entropy

MDPI

Article

Adaptive Path Integral Diffusion: AdaPID

Michael Chertkov * and Hamidreza Behjoo

Program in Applied Mathematics, Department of Mathematics, University of Arizona, Tucson, AZ 85721, USA; hamidreza.behjoo@gmail.com
* Correspondence: chertkov@arizona.edu

Abstract

Harmonic Path Integral Diffusion (H-PID) provides an analytically tractable framework for sampling from a target density $p^{(\mathrm{tar})}(x) \propto \exp(-E(x))$. H-PID can be viewed as a diffusion bridge model solving a stochastic optimal transport problem from a δ-density at $t = 0$ to the target density at $t = 1$. The dynamics are governed by a controlled stochastic differential equation, and the corresponding variational stochastic optimal transport objective combines a time-dependent quadratic potential, $\beta_t \|x_t\|^2/2$, with a kinetic control cost, $\|u(t; x_t)\|^2/2$. The focus of this paper is the design of the temporal stiffness protocol β_t, which enables explicit control of intermediate sampling dynamics when the terminal density is fixed. We exploit the central advantage of H-PID—its integrability—which yields an explicit representation of the optimal control in terms of the target density and Green functions of the associated linear forward and backward diffusion-in-a-potential problems. Our main contribution is to convert this integrable structure into a practical methodology for protocol optimization. Specializing to piecewise-constant stiffness schedules and Gaussian-mixture targets, we develop two complementary optimization principles: The first is a deterministic one, relying on explicit evaluation of the dynamic marginals, and exemplified on a velocity-gradient-sensitivity objective, which provides a computationally controlled framework for optimizing transport regularity and stiffness. The second is a stochastic one, implemented via sampling, and exemplified on sharpness-based temporal-memory objective regularized to favor transitions within a prescribed time window that targets the temporal organization of the sampling path. These two objectives illuminate different aspects of the same protocol-design problem. The velocity-gradient-sensitivity objective serves as a clean methodological backbone and supports interpretable optimization and scaling studies. The sharpness-based objective reveals that schedule quality is target-dependent, and that the dependence on β is not universal: different target geometries may favor different stiffness regimes and qualitatively different transient organizations. Examples with low- and moderate-dimensional Gaussian mixtures demonstrate that the proposed approach can control not only the terminal sampling accuracy but also the transient evolution of probability mass, while remaining computationally light and theoretically transparent.

Keywords: generative models; diffusion models; diffusion bridge; optimal transport; path integrals; stochastic optimal control; schedule optimization; Gaussian mixtures; sampling diagnostics

Check for updates

Academic Editors: Pan Zhang and Antonio M. Scarfone

Received: 15 January 2026
Revised: 8 April 2026
Accepted: 20 April 2026
Published: 24 April 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Diffusion-based sampling methods [1–5] have recently emerged as one of the most powerful paradigms for generative modeling and stochastic inference. A number of closely

related frameworks have been developed in this context, including denoising diffusion probabilistic models (DDPMs) [2], score-based diffusion models [3], Schrödinger bridge formulations [4], and flow-based generative models [5]. Despite their different formulations, these approaches share a common conceptual objective: to construct a stochastic or deterministic transport process that progressively transforms a simple reference distribution into a complex target distribution.

A fundamental feature shared by diffusion-based generative methods is that the target distribution alone does not uniquely determine the dynamics of the sampling process. There typically exists a large family of stochastic dynamics whose terminal distribution coincides with the desired target distribution. Consequently, the design of the diffusion dynamics itself becomes a nontrivial modeling choice.

Most existing work evaluates such constructions primarily in terms of their ability to reproduce the target distribution asymptotically. However, this criterion alone does not distinguish between the many possible transport dynamics that achieve the same terminal distribution. An important question therefore arises: **how should one choose among these admissible dynamics?**

In this work, we argue that the choice of dynamics should be guided not only by the correctness of the terminal distribution but also by the properties of the **transport trajectory** connecting the reference and target distributions. Different dynamics correspond to different transport paths and can lead to substantially different transient sampling behavior. This motivates the development of systematic criteria for evaluating diffusion dynamics beyond asymptotic correctness.

