'École Normale Supérieure*, 13(1):95–164, 1980.

Belinda Tzen and Maxim Raginsky. Theoretical guarantees for sampling and inference in generative models with latent diffusions. In *Conference on Learning Theory*, pp. 3084–3114. PMLR, 2019.

A. S. Üstünel. Martingale representation for degenerate diffusions. *J. Funct. Anal.*, 276(11):3468–3483, 2019.

Jin Wang, Kun Zhang, Li Xu, and Erkang Wang. Quantifying the Waddington landscape and biological paths for development and differentiation. *Proceedings of the National Academy of Sciences*, 108(20):8257–8262, 2011.

Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*, 2017.

Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 3813–3824. IEEE, 2023.

A A Primer on Stochastic Analysis

We recall the notions of weak and strong well-posedness used in the paper.

Definition 1 (Brownian motion)**.** *Let $(\Omega, \mathcal{F}, (\mathcal{G}_t)_t, \mathbb{P})$ be a complete filtered probability space. A d-dimensional Brownian motion is a $(\mathcal{G}_t)$-adapted process B with continuous paths such that*

1. *$B_0 = 0$ $\mathbb{P}$-a.s.;*
2. *for every $0 \leq s < t$, the increment $B_t - B_s$ is independent of $\mathcal{G}_s$;*
3. *for every $0 \leq s < t$, $B_t - B_s \sim \mathcal{N}(0, (t-s)I_d)$.*

Definition 2 (Weak solution)**.** *A weak solution of equation 1 consists of a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{G}_t)_t, \mathbb{P})$, a $(\mathcal{G}_t)$-Brownian motion B, and a continuous adapted process X with the prescribed initial law such that*

$$X_t = X_0 + \int_0^t b(s, X_s)\, ds + \int_0^t \sigma(s, X_s)\, dB_s \quad dt \otimes \mathbb{P}\text{-a.s.}$$

Remark 6. *In a weak solution, X need not be adapted to the filtration generated by B alone; the filtration generated by B may contain additional randomness.*

Definition 3 (Weak uniqueness)**.** *Equation 1 has* weak uniqueness*, or uniqueness in law, if any two weak solutions with the same initial law induce the same law for the state process. When needed, joint uniqueness refers to the law of (X, B).*

Definition 4 (Strong solution)**.** *A solution is* strong *if X is adapted to the usual augmentation of $\sigma(X_0) \vee \mathcal{F}_t(B)$; equivalently, it is constructed from the given initial variable and driving Brownian motion without additional randomness.*

Definition 5 (Pathwise uniqueness)**.** *Equation 1 has* pathwise uniqueness *if any two solutions on the same filtered probability space, driven by the same Brownian motion and with the same initial variable, are indistinguishable.*

Theorem 8 (Yamada-Watanabe)**.** *A weak solution exists and pathwise uniqueness holds if, and only if, a strong solution exists and uniqueness in law holds.*

The classical statement is for deterministic initial conditions; extensions to more general stochastic equations are discussed in Kurtz (2007) and the references therein.

In general, weak uniqueness need not imply pathwise uniqueness or strong existence. A standard example is Tanaka's equation

$$dX_t = \operatorname{sign}(X_t) dB_t.$$

Another example is Tsirelson's equation, which instead has a nonanticipative path-dependent drift, and has the following form

$$dX_t = b(t, X_{[0,t]})\, dt + dB_t.$$

Our martingale representation theorem requires weak existence and uniqueness in law for almost every deterministic initial state x under the conditioned measure $\mathbb{P}(\cdot | X_0 = x)$; thus, it is applicable for the SDEs above although they do not admit a strong solution.

Definition 6. *Let $(\mathcal{F}_t^0)_{t\geq 0}$ be a filtration and let $\mathcal{N}^{\mathbb{P}}$ denote the collection of all subsets of $\mathbb{P}$-null sets in $\mathcal{F}_\infty^0$. The* $\mathbb{P}$-usual augmentation *of $(\mathcal{F}_t^0)_{t\geq 0}$ is the filtration $(\mathcal{F}_t^{\mathbb{P}})_{t\geq 0}$ defined by*

$$\mathcal{F}_t^{\mathbb{P}} := \bigcap_{s>t} \left(\mathcal{F}_s^0 \vee \mathcal{N}^{\mathbb{P}}\right), \qquad t \geq 0.$$

It is the smallest right-continuous and $\mathbb{P}$-complete filtration containing $(\mathcal{F}_t^0)_{t\geq 0}$.

Definition 7 (Girsanov shift)**.** *Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ support a d-dimensional Brownian motion B, and let $\mathbb{Q} \ll \mathbb{P}$ on $\mathcal{F}_T$ with density process*

$$Z_t := \mathbb{E}^{\mathbb{P}}\left[\left.\frac{d\mathbb{Q}}{d\mathbb{P}}\right| \mathcal{F}_t\right].$$

An $(\mathcal{F}_t)$-progressively measurable process $u = (u_t)_{t\in[0,T]}$ is called a Girsanov shift *of $\mathbb{Q}$ relative to $\mathbb{P}$ if*

$$\int_0^T |u_t|^2\, dt < \infty \qquad \mathbb{Q}\text{-a.s., and } B_t^{\mathbb{Q}} := B_t - \int_0^t u_s\, ds, \qquad 0 \leq t \leq T,$$

is a Brownian motion under $\mathbb{Q}$.

B A Discussion of the Tweedie Formulae

We first recall the Kunita–Watanabe decomposition, which explains the potential orthogonal martingale component when the diffusion is degenerate.

Theorem 9 (Kunita–Watanabe decomposition). *Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ satisfy the usual conditions. Let $M = (M^1, \dots, M^d)$ be a continuous square-integrable martingale and let N be a square-integrable martingale. Then, there exist a predictable M-integrable process H and a square-integrable martingale L such that*

$$N_t = N_0 + \int_0^t \langle H_s, dM_s \rangle + L_t,$$

with

$$\langle L, M^i \rangle_t = 0, \qquad i = 1, \dots, d, \quad t \geq 0.$$

The decomposition is unique up to the usual $d\langle M \rangle \otimes d\mathbb{P}$ equivalence for H and indistinguishability for L, where $\langle A, B \rangle_t$ denotes the quadratic variation between A and B.

When $(\mathcal{F}_t)_t$ is a Brownian filtration, the Brownian martingale-representation theorem of Itô forces $L \equiv 0$ when $M = B$. For a degenerate equation, however, the filtration $(\mathcal{F}_t(X))_t$ generated by X can be strictly smaller than the filtration generated by the driving Brownian motion B. It is therefore not obvious that the orthogonal term L vanishes after conditioning. The martingale representation result for degenerate diffusions shows that the relevant Brownian integrand is unique; its image $\sigma(t, X_t)u_t$ is the added state drift in equation 2, where $(u_t)_t$ is $(\mathcal{F}_t(X))_t$-adapted. Thus, adaptiveness of $(u_t)_t$ is not sufficient for a stochastic integral of the form $\int_0^{\cdot} \langle u_s, dB_s \rangle$ to be $(\mathcal{F}_t(X))_t$-adapted.

C Proofs of the Theorems in Section 6

We first show that we can choose the projection $(P_t(X))_t$ as a Borel function of the coefficient. This choice is independent of any completion of the filtration generated by the state-process X that solves equation 1 and can therefore be used under both $\mathbb{P}$ and the conditional measures $\mathbb{P}_x = \mathbb{P}(\cdot \mid X_0 = x)$. This distinction is important, because in the original results of Üstünel (2019), the process $(P_t(X))_t$ is chosen via a measurable selection theorem; thus, it is not implementable in practice.

C.1 Auxiliary Results

Lemma 1. *Let $(\mathcal{G}_t)_t$ be a filtration. Let $K = (K_t)$ be a $(\mathcal{G}_t)_t$-predictable process. Let $z(\cdot, \cdot) : [0, T] \times \mathbb{R}^n \to \mathbb{R}^{n \times d}$ be Borel measurable. Define*

$$\mathfrak{q}_t(\cdot) := z(t, K_t(\cdot))^\top (z(t, K_t(\cdot)) z(t, K_t(\cdot))^\top)^\dagger z(t, K_t(\cdot)).$$

Then, $\mathfrak{q} = (\mathfrak{q}_t)_t$ is $(\mathcal{G}_t)$-predictable. Moreover, $\mathfrak{q}_t(\cdot)$ is the orthogonal projection onto $\mathrm{Ran}(z(t, K_t(\cdot))^\top) = (\ker z(t, K_t(\cdot)))^\perp$.

Proof. For $a \in \mathbb{R}^{n \times d}$, recall that $\Pi(a) := a^\top (aa^\top)^\dagger a$, where $(aa^\top)^\dagger$ is the Moore–Penrose inverse of $aa^\top$ (Horn & Johnson, 2013, 7.3P7). By (Horn & Johnson, 2013, 7.3P15),

$$\Pi(a) = \lim_{k \to \infty} a^\top \left(aa^\top + k^{-1} I_n \right)^{-1} a,$$

and for each k, the map on the right-hand side is continuous in a. Consequently, Π is Borel measurable. Thus, $\mathfrak{q} = (\Pi(z(t, K_t)))_t$ is predictable (Revuz & Yor, 2013, Chapter IX, Proposition 1.1). The projection property follows directly from the Moore–Penrose pseudoinverse identities (Horn & Johnson, 2013, 7.3P7). □

Remark 7. *Since $(t, \omega) \mapsto (t, X_t(\omega))$ is predictable and σ is Borel, $(\sigma(t, X_t))_t$ is predictable. The projection process $P_\cdot(X)$ need not be continuous when the rank of σ is not constant, but predictability is sufficient to have well-defined stochastic integrals. We also remark that, we did not assume that $(\mathcal{G}_t)_t$ is augmented.*

Second, for completeness, we show that $(Z_t(X))_t$ is a martingale.

Lemma 2. *Suppose that Assumptions 1–2 hold. For $\mathcal{L}_{\mathbb{P}}(Y)$-a.e. y, the process*

$$Z_t(X) := \frac{d(\mathbb{P}^y|_{\mathcal{F}_t(X)})}{d(\mathbb{P}|_{\mathcal{F}_t(X)})}, \qquad 0 \le t < T,$$

is a nonnegative $(\mathcal{F}_t(X), \mathbb{P})$-martingale. Moreover, for every $R < T$, it admits a continuous version of the form

$$Z_t(X) = Z_0(X) + \int_0^t \langle \xi_s, dB_s \rangle, \qquad 0 \le t \le R,$$

where $P_s(X)\xi_s = \xi_s\, dt \otimes d\mathbb{P}$-a.e., and

$$Z_0(X) = \frac{d(\mathbb{P}^y|_{\mathcal{F}_0(X)})}{d(\mathbb{P}|_{\mathcal{F}_0(X)})}.$$

Proof. For $0 \le s \le t < T$ and $A \in \mathcal{F}_s(X)$, using the tower property and Assumption 2, we obtain

$$\mathbb{E}^{\mathbb{P}}\left[Z_t(X)\mathbf{1}_A\right] = \mathbb{P}^y(A) = \mathbb{E}^{\mathbb{P}}\left[Z_s(X)\mathbf{1}_A\right].$$

Hence

$$\mathbb{E}^{\mathbb{P}}\left[Z_t(X) \mid \mathcal{F}_s(X)\right] = Z_s(X) \qquad \mathbb{P}\text{-a.s.},$$

and thus Z is a nonnegative $(\mathcal{F}_t(X), \mathbb{P})$-martingale.

Fix $R < T$. By Assumption 2, $Z_R(X) \in L^2(\mathcal{F}_R(X))$. Therefore, Theorem 4 gives an $\mathcal{F}(X)$-predictable process $\xi^R \in L^2_a(\mathcal{F}(X), \mathbb{P}; \mathbb{R}^d)$ such that

$$Z_R(X) = \mathbb{E}^{\mathbb{P}}[Z_R(X) \mid X_0] + \int_0^R \langle \xi^R_s, dB_s \rangle.$$

By the martingale property,

$$\mathbb{E}^{\mathbb{P}}[Z_R(X) \mid X_0] = Z_0(X).$$

The stochastic integral is an $\mathcal{F}(X)$-martingale. Taking conditional expectation with respect to $\mathcal{F}_t(X)$ therefore yields

$$Z_t(X) = Z_0(X) + \int_0^t \langle \xi^R_s, dB_s \rangle, \qquad 0 \le t \le R.$$

The right-hand side is continuous. Uniqueness of the integrand shows that these representations agree on overlapping intervals, and thus they define a single locally square-integrable process ξ on $[0, T)$. □

C.2 Proof of Theorem 4

Conditional versions of (Üstünel, 2019, Theorems 1 and 2), even under sufficiently regular assumptions, require significant technical care because the projection process $(P_t(X_t))_t$ need not be continuous. In particular, switching between probability measures changes the corresponding completed filtrations and therefore requires care when passing to enlargements of the filtration generated by the diffusion process. In the proof of Theorem 4, for this reason we use predictable σ-field associated with the raw filtration generated by X to obtain the desired predictable representation property of X.

The proof is organized into several parts. We start from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that a weak solution (X, B) to equation 1 lies on. Then, we show that (X, B) is also a solution to equation 1 under the conditioned measure $\mathbb{P}_x(\,\cdot\,) = \mathbb{P}(\,\cdot \mid X_0 = x)$. This requires B to be a Brownian motion under $\mathbb{P}_x$ too. Proofs of these statements are relatively standard; however, since we were not able to find a source for these statements, we provide complete proofs.

Lemma 3. *Let (X, B) be defined on $(\Omega, \mathcal{F}, (\mathcal{G}_t)_{t\le T}, \mathbb{P})$, where B is a d-dimensional $(\mathcal{G}_t, \mathbb{P})$-Brownian motion, X is continuous and $(\mathcal{G}_t)$-adapted, and X_0 is $\mathcal{G}_0$-measurable. Let $\mu_0 := \mathcal{L}_{\mathbb{P}}(X_0)$, and let $(\mathbb{P}_x)_{x\in\mathbb{R}^n}$ be a regular conditional probability of $\mathbb{P}$ given X_0. Define the raw σ-algebra*

$$\mathcal{G}^0_t := \sigma(X_s, B_s : 0 \le s \le t), \tag{22}$$

and let $(\mathcal{G}^x_t)_{t\le T}$ be the $\mathbb{P}_x$-usual augmentation of $(\mathcal{G}^0_t)_{t\le T}$. Then, for μ_0-a.e. x, B is a d-dimensional $(\mathcal{G}^x_t, \mathbb{P}_x)$-Brownian motion.

Proof. Since X and B are $(\mathcal{G}_t)$-adapted, $\mathcal{G}_t^0 \subseteq \mathcal{G}_t$ for every $t \le T$. Hence, under $\mathbb{P}$, every increment $B_t - B_s$ is independent of $\mathcal{G}_s^0$.

For each rational $s \in [0,T]$, choose a countable π-system $\mathcal{C}_s$ generating $\mathcal{G}_s^0$. Such a class may be constructed from finite-dimensional cylinder events of (X_q, B_q), $q \in \mathbb{Q} \cap [0,s]$, since X and B have continuous paths Kuo (2006). Fix rational $0 \le s < t \le T$, $A \in \mathcal{C}_s$, and $\theta \in \mathbb{Q}^d$. For every bounded Borel function $f : \mathbb{R}^n \to \mathbb{R}$, the random variable $f(X_0)\mathbf{1}_A$ is $\mathcal{G}_s$-measurable. Therefore,

$$\mathbb{E}^{\mathbb{P}}\left[f(X_0)\mathbf{1}_A e^{i\langle\theta, B_t - B_s\rangle}\right] = e^{-\frac{1}{2}|\theta|^2(t-s)}\mathbb{E}^{\mathbb{P}}\left[f(X_0)\mathbf{1}_A\right]. \tag{23}$$

Disintegrating with respect to X_0 yields

$$\int_{\mathbb{R}^n} f(x)\mathbb{E}^{\mathbb{P}_x}\left[\mathbf{1}_A e^{i\langle\theta, B_t - B_s\rangle}\right]\mu_0(dx) = \int_{\mathbb{R}^n} f(x)e^{-\frac{1}{2}|\theta|^2(t-s)}\mathbb{P}_x(A)\,\mu_0(dx). \tag{24}$$

Since this holds for every bounded Borel f,

$$\mathbb{E}^{\mathbb{P}_x}\left[\mathbf{1}_A e^{i\langle\theta, B_t - B_s\rangle}\right] = e^{-\frac{1}{2}|\theta|^2(t-s)}\mathbb{P}_x(A) \tag{25}$$

for μ_0-a.e. x.

There are only countably many choices of

$$(s,t,A,\theta) \in \mathbb{Q}\cap[0,T] \times \mathbb{Q}\cap[0,T] \times \mathcal{C}_s \times \mathbb{Q}^d. \tag{26}$$

Consequently, there exists a single μ_0-null set N such that, for every $x \notin N$, all of the preceding identities hold simultaneously. Enlarging N if necessary, disintegration of the $\mathbb{P}$-a.s. events $\{B_0 = 0\}$ and $\{B \text{ has continuous paths}\}$ also gives, for every $x \notin N$, $\mathbb{P}_x(B_0 = 0) = 1$ and B has continuous paths $\mathbb{P}_x$-a.s.

Fix $x \notin N$. For rational $s < t$, the monotone class theorem extends the preceding identity from $A \in \mathcal{C}_s$ to every $A \in \mathcal{G}_s^0$. Continuity in θ then extends it from $\theta \in \mathbb{Q}^d$ to every $\theta \in \mathbb{R}^d$. Thus,

$$\mathbb{E}^{\mathbb{P}_x}\left[\mathbf{1}_A e^{i\langle\theta, B_t - B_s\rangle}\right] = \mathbb{P}_x(A)e^{-\frac{1}{2}|\theta|^2(t-s)}, \qquad A \in \mathcal{G}_s^0. \tag{27}$$

By uniqueness of characteristic functions, $B_t - B_s \sim \mathcal{N}(0,(t-s)I_d)$ under $\mathbb{P}_x$, and $B_t - B_s$ is independent of $\mathcal{G}_s^0$.

Now let $0 \le s < t \le T$ be arbitrary. Choose rational sequences $s_k \downarrow s$ and $t_k \to t$ such that $s_k < t_k$. If $A \in \mathcal{G}_s^0$, then $A \in \mathcal{G}_{s_k}^0$, and hence

$$\mathbb{E}^{\mathbb{P}_x}\left[\mathbf{1}_A e^{i\langle\theta, B_{t_k} - B_{s_k}\rangle}\right] = \mathbb{P}_x(A)e^{-\frac{1}{2}|\theta|^2(t_k - s_k)}. \tag{28}$$

By continuity of B and dominated convergence, letting $k \to \infty$ gives

$$\mathbb{E}^{\mathbb{P}_x}\left[\mathbf{1}_A e^{i\langle\theta, B_t - B_s\rangle}\right] = \mathbb{P}_x(A)e^{-\frac{1}{2}|\theta|^2(t-s)}. \tag{29}$$

Thus $B_t - B_s$ is independent of $\mathcal{G}_s^0$ and has law $\mathcal{N}(0,(t-s)I_d)$ for every $0 \le s < t \le T$.

Finally, these properties are preserved under the $\mathbb{P}_x$-usual augmentation of the raw filtration. Therefore, B is a d-dimensional $(\mathcal{G}_t^x, \mathbb{P}_x)$-Brownian motion. □

Next, we show that a solution of (X,B) of equation 1 under $\mathbb{P}$ is also a solution under $\mathbb{P}_x$. This essentially amounts to picking predictable approximations of (X,B) under the raw filtration, and showing that these approximations also work under $\mathbb{P}_x$

Lemma 4. *Suppose, in addition to the hypotheses of Lemma 3, that $b : [0,T]\times\mathbb{R}^n \to \mathbb{R}^n$ and $\sigma : [0,T]\times\mathbb{R}^n \to \mathbb{R}^{n\times d}$ are bounded Borel functions and that*

$$X_t = X_0 + \int_0^t b(s,X_s)\,ds + \int_0^t \sigma(s,X_s)\,dB_s, \qquad 0 \le t \le T, \tag{30}$$

holds $\mathbb{P}$-a.s. Then, for μ_0-a.e. x,

$$\mathbb{P}_x(X_0 = x) = 1,$$

and, with respect to the filtration $(\mathcal{G}_t^x)_{t\le T}$*,*

$$X_t = x + \int_0^t b(s, X_s)\, ds + \int_0^t \sigma(s, X_s)\, dB_s, \qquad 0 \le t \le T, \quad \mathbb{P}_x\text{-a.s.} \tag{31}$$

Consequently, for μ_0*-a.e.* x*,* (X, B) *is a weak solution of equation 1 with deterministic initial condition* $X_0 = x$ *under* $\mathbb{P}_x$*.*

Proof. We first identify the initial condition. Since $(\mathbb{P}_x)_x$ is a regular conditional probability of $\mathbb{P}$ given X_0, for every bounded Borel function $f : \mathbb{R}^n \to \mathbb{R}$, $\mathbb{E}^{\mathbb{P}_x}[f(X_0)] = f(x)$ for μ_0-a.e. x. Taking f from a countable determining class for $\mathcal{B}(\mathbb{R}^n)$ and removing a single μ_0-null set gives $\mathcal{L}_{\mathbb{P}_x}(X_0) = \delta_x$, and hence $\mathbb{P}_x(X_0 = x) = 1$ for μ_0-a.e. x.

It remains to verify that the stochastic integral appearing in the original $\mathbb{P}$-SDE agrees, on almost every fiber, with the Itô integral defined under $\mathbb{P}_x$. Since X is continuous and adapted to the raw filtration $(\mathcal{G}_t^0)_t$, the process $\sigma = (\sigma(t, X_t))_t$ is $\mathcal{G}^0$-predictable (Revuz & Yor, 2013, Chapter IX, Proposition 1.1). Since σ is bounded, there exists a sequence of bounded elementary $\mathcal{G}^0$-predictable matrix-valued processes $(\sigma^m)_{m\ge1}$ such that, after passing to a subsequence,

$$\sum_{m=1}^{\infty} \mathbb{E}^{\mathbb{P}} \int_0^T \|\sigma^m(s, X_s) - \sigma(s, X_s)\|_{\mathrm{F}}^2\, ds < \infty. \tag{32}$$

For each m, define $I_t^m := \int_0^t \sigma^m(s, X_s)\, dB_s$. Since σ^m is an elementary predictable process, I_t^m is given by a finite sum of the form

$$I_t^m = \sum_j H_j^m \big(B_{t\wedge t_{j+1}} - B_{t\wedge t_j}\big),$$

and therefore represents the same random variable whether it is viewed under $\mathbb{P}$ or under $\mathbb{P}_x$.

Let

$$I_t := \int_0^t \sigma(s, X_s)\, dB_s$$

denote the Itô integral under $\mathbb{P}$. By Itô's isometry and equation 32,

$$\sum_{m=1}^{\infty} \mathbb{E}^{\mathbb{P}} |I_t^m - I_t|^2 \le \sum_{m=1}^{\infty} \mathbb{E}^{\mathbb{P}} \int_0^T \|\sigma_s^m - \sigma_s\|_{\mathrm{F}}^2\, ds < \infty \tag{33}$$

for every $t \le T$.

Disintegrating equation 32 with respect to X_0 and using Tonelli's theorem gives

$$\int_{\mathbb{R}^n} \sum_{m=1}^{\infty} \mathbb{E}^{\mathbb{P}_x} \int_0^T \|\sigma^m(s, X_s) - \sigma(s, X_s)\|_{\mathrm{F}}^2\, ds\, \mu_0(dx) = \sum_{m=1}^{\infty} \mathbb{E}^{\mathbb{P}} \int_0^T \|\sigma^m(s, X_s) - \sigma(s, X_s)\|_{\mathrm{F}}^2\, ds < \infty. \tag{34}$$

Hence, outside a μ_0-null set,

$$\mathbb{E}^{\mathbb{P}_x} \int_0^T \|\sigma^m(s, X_s) - \sigma(s, X_s)\|_{\mathrm{F}}^2\, ds \longrightarrow 0. \tag{35}$$

By Lemma 3, after enlarging the exceptional set if necessary, B is a $(\mathcal{G}_t^x, \mathbb{P}_x)$-Brownian motion. Therefore, σ is square-integrable and predictable under $\mathbb{P}_x$. Itô's isometry under $\mathbb{P}_x$ and equation 35 imply

$$I_t^m \longrightarrow I_t^x := \int_0^t \sigma_s\, dB_s \quad \text{in } L^2(\mathbb{P}_x) \tag{36}$$

for every $t \le T$.

We now identify I_t^x with the stochastic integral I_t appearing in the original $\mathbb{P}$-SDE. Fix $t \in \mathbb{Q}\cap[0, T]$. Disintegrating equation 33 yields

$$\int_{\mathbb{R}^n} \sum_{m=1}^{\infty} \mathbb{E}^{\mathbb{P}_x} |I_t^m - I_t|^2\, \mu_0(dx) = \sum_{m=1}^{\infty} \mathbb{E}^{\mathbb{P}} |I_t^m - I_t|^2 < \infty. \tag{37}$$

Thus, for μ_0-a.e. x,

$$I_t^m \longrightarrow I_t \quad \text{in } L^2(\mathbb{P}_x).$$

Comparing this with equation 36 and using uniqueness of the $L^2(\mathbb{P}_x)$ limit gives

$$I_t = \int_0^t \sigma(s, X_s)\, dB_s \qquad \mathbb{P}_x\text{-a.s.} \tag{38}$$

for μ_0-a.e. x. Since $\mathbb{Q} \cap [0, T]$ is countable, a single μ_0-null set can be chosen so that equation 38 holds simultaneously for all rational $t \in [0, T]$.