The Path Integral Diffusion (PID) framework introduced in [6] provides a particularly convenient setting for studying such questions. PID can be viewed as a diffusion bridge formulation in which the sampling process is constructed as a controlled stochastic trajectory that transports point distribution to a target distribution in unit time. Within this formulation, the score function appears naturally as an optimal control that steers the diffusion between the two distributions.

A distinctive feature of PID, generalizing the path integral control [7,8] to diffusive sampling, is that the resulting control problem is analytically integrable. Specifically, the optimal control (equivalently, the score function) admits an explicit representation in terms of a convolution between the target distribution and a ratio of Green functions associated with a pair of forward and backward linear partial differential equations. In the Harmonic PID construction, these Green functions can be written in closed form, making the transport dynamics analytically tractable while still allowing for nontrivial target distributions.

This integrable diffusion bridge structure provides a powerful analytic framework for studying diffusion dynamics. In particular, it allows one to explore how different choices of the transport protocol influence the behavior of the sampling trajectory, while maintaining an explicit connection to the target distribution.

In the original work on Harmonic Path Integral Diffusion (H-PID) [6] we introduced a constant stiffness protocol and demonstrated that the resulting dynamics admits a useful analytic structure that enables detailed analysis of the sampling process. While constant schedules provide a convenient baseline, they are not necessarily optimal for finite-time sampling, and they are also not sufficiently flexible for exploring details of transient dynamics, which can depend strongly on how the stiffness parameter varies over time.

Motivated by these observations, we ask the following question: once the terminal target law is fixed, how much freedom remains in the dynamics of sample generation, and can this freedom be used in a systematic and interpretable way? Our answer is affirmative in the analytically tractable setting of Harmonic Path Integral Diffusion (PID), where the diffusion dynamics are driven by a time-dependent quadratic potential and the target

distribution is a Gaussian Mixture Model (GMM). In this setting, one can explicitly parameterize a stiffness protocol, evaluate multiple notions of schedule quality, and optimize the protocol so as to shape intermediate-time behavior rather than merely the endpoint.

We formulate a practical workflow for protocol design in Harmonic PID: how to parameterize schedules, how to distinguish deterministic and stochastic objectives, how to optimize them, and how to interpret the resulting nontrivial schedules. The central message is not simply that different objectives lead to different schedules, although this is true and important. Rather, the main point is that Harmonic PID provides a rich and controllable methodology for shaping sample-generation dynamics in a way that is both theoretically transparent and computationally efficient.

The GMM/Harmonic PID setting is particularly attractive because it provides exact access to the time-dependent marginals $p_t(x)$. This creates a clean separation between two classes of protocol diagnostics. One-time quantities, such as spread or sensitivity proxies based on the control field, can be evaluated deterministically from exact marginals. In contrast, two-time observables that explicitly couple the current state to the endpoint remain genuinely pathwise and must be evaluated stochastically from trajectories. This separation is conceptually important and practically useful: it allows us to combine deterministic exact-marginal calculations with controlled stochastic estimation of temporal-memory observables.

Within this framework we study piecewise-constant (PWC) stiffness schedules and develop two complementary optimization principles: The first is deterministic and is based on a velocity-gradient-sensitivity (VGS) criterion—defined later in Equation (11)—that measures control regularity and dynamical stiffness. The second is stochastic and is based on a dynamic sharpness of the transient objective regularized by the requirement for the transition to occur at a prescribed time in the middle of the temporal interval—defined later in Equation (13). Together, they provide two concrete routes to schedule design. They also reveal that controlling the transient dynamics is not only possible but sufficiently rich to generate nontrivial, interpretable protocols, including higher-dimensional PWC schedules.

The examples in this paper should therefore be read as illustrations of a broader methodology. We do emphasize concrete low-dimensional and moderately high-dimensional benchmarks, because these are necessary to make the discussion explicit, reproducible, and delivering a sharp experimental message and a clear statement of scaling capabilities. However, the intended contribution is the methodology itself: an analytically controlled, computationally light, and conceptually flexible framework for protocol optimization in diffusion-bridge-based sampling.