For each rational $t \in [0, T]$, the original SDE identity

$$X_t = X_0 + \int_0^t b(s, X_s)\, ds + I_t$$

holds $\mathbb{P}$-a.s. Disintegrating this probability-one event with respect to X_0 and using countability of the rational t gives, outside one μ_0-null set,

$$X_t = x + \int_0^t b(s, X_s)\, ds + \int_0^t \sigma(s, X_s)\, dB_s$$

for every rational $t \in [0, T]$, $\mathbb{P}_x$-a.s., which follows from the definition of regular conditional probability directly.

Finally, X has a.s. continuous paths, $t \longmapsto \int_0^t b(s, X_s)\, ds$ is continuous, and $t \longmapsto \int_0^t \sigma(s, X_s)\, dB_s$ has a continuous version under $\mathbb{P}_x$. Hence the preceding identity extends from rational t to every $t \in [0, T]$. Since X is adapted to $(\mathcal{G}_t^x)_t$ and, by Lemma 3, B is a $(\mathcal{G}_t^x, \mathbb{P}_x)$-Brownian motion, (X, B) is a weak solution of equation 1 under $\mathbb{P}_x$ with initial condition x. □

Proof of Theorem 4. Let (X, B) be the given weak solution and let $(\mathbb{P}_x)_{x \in \mathbb{R}^n}$ be a regular conditional probability of $\mathbb{P}$ given X_0:

$$\mathbb{P}_x(\,\cdot\,) = \mathbb{P}(\,\cdot \mid X_0 = x),$$

whose existence follows from the running assumptions, see (Ikeda & Watanabe, 2014, Theorem I.3.3).

For each $x \in \mathbb{R}^n$, let $\mathcal{G}_t^0 := \sigma(X_s, B_s : 0 \le s \le t)$, and let $(\mathcal{G}_t^x)_{t \le T}$ be its $\mathbb{P}_x$-usual augmentation. By Lemma 3, for μ_0-a.e. x, B is a d-dimensional $(\mathcal{G}_t^x, \mathbb{P}_x)$-Brownian motion. Moreover, by Lemma 4, after removing a further μ_0-null set, $\mathbb{P}_x(X_0 = x) = 1$ and

$$X_t = x + \int_0^t b(s, X_s)\, ds + \int_0^t \sigma(s, X_s)\, dB_s, \qquad 0 \le t \le T, \quad \mathbb{P}_x\text{-a.s.}$$

Therefore, for μ_0-a.e. x, (X, B) is a weak solution of equation 1 with initial condition x on $(\Omega, \mathcal{F}, (\mathcal{G}_t^x)_{t \le T}, \mathbb{P}_x)$.

For any given probability measure $\mathbb{Q}$, write $\mathcal{F}^{\mathbb{Q}}(X)$ for the usual augmentation of the raw state filtration $(\mathcal{F}_t^0(X))_t$ under $\mathbb{Q}$.

We argue by orthogonality as in (Üstünel, 2019, Theorem 1), now disintegrating with respect to $\mathbb{P}_x$. Let $\mathcal{K}$ be the closed subspace of $L^2(\mathcal{F}_T^{\mathbb{P}}(X), \mathbb{P}; \mathbb{R})$ generated by random variables of the form

$$g(X_0) + \int_0^T \langle H_t\,, dB_t \rangle,$$

where $g(X_0) \in L^2(\mathcal{F}_0^{\mathbb{P}}(X))$, H is $\mathcal{F}^{\mathbb{P}}(X)$-predictable, square-integrable w.r.t. Lebesgue measure on $[0, T]$, and $H_t = P_t(X) H_t\, dt \otimes d\mathbb{P}$-a.e. It suffices to show $\mathcal{K}^\perp = \{0\}$, where $\mathcal{K}^\perp$ is the orthogonal complement of $\mathcal{K}$ under the L^2-norm. Fix $Z \in \mathcal{K}^\perp$. Choose an $\mathcal{F}_T^0(X)$-measurable representative equal to Z $\mathbb{P}$-a.s. and continue to denote it by Z. Disintegration transfers this equality to $\mathbb{P}_x$-a.s. for μ_0-a.e. x; in particular, Z is $\mathcal{F}_T^{\mathbb{P}_x}(X)$-measurable on those fibers. Since $\mathcal{F}_0^{\mathbb{P}}(X) = \sigma(X_0) \vee \mathcal{N}_{\mathbb{P}}$, orthogonality to every $g(X_0)$ gives $\mathbb{E}^{\mathbb{P}}[Z \mid X_0] = 0$.

Let $\mathscr{P}^0(X)$ denote the predictable σ-field associated with the raw filtration $(\mathcal{F}_t^0(X))_{t \in [0,T]}$ (Revuz & Yor, 2013, Chapter I, Exercise 4.20). Since the continuous path space $C([0, T]; \mathbb{R}^n)$ is Polish, the

raw state filtration $(\mathcal{F}_t(X))_t$ and its predictable σ-field are countably generated (Nikeghbali, 2006, Proposition 2.19). Hence, we may choose a countable family of bounded, $\mathbb{R}^d$-valued, $\mathscr{P}^0(X)$-measurable simple processes generated by a countable algebra (because the underlying L^2 space is separable). We define $(J^m)_{m\geq 1}$ to be an enumeration of all rational finite linear combinations of this family. Then, $(J^m)_{m\geq 1}$ itself is dense in

$$L^2\big(\mathscr{P}^0(X), dt \otimes \mathbb{Q}; \mathbb{R}^d\big)$$

for every probability measure $\mathbb{Q}$.

For $m \geq 1$, set

$$H_t^m := P_t(X) J_t^m.$$

Since X is continuous and adapted to its raw filtration $(\mathcal{F}_t^0(X))_{t\in[0,T]}$, it is $\mathcal{F}^0(X)$-predictable. Since $(t,x) \mapsto \sigma(t,x)$ is Borel measurable, $(\sigma(t,X_t))_t$ is $\mathscr{P}^0(X)$-measurable. Moreover, as argued in the proof of Lemma 1, the map

$$a \mapsto a^\top (aa^\top)^\dagger a$$

is Borel measurable. Hence $(P_t(X))_t$ is $\mathscr{P}^0(X)$-measurable (and predictable). Therefore, H^m is $\mathscr{P}^0(X)$-measurable, and hence predictable with respect to $\mathcal{F}^{\mathbb{P}}(X)$. Since J^m is bounded and $P_t(X)$ is an orthogonal projection, H^m is square-integrable. Furthermore,

$$P_t(X) H_t^m = P_t(X)^2 J_t^m = P_t(X) J_t^m = H_t^m, \qquad dt \otimes d\mathbb{P}\text{-a.e.}$$

For each m, since H^m is $\mathscr{P}^0(X)$-measurable and square-integrable, there exists a sequence of raw elementary predictable processes $(H^{m,k})_{k\geq 1}$ such that

$$\mathbb{E}^{\mathbb{P}} \int_0^T |H_t^{m,k} - H_t^m|^2 \, dt \longrightarrow 0,$$

see (Hairer, 2026, Proposition 4.6). Passing to a subsequence, we may assume that

$$\sum_{k=1}^{\infty} \mathbb{E}^{\mathbb{P}} \int_0^T |H_t^{m,k} - H_t^m|^2 \, dt < \infty.$$

Disintegration and Tonelli's theorem then show, simultaneously for all m, that $H^{m,k} \to H^m$ in $L^2(dt \otimes d\mathbb{P}_x)$ for μ_0-a.e. x. Choose the $\mathbb{P}$-version of $\int H^m \, dB$ as the $L^2(\mathbb{P})$ limit of the corresponding raw elementary integrals. Applying the same summable subsequence argument to the integral errors and using Itô's isometry under both $\mathbb{P}$ and $\mathbb{P}_x$ shows that this fixed version is also the $\mathbb{P}_x$-Itô integral for μ_0-a.e. x. Intersecting the resulting countably many full-measure sets makes this choice valid for every m below.

Hence, for every Borel set $A \subset \mathbb{R}^n$, L^2-orthogonality of Z to $\mathcal{K}$ gives

$$0 = \mathbb{E}^{\mathbb{P}}\left[Z \, 1_{\{X_0 \in A\}} \int_0^T \langle H_t^m \, , dB_t \rangle \right] = \mathbb{E}^{\mathbb{P}}\left[Z \int_0^T \langle 1_{\{X_0 \in A\}} \, H_t^m \, , dB_t \rangle \right].$$

Therefore,

$$\mathbb{E}^{\mathbb{P}}\left[Z \int_0^T \langle H_t^m \, , dB_t \rangle \,\middle|\, X_0 \right] = 0 \implies \mathbb{E}^{\mathbb{P}_x}\left[Z \int_0^T \langle H_t^m \, , dB_t \rangle \right] = 0. \tag{39}$$

Since the family $(H^m)_{m\geq 1}$ is countable, there exists a single set $N \subset \mathbb{R}^n$, with $\mu_0(N) = 0$, such that the preceding equality holds simultaneously for every $m \geq 1$ and every $x \notin N$.

Now fix $x \notin N$ such that $\mathbb{E}^{\mathbb{P}_x}[Z] = 0$, $Z \in L^2(\mathbb{P}_x)$, and the fixed-initial-condition hypotheses hold. Let $H \in L_a^2\big(\mathcal{F}^{\mathbb{P}_x}(X), \mathbb{P}_x; \mathbb{R}^d\big)$ satisfy $P_t(X) H_t = H_t \; dt \otimes d\mathbb{P}_x$-a.e., where $(P_t(X))_t$ is $\mathcal{F}^{\mathbb{P}_x}(X)$-adapted by Lemma 1. Since the predictable σ-field of the augmented filtration is the $dt \otimes \mathbb{P}_x$-completion of $\mathscr{P}^0(X)$, the process H admits a $\mathscr{P}^0(X)$-measurable representative. In particular, by density of $(J^m)_{m\geq 1}$, there exists a sequence $(J^{m_k})_{k\geq 1}$ such that

$$\mathbb{E}^{\mathbb{P}_x} \int_0^T |J_t^{m_k} - H_t|^2 \; dt \longrightarrow 0.$$

Since $P_t(X)$ is an orthogonal projection and $P_t(X)H_t = H_t\, dt \otimes d\mathbb{P}_x$-a.e.,

$$\mathbb{E}^{\mathbb{P}_x}\int_0^T |H_t^{m_k} - H_t|^2\, dt = \mathbb{E}^{\mathbb{P}_x}\int_0^T |P_t(X)J_t^{m_k} - P_t(X)H_t|^2\, dt$$
$$\leq \mathbb{E}^{\mathbb{P}_x}\int_0^T |J_t^{m_k} - H_t|^2\, dt \longrightarrow 0.$$

By Itô's isometry under $\mathbb{P}_x$,

$$\int_0^T \langle H_t^{m_k}, dB_t\rangle \longrightarrow \int_0^T \langle H_t, dB_t\rangle \qquad \text{in } L^2(\mathbb{P}_x).$$

Since $Z \in L^2(\mathbb{P}_x)$, by Cauchy-Schwarz and Itô isometry, it follows that

$$\mathbb{E}^{\mathbb{P}_x}\left[Z\int_0^T \langle H_t, dB_t\rangle\right] = \lim_{k\to\infty} \mathbb{E}^{\mathbb{P}_x}\left[Z\int_0^T \langle H_t^{m_k}, dB_t\rangle\right] = 0. \tag{40}$$

Thus Z is orthogonal under $\mathbb{P}_x$ to every stochastic integral $\int_0^T \langle H_s, dB_s\rangle$ with $H \in L_a^2(\mathcal{F}^{\mathbb{P}_x}(X), \mathbb{P}_x; \mathbb{R}^d)$ and $P_t(X)H_t = H_t\, dt \otimes d\mathbb{P}_x$-a.e.

Under $\mathbb{P}_x$, $X_0 = x$. The fixed-initial representation theorem (Üstünel, 2019, Theorem 1), together with equation 39–equation 40, therefore gives $Z = 0$ $\mathbb{P}_x$-a.s. for μ_0-a.e. x. Integrating with respect to μ_0,

$$\mathbb{P}(Z \neq 0) = \int \mathbb{P}_x(Z \neq 0)\, \mu_0(dx) = 0.$$

Thus $\mathcal{K}^\perp = \{0\}$. It remains to show that closure of $\mathcal{K}$ in L^2 consist of terms of the form

$$g(X_0) + \int_0^T \langle H_s, dB_s\rangle,$$

where $P_t(X)H_t = H_t\, dt \otimes d\mathbb{P}$-a.e.

To show this, we closely follow the proof of (Üstünel, 2019, Theorem 2).

Let $\widetilde{F} := F - \mathbb{E}^{\mathbb{P}}[F \mid X_0]$, i.e., $\mathbb{E}^{\mathbb{P}}\big[\widetilde{F} \mid X_0\big] = 0$. By the preceding density argument, there exists a sequence $(F^m)_{m\geq 1} \subset L^2(\mathcal{F}_T^{\mathbb{P}}(X), \mathbb{P})$ such that $F^m \to \widetilde{F}$ in $L^2(\mathbb{P})$, and, for every m,

$$F^m = \int_0^T \langle H_t^m, dB_t\rangle, \qquad H_t^m = P_t(X)H_t^m \quad dt \otimes d\mathbb{P}\text{-a.e.}$$

Since $F^m \to \widetilde{F}$ in $L^2(\mathbb{P})$, Itô's isometry gives

$$\lim_{m,k\to\infty} \mathbb{E}^{\mathbb{P}}\int_0^T |H_t^m - H_t^k|^2\, dt = \lim_{m,k\to\infty} \mathbb{E}^{\mathbb{P}}|F^m - F^k|^2 = 0.$$

Hence $(H^m)_{m\geq 1}$ converges in $L^2(dt \otimes d\mathbb{P}; \mathbb{R}^d)$ to some predictable process H. Since $P_t(X)$ is an orthogonal projection and $P_t(X)H_t^m = H_t^m$, we also have $P_t(X)H_t = H_t\, dt \otimes d\mathbb{P}$-a.e. Indeed,

$$\mathbb{E}^{\mathbb{P}}\int_0^T |P_t(X)H_t - H_t|^2\, dt \leq 4 \lim_{m\to\infty} \mathbb{E}^{\mathbb{P}}\int_0^T |H_t - H_t^m|^2\, dt = 0.$$

Applying Itô's isometry once more,

$$\int_0^T \langle H_t^m, dB_t\rangle \longrightarrow \int_0^T \langle H_t, dB_t\rangle \qquad \text{in } L^2(\mathbb{P}).$$

Therefore,

$$\widetilde{F} = \int_0^T \langle H_t, dB_t\rangle \qquad \mathbb{P}\text{-a.s.},$$

and consequently

$$F = \mathbb{E}^{\mathbb{P}}[F \mid X_0] + \int_0^T \langle H_t, dB_t\rangle, \qquad H_t = P_t(X)H_t \quad dt \otimes d\mathbb{P}\text{-a.e.}$$

This proves the assertion. □

C.3 Proof of Theorem 5

Proof of Theorem 5. Since X_0 is $\mathcal{G}_0$-measurable and B is a $(\mathcal{G}_t)$-Brownian motion,

$$X_0 \perp\!\!\!\perp \sigma(B_s : 0 \le s \le T).$$

Thus, the enlargement of the Brownian filtration by X_0 is independent, and the equivalence condition in (Amendinger, 2000, Theorem 4.2) is satisfied. Consequently, equation 18 follows in $(\mathcal{H}_t)_{t\le T}$; equivalently, one may apply the Brownian martingale representation theorem conditionally on X_0. For the rest of the proof, we follow the projection argument of (Üstünel, 2019, Theorem 3).

Let $\widehat{\xi}$ be the $L^2(dt \otimes d\mathbb{P})$ predictable projection of ξ onto the closed subspace of $\mathcal{F}(X)$-predictable processes (Revuz & Yor, 2013, Exercise 5.14). Then,

$$\mathbb{E}^{\mathbb{P}} \int_0^T |\widehat{\xi}_t|^2 dt \le \mathbb{E}^{\mathbb{P}} \int_0^T |\xi_t|^2 dt < \infty,$$

and, for every square-integrable $\mathcal{F}(X)$-predictable process H,

$$\mathbb{E}^{\mathbb{P}} \int_0^T \left\langle \xi_t - \widehat{\xi}_t, H_t \right\rangle dt = 0. \tag{41}$$

Define $K_t := P_t(X)\widehat{\xi}_t$. By Lemma 1, $P_t(X)$ is $\mathcal{F}(X)$-predictable. Hence K is an $\mathcal{F}(X)$-predictable process. Moreover, since $P_t(X)$ is an orthogonal projection with respect to the Euclidean metric, we have $|K_t| \le |\widehat{\xi}_t|$ $dt \otimes \mathbb{P}$-a.e., and therefore K is square-integrable. Also,

$$P_t(X)K_t = P_t(X)^2\widehat{\xi}_t = K_t \qquad dt \otimes d\mathbb{P}\text{-a.e.} \tag{42}$$

Set

$$\widetilde{F} := \mathbb{E}^{\mathbb{P}}[F \mid X_0] + \int_0^T \langle K_t, dB_t \rangle.$$

By Theorem 4, since K is $\mathcal{F}(X)$-predictable, square-integrable, and equation 42 holds, we have $\widetilde{F} \in L^2(\mathcal{F}_T(X), \mathbb{P})$. We claim that $\widetilde{F} = \mathbb{E}^{\mathbb{P}}[F \mid \mathcal{F}_T(X)]$. It is therefore enough to prove that

$$\mathbb{E}^{\mathbb{P}}[(F - \widetilde{F})Z] = 0 \qquad \text{for every } Z \in L^2(\mathcal{F}_T(X), \mathbb{P}; \mathbb{R}).$$

Fix such a Z. By Theorem 4, there exists a square-integrable $\mathcal{F}(X)$-predictable process H satisfying $P_t(X)H_t = H_t$ $dt \otimes d\mathbb{P}$-a.e. such that

$$Z = \mathbb{E}^{\mathbb{P}}[Z \mid X_0] + \int_0^T \langle H_t, dB_t \rangle. \tag{43}$$

On the other hand, by the initially enlarged-filtration Brownian martingale representation established above,

$$F = \mathbb{E}^{\mathbb{P}}[F \mid X_0] + \int_0^T \langle \xi_t, dB_t \rangle \implies F - \widetilde{F} = \int_0^T \langle \xi_t - K_t, dB_t \rangle.$$

In particular, $\mathbb{E}^{\mathbb{P}}[F - \widetilde{F} \mid X_0] = 0$ $\mathbb{P}$-a.e.

Using equation 43, we therefore obtain

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}[(F - \widetilde{F})Z] = \mathbb{E}^{\mathbb{P}}\left[(F - \widetilde{F})\mathbb{E}^{\mathbb{P}}[Z \mid X_0]\right] \\ + \mathbb{E}^{\mathbb{P}}\left[\left(\int_0^T \langle \xi_t - K_t, dB_t \rangle\right)\left(\int_0^T \langle H_t, dB_t \rangle\right)\right].\end{aligned}$$

The first term is zero by conditioning on X_0. By Itô's isometry, we have

$$\mathbb{E}^{\mathbb{P}}\left[\left(\int_0^T \langle \xi_t - K_t, dB_t \rangle\right)\left(\int_0^T \langle H_t, dB_t \rangle\right)\right] = \mathbb{E}^{\mathbb{P}} \int_0^T \langle \xi_t - K_t, H_t \rangle dt.$$

Since $K_t = P_t(X)\widehat{\xi}_t$, the symmetry of the orthogonal projection $P_t(X)$ and the invariance of H_t under $P_t(X)$ yield

$$\begin{aligned}\langle K_t, H_t\rangle &= \left\langle P_t(X)\widehat{\xi}_t, H_t\right\rangle \\ &= \left\langle \widehat{\xi}_t, P_t(X)H_t\right\rangle \\ &= \langle \widehat{\xi}_t, H_t\rangle.\end{aligned}$$

Consequently,

$$\mathbb{E}^{\mathbb{P}}[(F-\widetilde{F})Z] = \mathbb{E}^{\mathbb{P}}\int_0^T \langle \xi_t - \widehat{\xi}_t, H_t\rangle dt.$$

Since H is $\mathcal{F}(X)$-predictable, the defining orthogonality property of equation 41 gives

$$\mathbb{E}^{\mathbb{P}}[(F-\widetilde{F})Z] = 0.$$

Since this holds for every $Z \in L^2(\mathcal{F}_T(X))$ and $\widetilde{F} \in L^2(\mathcal{F}_T(X))$, the Hilbert-space characterization of conditional expectation implies $\widetilde{F} = \mathbb{E}^{\mathbb{P}}[F \mid \mathcal{F}_T(X)]$ $\mathbb{P}$-a.s. Therefore,

$$\mathbb{E}^{\mathbb{P}}[F \mid \mathcal{F}_T(X)] = \mathbb{E}^{\mathbb{P}}[F \mid X_0] + \int_0^T \left\langle P_t(X)\widehat{\xi}_t, dB_t\right\rangle.$$

This proves the result. □

C.4 Proof of Theorem 6

Proof of Theorem 6. Fix $r < T$ and choose R such that $r < R < T$. By Lemma 2, on $[0, R]$,

$$Z_t(X) = Z_0(X) + \int_0^t \langle \xi_s, dB_s\rangle, \qquad P_s(X)\xi_s = \xi_s.$$

Set

$$u_t := \frac{\xi_t}{Z_t(X)} \quad \text{on } \{Z_t(X) > 0\},$$

and set $u_t = 0$ on $\{Z_t(X) = 0\}$. This set is $dt \otimes d\mathbb{P}^y$-null because $\mathbb{P}^y(Z_t(X) = 0) = \mathbb{E}^{\mathbb{P}}[Z_t(X)\mathbf{1}_{\{Z_t(X)=0\}}] = 0$.

For $n \geq 1$, define the stopping time

$$\tau_n^R := R \wedge \inf\left\{t \geq 0 : Z_t(X) \notin \left(\frac{1}{n}, n\right)\right\},$$

with $\inf \varnothing = \infty$. On $[0, \tau_n^R]$, $1/n \leq Z_t(X) \leq n$, and

$$dZ_t(X) = Z_t(X)\langle u_t, dB_t\rangle.$$

The change of measure separates into an initial change by the $\mathcal{G}_0$-measurable density Z_0 and a dynamic change generated by u. The initial change leaves Brownian increments Brownian, since they are independent of $\mathcal{G}_0$; on $\{Z_0 > 0\}$, the normalized density Z_t/Z_0 generates the dynamic Girsanov shift. Thus, Girsanov's theorem over $(\mathcal{F}_t(X))_t$ shows that

$$M_{t\wedge\tau_n^R} := \int_0^{t\wedge\tau_n^R} P_s(X)\, dB_s - \int_0^{t\wedge\tau_n^R} u_s\, ds$$

is an $(\mathcal{F}_t(X), \mathbb{P}^y)$-local martingale. It is in fact square-integrable, since

$$\mathbb{E}^{\mathbb{P}^y}\int_0^{\tau_n^R} |u_s|^2 ds = \mathbb{E}^{\mathbb{P}}\int_0^{\tau_n^R} \frac{|\xi_s|^2}{Z_s(X)}\, ds \leq n\,\mathbb{E}^{\mathbb{P}}\int_0^R |\xi_s|^2 ds < \infty.$$

We next remove the localization on the fixed interval $[0, R]$. Since Z is a nonnegative continuous martingale, it remains zero after its first hitting time of zero. Therefore, for

$$A_R := \left\{\inf_{0\leq s\leq R} Z_s(X) = 0\right\},$$

we have $Z_R(X) = 0$ on A_R, and thus

$$\mathbb{P}^y(A_R) = \mathbb{E}^{\mathbb{P}}[Z_R(X)\mathbf{1}_{A_R}] = 0.$$

On A_R^c, continuity gives $0 < \inf_{s\le R} Z_s(X) \le \sup_{s\le R} Z_s(X) < \infty$. On A_R, the stopped values converge to zero, which equals Z_R because a nonnegative continuous martingale remains zero after its first hitting time of zero. Consequently, pointwise,

$$Z_{\tau_n^R}(X) \longrightarrow Z_R(X) \qquad \mathbb{P}\text{-a.s.},$$

and $\tau_n^R \nearrow R\,\mathbb{P}^y$-a.s.