2. Setting the Stage: Problem Formulation

We study finite-time diffusion-based transport from a point source to a prescribed target law. The basic optimization problem is

$$u_t^*(x;\beta_{0\to 1}) = \arg\min_{u_{0\to 1}} \int_0^1 \mathbb{E}\Big[\tfrac{1}{2}\|u_t\|^2 + V_t(x)\Big]\,dt, \tag{1}$$

$$\text{subject to} \qquad \begin{aligned} dx_t &= u_t(x_t)\,dt + dW_t, \qquad t \in [0,1], \\ x_0 &= 0, \qquad x_1 \sim p^{(\mathrm{tar})}, \end{aligned} \tag{2}$$

where $x_t \in \mathbb{R}^d$ and the target density

$$p^{(\mathrm{tar})}(y) = \left(Z^{(\mathrm{tar})}\right)^{-1} \exp(-E(y)), \qquad Z^{(\mathrm{tar})} \doteq \int \exp(-E(y'))\,dy', \tag{3}$$

is defined up to a normalization factor (partition function).

The drift (or control) field $u_t(x)$ in Equation (2) plays the role of a time-dependent drift (in the terminology of fluid mechanics) or score (in the terminology of generative AI) that transports probability mass from the initial point source at $t = 0$ to the target law at $t = 1$. The central question of this paper is not whether such a transport can be constructed—this is guaranteed by the PID framework—but how the time-dependent protocol can be designed so as to shape the **transient** dynamics of sampling.

2.1. Harmonic PID with Time-Dependent Stiffness

Throughout this paper, we work with a quadratic confining potential

$$V_t(x) = \frac{\beta_t}{2}\|x\|^2, \qquad \beta_{0\to 1} \doteq \{\beta_t\}_{t\in[0,1]}. \tag{4}$$

The scalar stiffness protocol β_t is the main design variable in this paper. In [6], β was taken to be constant. Here, we allow it to vary in time, thereby introducing a controlled tradeoff between exploration and contraction along the sampling path.

In the experiments, we restrict attention to piecewise-constant (PWC) schedules:

$$\beta_t = \beta^{(k)}, \qquad t \in [t_{k-1}, t_k), \qquad k = 1, \ldots, K_\beta. \tag{5}$$

This family is expressive enough to produce nontrivial schedules, while remaining low-dimensional enough to be optimized robustly with modest computational cost. It also preserves the analytic tractability of Harmonic PID.

The key structural fact is that, for any choice of protocol $\beta_{0\to 1}$, the associated stochastic optimal transport problem admits a closed-form solution for the optimal control:

$$u_t^*(x;\beta_{0\to 1}) = \nabla_x \log\Big(Z^{(\mathrm{probe})}(t;x;\beta_{0\to 1})\Big), \tag{6}$$

$$Z^{(\mathrm{probe})}(t;x;\beta_{0\to 1}) \doteq \int_{\mathbb{R}^d} q^{(\mathrm{probe})}(y|t;x;\beta_{0\to 1})\, dy, \tag{7}$$

$$q^{(\mathrm{probe})}(y|t;x;\beta_{0\to 1}) \doteq p^{(\mathrm{tar})}(y)\, \exp\big(-\Delta(t;x;y;\beta_{0\to 1})\big), \tag{8}$$

$$\Delta(t;x;y;\beta_{0\to 1}) \doteq -\log \frac{G_t^{(-)}(x \mid y;\beta_{0\to 1})}{G_1^{(+)}(y \mid 0;\beta_{0\to 1})}. \tag{9}$$

Here, $q^{(\mathrm{probe})}(y|t;x;\beta_{0\to 1})$ may be interpreted as an unnormalized density of terminal state y conditioned on observing state x at the earlier time t, and $G_t^{(\pm)}$ denotes the forward/backward Green functions of the auxiliary linear PDEs associated with the harmonic potential. The main significance of Equations (6)–(9) is that they make the dependence of the controlled dynamics on the protocol $\beta_{0\to 1}$ explicit. Thus, the protocol is not a passive parameter but a genuine design variable that can be optimized to shape the intermediate dynamics of sample generation. Further technical details, including the specialization to PWC schedules, can be found in Appendix A.