Moreover, Itô's formula for $\log Z$ on $[0, \tau_n^R]$ gives

$$\frac{1}{2}\mathbb{E}^{\mathbb{P}^y}\int_0^{\tau_n^R} |u_s|^2 ds = \mathbb{E}^{\mathbb{P}^y}\left[\log\frac{Z_{\tau_n^R}(X)}{Z_0(X)}\right] = \mathbb{E}^{\mathbb{P}}\left[Z_{\tau_n^R}(X)\log Z_{\tau_n^R}(X)\right] - \mathbb{E}^{\mathbb{P}}\left[Z_0(X)\log Z_0(X)\right].$$

By Doob's Optional Sampling Theorem, we obtain $Z_{\tau_n^R} = \mathbb{E}^{\mathbb{P}}[Z_R \mid \mathcal{F}_{\tau_n^R}(X)]$. In particular, conditional Jensen's inequality gives

$$Z^2_{\tau_n^R} \le \mathbb{E}^{\mathbb{P}}[Z_R^2 \mid \mathcal{F}_{\tau_n^R}(X)].$$

The conditional expectations on the right form a uniformly integrable family. Hence $(Z^2_{\tau_n^R})_n$ is uniformly integrable. Since $(z\log z)^+ \le z^2$ and $(z\log z)^- \le e^{-1}$, the family $(Z_{\tau_n^R}\log Z_{\tau_n^R})_n$ is uniformly integrable as well. Together with the almost-sure convergence above, by Vitali's convergence theorem,

$$\mathbb{E}^{\mathbb{P}}[Z_{\tau_n^R}\log Z_{\tau_n^R}] \longrightarrow \mathbb{E}^{\mathbb{P}}[Z_R\log Z_R].$$

Using monotone convergence on $\int_0^{\tau_n^R}|u_s|^2ds = \int_0^{\infty}|u_s|^2 1_{\{s\le\tau_n^R\}}ds$, we obtain $\lim_{n\to\infty}\int_0^{\tau_n^R}|u_s|^2ds = \int_0^R |u_s|^2 ds$. Thus, we obtain both square integrability and the exact identity

$$\frac{1}{2}\mathbb{E}^{\mathbb{P}^y}\int_0^R |u_s|^2 ds = \mathbb{E}^{\mathbb{P}}[Z_R\log Z_R] - \mathbb{E}^{\mathbb{P}}[Z_0\log Z_0]. \tag{44}$$

Since $\tau_n^R \nearrow R\,\mathbb{P}^y$-a.s. and $M^{\tau_n^R}$ is a martingale for every n, the process

$$M_t := \int_0^t P_s(X)\,dB_s - \int_0^t u_s\,ds, \qquad 0 \le t \le R,$$

is a continuous $(\mathcal{F}_t(X), \mathbb{P}^y)$-local martingale. Its matrix quadratic variation is

$$\langle M\rangle_t = \int_0^t P_s(X)P_s(X)^\top ds = \int_0^t P_s(X)\,ds \implies \operatorname{tr}\langle M\rangle_R = \int_0^R \operatorname{tr}\big(P_s(X)\big)\,ds \le d\,R.$$

Hence, by (Revuz & Yor, 2013, Chapter IV, Proposition 1.23) (applied componentwise), M is a square-integrable $(\mathcal{F}_t(X), \mathbb{P}^y)$-martingale on $[0, R]$.

Define the change of measure $\mathbb{Q}^{y,R}$ obtained on $\mathcal{G}_R$ by

$$\frac{d(\mathbb{Q}^{y,R}|_{\mathcal{G}_R})}{d(\mathbb{P}|_{\mathcal{G}_R})} = Z_R(X).$$

The representation of Z is a square-integrable $(\mathcal{G}_t, \mathbb{P})$-martingale, and thus $Z_t(X)$ is the density process of $\mathbb{Q}^{y,R}$. Moreover, $\mathbb{Q}^{y,R}|_{\mathcal{F}_R(X)} = \mathbb{P}^y|_{\mathcal{F}_R(X)}$. The preceding localization and Girsanov's theorem show that

$$B_t^{y,R} := B_t - \int_0^t u_s\,ds, \qquad 0 \le t \le R,$$

is a $(\mathcal{G}_t, \mathbb{Q}^{y,R})$-Brownian motion. Since $\mathbb{Q}^{y,R} \ll \mathbb{P}$ on $\mathcal{G}_R$, the original SDE formulation (equation 1) is preserved. Substitution gives

$$dX_t = \big(b(t, X_t) + \sigma(t, X_t)u_t\big)dt + \sigma(t, X_t)\,dB_t^{y,R}, \qquad 0 \le t \le R. \tag{45}$$

Consequently,

$$\mathcal{L}_{\mathbb{Q}^{y,R}}(X_{[0,R]}) = \mathcal{L}_{\mathbb{P}}(X_{[0,R]} \mid Y = y).$$

For $0 < \varepsilon < R - r$ and $0 \le t \le r$, the martingale property of M together with Theorem 5 give

$$\mathbb{E}^{\mathbb{P}^y}\left[\frac{1}{\varepsilon}\int_t^{t+\varepsilon} P_s(X)\,dB_s \middle| X_t\right] = \mathbb{E}^{\mathbb{P}^y}\left[\frac{1}{\varepsilon}\int_t^{t+\varepsilon} u_s\,ds \middle| X_t\right].$$

Assumption 3 gives the $\sigma(X_t)$-measurable version $u_t = u_t(X_t, y)$. As $u \in L^2([0,R]\times\Omega, dt\otimes d\mathbb{P}^y)$, by the Lebesgue differentiation theorem,

$$\frac{1}{\varepsilon}\int_t^{t+\varepsilon} u_s\,ds \longrightarrow u_t \quad \text{in } L^2([0,r]\times\Omega, dt\otimes d\mathbb{P}^y).$$

Since u_t is $\sigma(X_t)$-measurable and conditional expectation is an L^2-contraction, the same convergence holds after conditioning. Therefore

$$\lim_{\varepsilon\downarrow 0}\mathbb{E}^{\mathbb{P}^y}\left[\frac{1}{\varepsilon}\int_t^{t+\varepsilon} P_s(X)\,dB_s \middle| X_t\right] = u_t = \frac{\xi_t}{Z_t(X)}$$

in $L^2([0,r]\times\Omega, dt\otimes d\mathbb{P}^y)$. Since $r < T$ was arbitrary, the result follows on every compact subinterval of $[0,T)$, as desired. □

C.5 Proof of Theorem 7

Proof of Theorem 7. Fix $r < T$. Throughout the proof, let $(\Omega, \mathcal{F}, (\mathcal{G}_t)_{0\le t\le r}, \mathbb{P})$ be the filtered probability space carrying a weak solution (X, B) of equation 1, where B is a $(\mathcal{G}_t)$-Brownian motion and

$$\mathcal{F}_t(X) \subseteq \mathcal{G}_t, \qquad 0 \le t \le r.$$

Thus, under $\mathbb{P}$,

$$dX_t = b(t, X_t)\,dt + \sigma(t, X_t)\,dB_t. \tag{46}$$

Let

$$Z_t(X) = \frac{d(\mathbb{P}^y|_{\mathcal{F}_t(X)})}{d(\mathbb{P}|_{\mathcal{F}_t(X)})}, \qquad u_t := \frac{\xi_t}{Z_t(X)} \quad \text{on } \{Z_t(X) > 0\}.$$

By Lemma 2 and Theorem 4, there exists a square-integrable integrand $\xi = (\xi_s)_s$ on $[0,r]$ such that

$$Z_t(X) = Z_0(X) + \int_0^t \langle \xi_s, dB_s\rangle, \qquad Z_0(X) = \frac{d(\mathbb{P}^y|_{\mathcal{F}_0(X)})}{d(\mathbb{P}|_{\mathcal{F}_0(X)})}, \qquad P_s(X)\xi_s = \xi_s.$$

In particular, since the integrand is $(\mathcal{G}_t)$-adapted, $(Z_t(X))_{0\le t\le r}$ is also a $(\mathcal{G}_t, \mathbb{P})$- martingale.

We now construct a probability measure under which the state process has the desired conditional law. Define $\mathbb{Q}^{y,r}$ on $(\Omega, \mathcal{G}_r)$ by

$$\frac{d(\mathbb{Q}^{y,r}|_{\mathcal{G}_r})}{d(\mathbb{P}|_{\mathcal{G}_r})} := Z_r(X). \tag{47}$$

Since $Z_r(X)$ is $\mathcal{F}_r(X)$-measurable and $\mathbb{E}^{\mathbb{P}}[Z_r(X)] = 1$, this defines a probability measure. Moreover, for every $A \in \mathcal{F}_r(X)$,

$$\mathbb{Q}^{y,r}(A) = \mathbb{E}^{\mathbb{P}}[Z_r(X)\mathbf{1}_A] = \mathbb{P}^y(A).$$

Hence

$$\mathbb{Q}^{y,r}|_{\mathcal{F}_r(X)} = \mathbb{P}^y|_{\mathcal{F}_r(X)}. \tag{48}$$

Furthermore, because Z is a $(\mathcal{G}_t, \mathbb{P})$-martingale with terminal value $Z_r(X)$,

$$Z_t(X) = \mathbb{E}^{\mathbb{P}}\left[Z_r(X) \mid \mathcal{G}_t\right], \qquad 0 \le t \le r.$$

Thus, Z is the Radon-Nikodym derivative of $\mathbb{Q}^{y,r}$ with respect to $\mathbb{P}$ on $(\mathcal{G}_t)_t$.

As proved by the localization argument in Theorem 6, under $\mathbb{Q}^{y,r}$, the process

$$B_t^{y,r} := B_t - \int_0^t u_s\,ds, \qquad 0 \le t \le r,$$

is a $(\mathcal{G}_t)$-Brownian motion. Since equation 46 holds $\mathbb{P}$-a.s. and $\mathbb{Q}^{y,r} \ll \mathbb{P}$, the same stochastic identity holds $\mathbb{Q}^{y,r}$-a.s. Substituting $dB_t = dB_t^{y,r} + u_t dt$ gives

$$dX_t = \big(b(t, X_t) + \sigma(t, X_t)u_t\big)dt + \sigma(t, X_t)dB_t^{y,r}. \tag{49}$$

Consequently,

$$\big(\Omega, \mathcal{F}, (\mathcal{G}_t)_{t\le r}, \mathbb{Q}^{y,r}, X, B^{y,r}\big)$$

is a weak solution of equation 2 with control u.

By equation 48,

$$\mathcal{L}_{\mathbb{Q}^{y,r}}(X_{[0,r]}) = \mathcal{L}_{\mathbb{P}^y}(X_{[0,r]}) = \mathcal{L}_{\mathbb{P}}(X_{[0,r]} \mid Y = y).$$

Thus, the solution of the controlled equation we have obtained via Girsanov shift has the conditioned state-path law $\mathcal{L}_{\mathbb{P}}(X_{[0,r]} \mid Y = y)$.

We next prove its minimum-energy property. Applying the exact limit identity equation 44 with $R = r$ yields

$$D_{\mathrm{KL}}\left(\mathbb{P}^y|_{\mathcal{F}_r(X)} \,\middle\|\, \mathbb{P}|_{\mathcal{F}_r(X)}\right) = D_{\mathrm{KL}}\left(\mathbb{P}^y|_{\mathcal{F}_0(X)} \,\middle\|\, \mathbb{P}|_{\mathcal{F}_0(X)}\right) + \frac{1}{2}\mathbb{E}^{\mathbb{P}^y}\int_0^r |u_t|^2 dt. \tag{50}$$

Since $\int_0^r |u_t|^2 dt$ is $\mathcal{F}_r(X)$-measurable and $\mathbb{Q}^{y,r}$ and $\mathbb{P}^y$ agree on $\mathcal{F}_r(X)$, equivalently,

$$\frac{1}{2}\mathbb{E}^{\mathbb{Q}^{y,r}}\int_0^r |u_t|^2 dt = D_{\mathrm{KL}}\left(\mathbb{P}^y|_{\mathcal{F}_r(X)} \,\middle\|\, \mathbb{P}|_{\mathcal{F}_r(X)}\right) - D_{\mathrm{KL}}\left(\mathbb{P}^y|_{\mathcal{F}_0(X)} \,\middle\|\, \mathbb{P}|_{\mathcal{F}_0(X)}\right).$$

Let $(\mathbb{Q}, v) \in \mathcal{A}_r^y$ be such that $\mathcal{L}_{\mathbb{Q}}(X_{[0,r]}) = \mathcal{L}_{\mathbb{P}}(X_{[0,r]}|Y = y)$. Then $\mathbb{Q}|_{\mathcal{F}_r(X)} = \mathbb{P}^y|_{\mathcal{F}_r(X)}$, and data processing inequality gives

$$\begin{aligned} D_{\mathrm{KL}}\left(\mathbb{P}^y|_{\mathcal{F}_r(X)} \,\middle\|\, \mathbb{P}|_{\mathcal{F}_r(X)}\right) &\le D_{\mathrm{KL}}(\mathbb{Q}\|\mathbb{P}) \\ &= H_0^y + \frac{1}{2}\mathbb{E}^{\mathbb{Q}}\int_0^r |v_t|^2 dt. \end{aligned}$$

Combining this inequality with equation 50 yields

$$\mathbb{E}^{\mathbb{Q}}\int_0^r |v_t|^2 dt \ge \mathbb{E}^{\mathbb{P}^y}\int_0^r |u_t|^2 dt.$$

The pair $(\mathbb{Q}^{y,r}, u)$ has initial density $Z_0(X)$, and the construction above shows that $B^{y,r} = B - \int_0^\cdot u_s\, ds$ is Brownian under $\mathbb{Q}^{y,r}$; indeed, $\mathbb{E}^{\mathbb{P}}[Z_r(X) \mid \mathcal{G}_0] = Z_0(X)$. Identity present in equation 50 gives the required entropy identity, and thus $(\mathbb{Q}^{y,r}, u) \in \mathcal{A}_r^y$ and it attains equality.

If another admissible adapted Girsanov shift attains the same value, then the data-processing inequality used above must hold with equality. By the equality case of conditional Jensen's inequality,

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \mathbb{E}^{\mathbb{P}}\left[\frac{d\mathbb{Q}}{d\mathbb{P}}\middle|\, \mathcal{F}_r(X)\right] \qquad \mathbb{P}\text{-a.s.}$$

Thus, $d\mathbb{Q}/d\mathbb{P}$ admits an $\mathcal{F}_r(X)$-measurable version. Since its restriction to $\mathcal{F}_r(X)$ has density $Z_r(X)$, it follows that

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = Z_r(X) \qquad \mathbb{P}\text{-a.s.},$$

and hence

$$\mathbb{Q} = \mathbb{Q}^{y,r} \qquad \text{on } \mathcal{G}_r.$$

Theorem 4 was used above to identify the Girsanov shift u through the martingale representation of the density process, and the corresponding Girsanov argument shows that $B - \int_0^\cdot u_s\, ds$ is a Brownian motion under $\mathbb{Q}^{y,r}$. By admissibility, $B - \int_0^\cdot v_s\, ds$ is a Brownian motion under $\mathbb{Q}$. Under the common measure $\mathbb{Q} = \mathbb{Q}^{y,r}$, uniqueness of the continuous semimartingale decomposition therefore gives

$$v_t = u_t \qquad dt \otimes d\mathbb{Q}^{y,r}\text{-a.e.}$$

This proves the minimum-energy property and uniqueness.

□

D STATEMENT AND PROOF OF THEOREM 3

First, we provide a formal version of Theorem 3. For simplicity, we assume that X_0 is deterministic.

Theorem 10. *Suppose that $X_0 = x_0$ is deterministic and Assumption 1 holds. Fix $r < T$, set $\mu_r := \mathcal{L}_{\mathbb{P}}(X_r)$, and let ν_r be a probability measure such that*

$$\nu_r \ll \mu_r, \qquad z_r := \frac{d\nu_r}{d\mu_r} \in L^2(\mu_r).$$

Let $\mathcal{A}_r$ be the class of pairs $(\mathbb{Q}, v)$ such that $\mathbb{Q} \ll \mathbb{P}$ on $\mathcal{G}_r$, v is $(\mathcal{G}_t)$-predictable,

$$B_t^{\mathbb{Q}} := B_t - \int_0^t v_s\, ds$$

is a $(\mathcal{G}_t, \mathbb{Q})$-Brownian motion, $\mathbb{E}^{\mathbb{Q}} \int_0^r |v_t|^2 dt < \infty$. Consider the stochastic optimal control problem

$$V(r, \nu_r) := \inf_{(\mathbb{Q},v)\in\mathcal{A}_r} \mathbb{E}^{\mathbb{Q}} \left[\frac{1}{2} \int_0^r |v_t|^2 dt - \log z_r \right], \tag{51}$$

where the objective is interpreted in the extended sense. Define

$$Z_t^{\nu,r}(X) := \mathbb{E}^{\mathbb{P}} \left[z_r \mid \mathcal{F}_t(X) \right], \qquad 0 \le t \le r.$$

Then, $Z_0^{\nu,r}(X) = 1$, $Z_r^{\nu,r}(X) = z_r$, and there exists a unique $\xi^{\nu,r} \in L_a^2([0,r], \mathcal{F}(X), \mathbb{P}; \mathbb{R}^d)$ such that

$$Z_t^{\nu,r}(X) = 1 + \int_0^t \langle \xi_s^{\nu,r}, dB_s \rangle, \qquad P_t(X)\xi_t^{\nu,r} = \xi_t^{\nu,r} \quad dt \otimes d\mathbb{P}\text{-a.e.}$$

Set

$$u_t^{\nu,r} := \frac{\xi_t^{\nu,r}}{Z_t^{\nu,r}(X)} \quad \text{on } \{Z_t^{\nu,r}(X) > 0\}, \qquad u_t^{\nu,r} := 0 \quad \text{otherwise},$$

and define

$$\frac{d(\mathbb{Q}^{\nu,r}|_{\mathcal{G}_r})}{d(\mathbb{P}|_{\mathcal{G}_r})} = z_r(X_r).$$

Then we have $(\mathbb{Q}^{\nu,r}, u^{\nu,r}) \in \mathcal{A}_r$, such that $P_t(X)u_t^{\nu,r} = u_t^{\nu,r}$ $dt \otimes d\mathbb{Q}^{\nu,r}$-a.e., and

$$B_t^{\nu,r} := B_t - \int_0^t u_s^{\nu,r}\, ds$$

is a $(\mathcal{G}_t, \mathbb{Q}^{\nu,r})$-Brownian motion. Under this measure, the coordinate process satisfies

$$dX_t = (b(t, X_t) + \sigma(t, X_t)u_t^{\nu,r})\, dt + \sigma(t, X_t) dB_t^{\nu,r}. \tag{52}$$

Moreover,

$$\mathcal{L}_{\mathbb{Q}^{\nu,r}}(X_r) = \nu_r, \qquad V(r, \nu_r) = 0,$$

and $u^{\nu,r}$ is the unique minimum-energy Girsanov shift realizing ν_r:

$$\frac{1}{2}\mathbb{E}^{\mathbb{Q}^{\nu,r}} \int_0^r |u_t^{\nu,r}|^2 dt = \inf_{\substack{(\mathbb{Q},v)\in\mathcal{A}_r: \\ \mathcal{L}_{\mathbb{Q}}(X_r)=\nu_r}} \frac{1}{2}\mathbb{E}^{\mathbb{Q}} \int_0^r |v_t|^2 dt.$$

In particular, under Assumption 2, one may take $\nu_r = \mathcal{L}_{\mathbb{P}}(X_r \mid Y = y)$. We then write z_r^y, $Z^{y,r}(X)$, $u^{y,r}$, and $\mathbb{Q}^{y,r}$ for the corresponding objects.

Remark 8. *Pairs induced by strictly positive true stochastic exponentials (a.k.a. Doléans-Dade exponentials) are special cases of $\mathcal{A}_r$. Allowing $\mathbb{Q}$ to be merely absolutely continuous, rather than equivalent, includes corresponding densities that may vanish under $\mathbb{P}$.*

Proof of Theorem 10. Since $z_r \in L^2(\mathcal{F}_r(X), \mathbb{P})$ and X_0 is deterministic,

$$Z_0^{\nu,r}(X) = 1, \qquad Z_r^{\nu,r}(X) = z_r.$$

Theorem 4 implies that there exists $\mathcal{F}(X)$-adapted square integrable $\xi^{\nu,r}$ such that

$$Z_t^{\nu,r}(X) = 1 + \int_0^t \langle \xi_s^{\nu,r}, dB_s \rangle, \qquad P_s(X)\xi_s^{\nu,r} = \xi_s^{\nu,r}.$$

Define

$$u_s^{\nu,r} := \frac{\xi_s^{\nu,r}}{Z_s^{\nu,r}(X)} \mathbf{1}_{\{Z_s^{\nu,r}(X)>0\}}.$$

Then

$$dZ_t^{\nu,r}(X) = Z_t^{\nu,r}(X)\langle u_t^{\nu,r}, dB_t \rangle, \qquad P_t(X)u_t^{\nu,r} = u_t^{\nu,r}.$$

Moreover,

$$\{Z_t^{\nu,r}(X) = 0\}$$

is $dt \otimes d\mathbb{Q}^{\nu,r}$-null.

The localization and uniform-integrability argument in the proof of Theorem 6 applies with terminal density $z_r(X_r)$. It shows that $B^{\nu,r}$ is Brownian under $\mathbb{Q}^{\nu,r}$ and that

$$\frac{1}{2}\mathbb{E}^{\mathbb{Q}^{\nu,r}} \int_0^r |u_t^{\nu,r}|^2 dt = D_{\mathrm{KL}}(\mathbb{Q}^{\nu,r} \| \mathbb{P}).$$

Thus $(\mathbb{Q}^{\nu,r}, u^{\nu,r}) \in \mathcal{A}_r$. Since $d\mathbb{Q}^{\nu,r}/d\mathbb{P} = z_r(X_r)$,

$$\mathcal{L}_{\mathbb{Q}^{\nu,r}}(X_r) = \nu_r.$$

The original SDE identity holds $\mathbb{Q}^{\nu,r}$-a.s.; substituting $dB_t = dB_t^{\nu,r} + u_t^{\nu,r}\, dt$ proves equation 52.

For any $(\mathbb{Q}, v) \in \mathcal{A}_r$, let $\lambda_r := \mathcal{L}_{\mathbb{Q}}(X_r)$. Data processing inequality for relative entropy gives

$$D_{\mathrm{KL}}(\mathbb{Q}\|\mathbb{P}) \geq D_{\mathrm{KL}}(\lambda_r \| \mu_r).$$

Therefore, with the standard extended-value convention,

$$\begin{aligned}\mathbb{E}^{\mathbb{Q}}\left[\frac{1}{2}\int_0^r |v_t|^2 dt - \log z_r(X_r)\right] &= D_{\mathrm{KL}}(\mathbb{Q}\|\mathbb{P}) - \int \log z_r \, d\lambda_r \\ &\geq D_{\mathrm{KL}}(\lambda_r \| \nu_r) \geq 0.\end{aligned}$$

The corresponding pair $(\mathbb{Q}^{\nu,r}, u^{\nu,r})$ attains zero, and hence $V(r, \nu_r) = 0$.

Now restrict to competitors satisfying $\lambda_r = \nu_r$. Then

$$\frac{1}{2}\mathbb{E}^{\mathbb{Q}} \int_0^r |v_t|^2 dt \geq D_{\mathrm{KL}}(\nu_r \| \mu_r) = \frac{1}{2}\mathbb{E}^{\mathbb{Q}^{\nu,r}} \int_0^r |u_t^{\nu,r}|^2 dt.$$

If equality holds, equality also holds in data processing. Its equality case shows that $d\mathbb{Q}/d\mathbb{P}$ is $\sigma(X_r)$-measurable. Since its pushforward density is z_r, necessarily

$$\frac{d(\mathbb{Q}|_{\mathcal{G}_r})}{d(\mathbb{P}|_{\mathcal{G}_r})} = z_r = \frac{d(\mathbb{Q}^{\nu,r}|_{\mathcal{G}_r})}{d(\mathbb{P}|_{\mathcal{G}_r})}.$$

Thus $\mathbb{Q} = \mathbb{Q}^{\nu,r}$. Under this common measure, both $B - \int_0^\cdot v_s\, ds$ and $B - \int_0^\cdot u_s^{\nu,r}\, ds$ are Brownian motions. Uniqueness of the continuous semimartingale decomposition gives $v_t = u_t^{\nu,r}\ dt \otimes d\mathbb{Q}^{\nu,r}$-a.e.

Note that, for any bounded Borel function φ,

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}\left[\varphi(X_r)\mathbb{E}^{\mathbb{P}}[Z_r^y(X) \mid X_r]\right] &= \mathbb{E}^{\mathbb{P}}\left[\varphi(X_r) Z_r^y(X)\right] \\ &= \mathbb{E}^{\mathbb{P}^y}\left[\varphi(X_r)\right] \\ &= \int \varphi(x) z_r^y(x)\, \mathcal{L}_{\mathbb{P}}(X_r)(dx).\end{aligned}$$

Thus, $\mathcal{L}_{\mathbb{P}}(X_r \mid Y = y) \ll \mu_r$, and

$$z_r^y(X_r) = \mathbb{E}^{\mathbb{P}}[Z_r^y(X) \mid X_r].$$

The L^2-contraction and Assumption 2 give $z_r^y \in L^2(\mathbb{P})$, and thus all preceding conclusions apply. □

E PROOFS OF THE DEGENERATE TWEEDIE FORMULAE

E.1 PROOF OF DEGENERATE TWEEDIE FORMULA-1

Proof of Theorem 1. Apply Theorem 6. It gives equation 3 in the stated L^2 space with $u_t(X_t, y) = \xi_t/Z_t(X)$. Since $P_t(X)\xi_t = \xi_t\, dt \otimes d\mathbb{P}$-a.e., we also have $P_t(X)u_t(X_t, y) = u_t(X_t, y)\, dt \otimes d\mathbb{P}^y$-a.e., and uniqueness follows from the uniqueness of the integrand in the martingale representation by Theorem 4. The conditional-law and minimum-entropy claims on every $[0, r]$, $r < T$, follow directly from Theorem 7. □

E.2 PROOF OF DEGENERATE TWEEDIE FORMULA-2

Proof of Corollary 1. Fix $r < T$, $0 \le t \le r$, and $0 < \varepsilon < T - r$. We apply Theorem 5 componentwise to $B_{t+\varepsilon} - B_t$.

For each $j \in \{1, \dots, d\}$, j^{th}-component of $B_{t+\varepsilon} - B_t$ satisfies

$$(B_{t+\varepsilon} - B_t)^j = \int_0^T \left\langle \mathbf{1}_{(t,t+\varepsilon]}(s)e_j, dB_s \right\rangle,$$

and $\mathbb{E}^{\mathbb{P}}\left[(B_{t+\varepsilon} - B_t)^j \mid X_0\right] = 0$, since $B_{t+\varepsilon} - B_t$ is independent of $\mathcal{G}_t$, and hence of $\sigma(X_0)$. Therefore, Theorem 5 gives

$$\mathbb{E}^{\mathbb{P}}\left[(B_{t+\varepsilon} - B_t)^j \mid \mathcal{F}_T(X)\right] = \int_t^{t+\varepsilon} \langle P_s(X)e_j, dB_s\rangle.$$

Applying this componentwise yields

$$\mathbb{E}^{\mathbb{P}}\left[B_{t+\varepsilon} - B_t \mid \mathcal{F}_T(X)\right] = \int_t^{t+\varepsilon} P_s(X)\, dB_s. \tag{53}$$

Since $Y = G(X_T)$, $\sigma\{X_t, Y\} \subseteq \mathcal{F}_T(X)$. Hence, by the tower property and equation 53,

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}\left[B_{t+\varepsilon} - B_t \mid X_t, Y\right] &= \mathbb{E}^{\mathbb{P}}\left[\mathbb{E}^{\mathbb{P}}\left[B_{t+\varepsilon} - B_t \mid \mathcal{F}_T(X)\right] \middle| X_t, Y\right] \\ &= \mathbb{E}^{\mathbb{P}}\left[\int_t^{t+\varepsilon} P_s(X)\, dB_s \middle| X_t, Y\right].\end{aligned}$$

Thus, the result follows from Theorem 1. □

E.3 PROOF OF DEGENERATE TWEEDIE FORMULA-3

Proof of Theorem 2. Fix R with $r < R < T$. Under the corresponding measure $\mathbb{Q}^{y,R}$ of Theorem 7, for $0 < \varepsilon < R - r$ and $0 \le t \le r$,

$$X_{t+\varepsilon} - X_t = \int_t^{t+\varepsilon} \big(b(s, X_s) + \sigma(s, X_s)u_s(X_s, y)\big)ds + \int_t^{t+\varepsilon} \sigma(s, X_s)\, dB_s^{y,R}.$$

The stochastic integral is a square-integrable $(\mathcal{G}_t, \mathbb{Q}^{y,R})$-martingale increment, so its conditional expectation given X_t is zero. Since $\mathbb{Q}^{y,R}$ and $\mathbb{P}^y$ agree on $\mathcal{F}_R(X)$, this gives

$$\begin{aligned}\mathbb{E}^{\mathbb{P}^y}\left[\frac{X_{t+\varepsilon} - X_t}{\varepsilon} - b(t, X_t)\middle| X_t\right] &= \mathbb{E}^{\mathbb{P}^y}\left[\frac{1}{\varepsilon}\int_t^{t+\varepsilon} \big(b(s, X_s) - b(t, X_t)\big)\, ds \middle| X_t\right] \\ &\quad + \mathbb{E}^{\mathbb{P}^y}\left[\frac{1}{\varepsilon}\int_t^{t+\varepsilon} \sigma(s, X_s)u_s(X_s, y)\, ds \middle| X_t\right].\end{aligned}$$

By Assumption 1 and Theorem 1,

$$\mathbb{E}^{\mathbb{P}^y}\int_0^r |\sigma(t, X_t)u_t(X_t, y)|^2 dt \le \|\sigma\|_\infty^2 \mathbb{E}^{\mathbb{P}^y}\int_0^r |u_t(X_t, y)|^2 dt < \infty.$$

Thus, by Assumption 1, the bound above, and the L^2-Lebesgue differentiation theorem,

$$\frac{1}{\varepsilon}\int_t^{t+\varepsilon}\big(b(s,X_s)-b(t,X_t)\big)\,ds \longrightarrow 0,$$
$$\frac{1}{\varepsilon}\int_t^{t+\varepsilon}\sigma(s,X_s)u_s(X_s,y)\,ds \longrightarrow \sigma(t,X_t)u_t(X_t,y)$$

in $L^2([0,r]\times\Omega, dt\otimes d\mathbb{P}^y)$. For any such pair $H_t^\varepsilon \to H_t$,

$$\int_0^r \mathbb{E}^{\mathbb{P}^y}\left|\mathbb{E}^{\mathbb{P}^y}[H_t^\varepsilon \mid X_t]-H_t\right|^2 dt \le \int_0^r \mathbb{E}^{\mathbb{P}^y}|H_t^\varepsilon - H_t|^2 dt \longrightarrow 0.$$

Thus both convergences remain valid after conditioning on X_t. Passing to the limit in the preceding identity proves equation 5. □

F Terminal-Increment Regression for Additive Diffusions

For additive diffusions with a fixed noise matrix, the local Brownian target can be replaced exactly by a terminal-increment target. This removes the auxiliary limit $\varepsilon \downarrow 0$ required in the the definition of the loss function.

Theorem 11. *Let B be a d-dimensional $(\mathcal{G}_t)$-Brownian motion, let X_0 be $\mathcal{G}_0$-measurable, and suppose that*

$$X_t = X_0 + \int_0^t b(s)\,ds + \sigma B_t,$$

where $b:[0,T]\to\mathbb{R}^n$ is deterministic and integrable and $\sigma\in\mathbb{R}^{n\times d}$ is fixed, possibly rank deficient. Let $Y = G(X_T)$ for a Borel map G. Fix $r<T$, let $\mathscr{H}_r$ be the σ-field on $[0,r]\times\Omega$ generated by $(t,\omega)\mapsto(t,X_t(\omega),Y(\omega))$. For $Q_t := (B_T-B_t)/(T-t)$, define the product-space conditional expectation

$$\bar{u} := \mathbb{E}^{dt\otimes d\mathbb{P}}[Q \mid \mathscr{H}_r].$$

Choose a Borel representative $\bar{u}:[0,r]\times\mathbb{R}^n\times\mathsf{Y}\to\mathbb{R}^d$ for this conditional expectation. Then, $\bar{u}(t,X_t,Y)=\mathbb{E}^{\mathbb{P}}[Q_t\mid X_t,Y]$ $dt\otimes d\mathbb{P}$-a.e. For every $0<\varepsilon<T-r$,

$$\bar{u}(t,X_t,Y) = \mathbb{E}^{\mathbb{P}}\left[\frac{P(B_T-B_t)}{T-t}\middle| X_t,Y\right] = \mathbb{E}^{\mathbb{P}}\left[\frac{B_{t+\varepsilon}-B_t}{\varepsilon}\middle| X_t,Y\right] \tag{54}$$

for $dt\otimes d\mathbb{P}$-a.e. $(t,\omega)\in[0,r]\times\Omega$. Moreover,

$$\mathbb{E}^{\mathbb{P}}\int_0^r\left|\frac{B_T-B_t}{T-t}\right|^2 dt = d\log\frac{T}{T-r}<\infty,$$

and the regression problem

$$\inf_{v\in L^2(\mathscr{H}_r,dt\otimes d\mathbb{P};\mathbb{R}^d)}\mathbb{E}^{\mathbb{P}}\int_0^r\left|v_t-\frac{B_T-B_t}{T-t}\right|^2 dt \tag{55}$$

has the unique minimizer $\bar{u}$.

If Assumptions 1–2 also hold, let ξ^y be the integrand from Lemma 2. For $\mathcal{L}_{\mathbb{P}}(Y)$-a.e. y, let $u_t^{\mathrm{path},y} := \xi_t^y/Z_t^y(X)$ on $\{Z_t^y(X)>0\}$ and set it to zero otherwise. Then $\bar{u}(t,X_t,y)=u_t^{\mathrm{path},y}$ in $L^2([0,r]\times\Omega, dt\otimes d\mathbb{P}^y)$. Consequently, the path-adapted control admits this state-feedback version, and the terminal-increment loss recovers the corresponding controlled state law on $[0,r]$ and its minimum-energy property.

Proof. Fix $0\le t\le r$ and $0<\varepsilon<T-r$, and write

$$\Delta B_{t,T} := B_T - B_t, \qquad \Delta B_{t,\varepsilon} := B_{t+\varepsilon}-B_t.$$

The Brownian bridge decomposition is

$$\Delta B_{t,\varepsilon} = \frac{\varepsilon}{T-t}\Delta B_{t,T} + R_{t,\varepsilon}, \tag{56}$$

where $R_{t,\varepsilon}$ is centered Gaussian and independent of $\mathcal{G}_t \vee \sigma(\Delta B_{t,T})$. Since

$$X_T = X_t + \int_t^T b(s)\, ds + \sigma \Delta B_{t,T},$$

the pair (X_t, Y) is measurable with respect to $\mathcal{G}_t \vee \sigma(\Delta B_{t,T})$. Hence $\mathbb{E}^{\mathbb{P}}[R_{t,\varepsilon} \mid X_t, Y] = 0$, and conditioning equation 56 gives

$$\mathbb{E}^{\mathbb{P}}\left[\frac{B_{t+\varepsilon} - B_t}{\varepsilon} \middle| X_t, Y \right] = \mathbb{E}^{\mathbb{P}}\left[\frac{B_T - B_t}{T-t} \middle| X_t, Y \right].$$

Since P is the orthogonal projection onto $\mathrm{Ran}(\sigma^\top)$, we have $\sigma(I - P) = 0$. The Gaussian vectors $P\Delta B_{t,T}$ and $(I - P)\Delta B_{t,T}$ are independent, and the latter is jointly independent of $\mathcal{G}_t \vee \sigma(P\Delta B_{t,T})$. Since Y depends on the future increment only through $P\Delta B_{t,T}$,

$$\mathbb{E}^{\mathbb{P}}[(I - P)\Delta B_{t,T} \mid X_t, Y] = 0.$$

This proves all identities in equation 54.

The stated integrability follows from

$$\mathbb{E}^{\mathbb{P}}|B_T - B_t|^2 = d(T - t).$$

The regression claim is the orthogonal-projection characterization of conditional expectation in the Hilbert space $\mathcal{V}_r$. Finally, choose a deterministic sequence $\varepsilon_k \downarrow 0$ and disintegrate the exact identities in equation 54 simultaneously outside one $\mathcal{L}_{\mathbb{P}}(Y)$-null set. The same independence argument with conditioning on $\mathcal{F}_t(X) \vee \sigma(Y)$ gives

$$\mathbb{E}^{\mathbb{P}}\left[\frac{B_{t+\varepsilon_k} - B_t}{\varepsilon_k} \middle| \mathcal{F}_t(X) \vee \sigma(Y) \right] = \mathbb{E}^{\mathbb{P}}\left[\frac{B_T - B_t}{T-t} \middle| X_t, Y \right].$$

The path-feedback form of Corollary 1 therefore yields $\bar{u}(t, X_t, y) = u_t^{\mathrm{path},y}$ in the stated fiberwise L^2 space. In particular, the additive model supplies its own state-feedback version without Assumption 3. The controlled-law and minimum-energy conclusions follow from the path-feedback form of Theorem 7. □

G Implementation of the Training Procedure

This section studies the procedure for the convergence to the control u appearing in equation 3. We consider both the projected target used in Proposition 1 and the unprojected Brownian-increment target from Corollary 1.

G.1 Population regression

Assume throughout this subsection that Assumptions 1, 2, and 3 hold. Fix $r < T$ and $0 < \varepsilon < T - r$. Define

$$M_t^{\varepsilon,\mathrm{proj}} := \frac{1}{\varepsilon} \int_t^{t+\varepsilon} P_s(X)\, dB_s \text{ and } M_t^{\varepsilon,\mathrm{raw}} := \frac{B_{t+\varepsilon} - B_t}{\varepsilon}.$$

Let $\mathscr{H}_r$ be the product σ-field on $[0, r] \times \Omega$ generated by the map $(t, \omega) \mapsto (t, X_t(\omega), Y(\omega))$, and let

$$\mathcal{V}_r := L^2(\mathscr{H}_r, dt \otimes d\mathbb{P}; \mathbb{R}^d).$$

This is the closed regression space corresponding to state feedback controls that appears in equation 2.

For either choice $M_t^\varepsilon \in \{M_t^{\varepsilon,\mathrm{proj}}, M_t^{\varepsilon,\mathrm{raw}}\}$, consider the population loss

$$\mathcal{L}_\varepsilon(v) = \mathbb{E}^{\mathbb{P}} \int_0^r |v_t - M_t^\varepsilon|^2\, dt.$$

Its unique minimizer over $\mathcal{V}_r$ is the product-space conditional expectation. Choose a Borel representative $u_\varepsilon^* : [0, r] \times \mathbb{R}^n \times \mathsf{Y} \to \mathbb{R}^d$; then, $dt \otimes d\mathbb{P}$-a.e.,

$$u_\varepsilon^*(t, X_t, Y) = \mathbb{E}^{\mathbb{P}}\left[M_t^\varepsilon \mid X_t, Y \right].$$

When no arguments are displayed, u_ε^* denotes the realized process $u_\varepsilon^*(t, X_t, Y)$.

Indeed, the orthogonality property of conditional expectation gives

$$\mathcal{L}_\varepsilon(v) - \mathcal{L}_\varepsilon(u_\varepsilon^*) = \mathbb{E}^{\mathbb{P}} \int_0^r |v_t - u_\varepsilon^*(t, X_t, Y)|^2 \, dt. \tag{57}$$

Consequently, minimizing the squared regression loss is equivalent to approximating the appropriate conditional expectation.

After disintegration with respect to Y, Proposition 1 and Corollary 1 give, for $\mathcal{L}_{\mathbb{P}}(Y)$-a.e. y,

$$u_\varepsilon^{*,\mathrm{proj}} \longrightarrow u \quad \text{in } L^2([0,r] \times \Omega, dt \otimes d\mathbb{P}^y), \qquad u_\varepsilon^{*,\mathrm{raw}} \longrightarrow u \quad \text{in } L^2([0,r] \times \Omega, dt \otimes d\mathbb{P}^y).$$

Let $g_\varepsilon(y)$ be such that

$$\mathbb{E}^{\mathbb{P}} \int_0^r |u_\varepsilon^*(t, X_t, Y) - u_t(X_t, Y)|^2 \, dt = \int g_\varepsilon(y) \, \mathcal{L}_{\mathbb{P}}(Y)(dy).$$

We additionally assume convergence in $L^2([0,r] \times \Omega, dt \otimes d\mathbb{P})$; it follows, for example, if $g_\varepsilon(y) \to 0$ for $\mathcal{L}_{\mathbb{P}}(Y)$-a.e. y and the family $(g_\varepsilon)_\varepsilon$ is uniformly integrable under $\mathcal{L}_{\mathbb{P}}(Y)$. Under this joint regularity conditions, both losses recover the same minimum-energy control.

G.2 Discretization of the projected target

In this subsection, we provide an approximation scheme for $M_{t_j}^{\varepsilon,\mathrm{proj}}$ that appears in equation 7, whose convergence is not obvious since the projection $(P_t(X))_t$ is not continuous.

Let $t_k = kh$ be a uniform grid, assume $T = Kh$ for an integer K, and suppose that $\varepsilon = qh$ for an integer $q \geq 1$. Let X^h be the Euler–Maruyama approximation of X, and let $\eta_h(0) = 0$ and $\eta_h(s) = t_k$ for $s \in (t_k, t_{k+1}]$. Define the predictable piecewise-constant approximation

$$\overline{P}_s^h := P_{\eta_h(s)}(X^h), \qquad 0 \leq s \leq T,$$

where $\overline{P}_0^h = P_0(X^h)$. Equivalently,

$$\overline{P}_s^h = P_{t_k}(X^h), \qquad s \in (t_k, t_{k+1}].$$

For $0 \leq t \leq r$, set

$$M_t^{\varepsilon,\mathrm{proj},h} := \frac{1}{\varepsilon} \int_t^{t+\varepsilon} \overline{P}_s^h \, dB_s \implies M_{t_j}^{\varepsilon,\mathrm{proj},h} = \frac{1}{qh} \sum_{k=j}^{j+q-1} P_{t_k}(X^h)\big(B_{t_{k+1}} - B_{t_k}\big).$$

By Itô's isometry and Fubini's theorem,

$$\mathbb{E}^{\mathbb{P}} \int_0^r \left| M_t^{\varepsilon,\mathrm{proj},h} - M_t^{\varepsilon,\mathrm{proj}} \right|^2 dt \leq \frac{1}{\varepsilon} \mathbb{E}^{\mathbb{P}} \int_0^{r+\varepsilon} \left\| \overline{P}_s^h - P_s(X) \right\|_{\mathrm{F}}^2 ds, \tag{58}$$

where $\|\cdot\|_{\mathrm{F}}$ is the Frobenius norm. Therefore, for fixed ε, the discrete target converges whenever

$$\mathbb{E}^{\mathbb{P}} \int_0^{r+\varepsilon} \left\| \overline{P}_s^h - P_s(X) \right\|_{\mathrm{F}}^2 ds \longrightarrow 0.$$

For example, suppose that the state-dependent projection $P(t,x)$ satisfies

$$\|P(t,x) - P(s,z)\|_{\mathrm{F}} \leq L_P\big(|x - z| + |t - s|^{1/2}\big), \tag{59}$$

and that the numerical approximation of X has mean-square error of order h. Then,

$$\mathbb{E}^{\mathbb{P}} \int_0^{r+\varepsilon} \left\| \overline{P}_s^h - P_s(X) \right\|_{\mathrm{F}}^2 ds \leq Ch \implies \mathbb{E}^{\mathbb{P}} \int_0^r \left| M_t^{\varepsilon,\mathrm{proj},h} - M_t^{\varepsilon,\mathrm{proj}} \right|^2 dt \leq C \frac{h}{\varepsilon}.$$

The displayed bound is driven to zero by the sufficient conditions $\varepsilon \downarrow 0$ and $h/\varepsilon \to 0$.

Below, we provide a sufficient condition for equation 59.

Proposition 2. *Let X^h be the Euler–Maruyama approximation of X. Suppose that b and σ are globally Lipschitz in the state variable, $1/2$-Hölder continuous in time, and of linear growth, with $\mathbb{E}^{\mathbb{P}}|X_0|^2 < \infty$. Assume further that, for some $\rho \geq 1$ and $\kappa > 0$,*

$$\operatorname{rank}\sigma(t,x) = \rho, \qquad s_\rho\big(\sigma(t,x)\big) \geq \kappa$$

for every (t,x), where $s_\rho(\sigma)$ denotes the smallest nonzero singular value of σ. Then, there is a constant C, independent of h and ε, such that

$$\mathbb{E}^{\mathbb{P}} \int_0^{r+\varepsilon} \left\| \overline{P}_s^h - P_s(X) \right\|_{\mathrm{F}}^2 ds \leq Ch.$$

Consequently,

$$\mathbb{E}^{\mathbb{P}} \int_0^r \left| M_t^{\varepsilon,\mathrm{proj},h} - M_t^{\varepsilon,\mathrm{proj}} \right|^2 dt \leq C\frac{h}{\varepsilon}.$$

Proof. The constant-rank assumption and the uniform lower bound on the smallest nonzero singular value imply that the orthogonal projection $\Pi(a) = a^\dagger a$ onto $\operatorname{Ran}(a^\top)$ is Lipschitz continuous as a function of a on this class of matrices. For $a, \widetilde{a} \in \mathbb{R}^{n\times d}$ satisfying

$$\operatorname{rank}(a) = \operatorname{rank}(\widetilde{a}) = \rho, \qquad s_\rho(a),\, s_\rho(\widetilde{a}) \geq \kappa,$$

let

$$\Pi(a) := a^\dagger a, \qquad \Pi(\widetilde{a}) := \widetilde{a}^\dagger \widetilde{a}$$

denote the orthogonal projections onto $\operatorname{Ran}(a^\top)$ and $\operatorname{Ran}(\widetilde{a}^\top)$, respectively. Since

$$\widetilde{a}\big(I - \Pi(\widetilde{a})\big) = 0,$$

we have

$$\Pi(a)\big(I - \Pi(\widetilde{a})\big) = a^\dagger(a - \widetilde{a})\big(I - \Pi(\widetilde{a})\big).$$

Moreover,

$$\|a^\dagger\|_{\mathrm{op}} = \frac{1}{s_\rho(a)} \leq \frac{1}{\kappa},$$

and therefore

$$\|\Pi(a)(I - \Pi(\widetilde{a}))\|_{\mathrm{F}} \leq \frac{1}{\kappa}\|a - \widetilde{a}\|_{\mathrm{F}} \implies \|(I - \Pi(a))\Pi(\widetilde{a})\|_{\mathrm{F}} \leq \frac{1}{\kappa}\|a - \widetilde{a}\|_{\mathrm{F}}.$$

Since

$$\Pi(a) - \Pi(\widetilde{a}) = \Pi(a)(I - \Pi(\widetilde{a})) - (I - \Pi(a))\Pi(\widetilde{a}),$$

we obtain

$$\|\Pi(a) - \Pi(\widetilde{a})\|_{\mathrm{F}} \leq \frac{2}{\kappa}\|a - \widetilde{a}\|_{\mathrm{F}} \implies \|\widetilde{\Pi}(t,\sigma(\cdot,x)) - \widetilde{\Pi}(s,\sigma(\cdot,z))\|_{\mathrm{F}} \leq \frac{2}{\kappa}\|\sigma(t,x) - \sigma(s,z)\|_{\mathrm{F}}.$$

Using the Lipschitz continuity of σ in the state variable and its $1/2$-Hölder continuity in time, we conclude that

$$\|\widetilde{\Pi}(t,\sigma(\cdot,x)) - \widetilde{\Pi}(s,\sigma(\cdot,z))\|_{\mathrm{F}} \leq C\big(|x - z| + |t - s|^{1/2}\big),$$

where C depends only on κ and the corresponding regularity constants of σ.

The standard strong Euler–Maruyama estimate and the increment estimate for X yield

$$\mathbb{E}^{\mathbb{P}} \left| X_s - X^h_{\eta_h(s)} \right|^2 \leq Ch,$$

see (Kloeden & Platen, 1992, Theorem 10.2.2.). Hence,

$$\begin{aligned}\mathbb{E}^{\mathbb{P}} \left\| \overline{P}_s^h - P_s(X) \right\|_{\mathrm{F}}^2 &\leq C\left(\mathbb{E}^{\mathbb{P}} \left| X^h_{\eta_h(s)} - X_s \right|^2 + |s - \eta_h(s)| \right) \\ &\leq Ch.\end{aligned}$$

Integrating over $[0, r+\varepsilon]$ proves the first assertion.

Finally, Itô's isometry and Fubini's theorem give

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}\int_0^r \left|M_t^{\varepsilon,\mathrm{proj},h} - M_t^{\varepsilon,\mathrm{proj}}\right|^2 dt &= \frac{1}{\varepsilon^2}\int_0^r \mathbb{E}^{\mathbb{P}}\int_t^{t+\varepsilon} \left\|\overline{P}_s^h - P_s(X)\right\|_{\mathrm{F}}^2 ds\, dt \\ &\leq \frac{1}{\varepsilon}\mathbb{E}^{\mathbb{P}}\int_0^{r+\varepsilon} \left\|\overline{P}_s^h - P_s(X)\right\|_{\mathrm{F}}^2 ds \\ &\leq C\frac{h}{\varepsilon}.\end{aligned}$$

□

G.3 The unprojected loss function

When Corollary 1 is used, the discrete target is

$$M_{t_j}^{\varepsilon,\mathrm{raw}} = \frac{B_{t_{j+q}} - B_{t_j}}{qh} = \frac{1}{qh}\sum_{k=j}^{j+q-1}\left(B_{t_{k+1}} - B_{t_k}\right). \tag{60}$$

This equality is exact. Hence, unlike the projected case, there is no stochastic-integral discretization error and no regularity assumption on $P_t(X)$ is needed. A time discretization is still required for the integral in the loss and, when X is simulated numerically, for the state process itself.

G.4 Monte Carlo and optimization consistency

To approximate the loss functions we defined for the Tweedie's formula, a standard approach is to use neural networks. We refer to (Tzen & Raginsky, 2019, Section 3) for a theoretical study of the feasibility of this approach. Their analysis is directly applicable in our setting.