2.2. Why Gaussian Mixtures?

Gaussian-mixture (GM) targets play a central role in this work, not because we wish to restrict attention to toy examples, but because they provide an exceptional degree of analytic and computational control. In the H-PID setting, Gaussian-mixture targets allow explicit representations for the following:

- The forward and backward Green-function messages;
- The optimal control field;
- The time-dependent marginal density $p_t(x)$;
- Several moment-based quantities derived from these marginals.

As a result, the full marginal law $p_t(x)$ is available explicitly at every time t. This is a major computational advantage, because it allows deterministic evaluation (in terms of simple integrals or elementary functions) of all **one-time** observables, $\mathbb{E}\|f(x_t)\|^2$ or sensitivity measures derived from the control field.

At the same time, not all quantities of interest collapse to one-time exact-marginal expressions. Temporal-memory observables such as $\hat{A}(t)$ explicitly couple the current state to the endpoint and, therefore, remain genuinely pathwise, multi-time quantities. This distinction between one-time exact-marginal observables and two-time path observables is one of the main methodological themes of this paper. It is what ultimately motivates our use of two complementary optimization principles: one deterministic and one stochastic. Additional technical details on the Gaussian-mixture simplifications are given in Appendix C.

2.3. What Is Being Optimized?

The object being optimized in this paper is the stiffness protocol $\beta_{0\to1}$, parameterized through the finite-dimensional PWC coefficients $\beta^{(k)}$. We refer to this as "outer" optimization in order to distinguish it from the inner stochastic optimal transport problem that defines the optimal control for a fixed protocol. The point is not to re-optimize the terminal target law, which remains fixed throughout, but rather to optimize **how** the target is assembled over time.

Accordingly, the outputs of interest are not limited to a single best scalar parameter. They include the following:

- Constant schedules, used as simple and interpretable baselines;
- Nontrivial time-dependent PWC schedules;
- Inn higher-dimensional studies, families of optimized schedules that vary systematically with dimension or target complexity.

The emphasis throughout is on finite-budget intermediate-time behavior. We are interested in how target structure emerges in time, how sensitive the resulting controlled dynamics are, and how the protocol $\beta(t)$ can be designed so as to shape this transient evolution in an interpretable way. This viewpoint leads naturally to the two optimization principles introduced next: a deterministic exact-marginal objective based on control sensitivity, and a stochastic pathwise objective based on temporal memory.

3. Methodology for Schedule Design

This section describes how stiffness schedules are parameterized, evaluated, and optimized once the H-PID problem is fixed. The central methodological distinction is between objectives that can be evaluated deterministically from exact marginals and objectives that remain intrinsically pathwise and, therefore, require stochastic estimation. This distinction leads naturally to two complementary optimization principles: a deterministic objective based on velocity-gradient sensitivity, and a stochastic objective based on temporal sharpness.

3.1. Protocol Parameterization

Throughout this paper, we work with piecewise-constant (PWC) stiffness schedules:

$$\beta_t = \beta^{(k)}, \qquad t \in [t_{k-1}, t_k), \qquad k = 1, \ldots, K_\beta. \tag{10}$$

The number of intervals K_β is kept small. In the experiments below, we typically use four intervals, which is sufficient to produce nontrivial schedules while keeping the optimization low-dimensional and interpretable. This low-dimensional parameterization

is important both computationally and conceptually: it allows the resulting protocols to be optimized robustly and still remain understandable as explicit temporal design choices.

3.2. One-Time Versus Two-Time Observables

A central methodological distinction in this paper is between one-time and two-time observables.

3.2.1. One-Time Observables

These depend only on the current marginal law $p_t(x)$ and can therefore be evaluated deterministically once exact marginals are available. The main examples in this work are as follows:

- Control regularity and sensitivity measures derived from the spatial Jacobian of the control field;
- Spread-like moment quantities such as $\mathbb{E}[\|x_t\|^2]$;
- Exact-marginal energy- or entropy-type quantities.

3.2.2. Two-Time Observables

These couple states at different times along the trajectory and, therefore, do not reduce to a one-time marginal quantity. The main example in this paper is the temporal-memory function $\hat{\mathcal{A}}(t)$, which couples the current state to the terminal state and must therefore be estimated from simulated trajectories.

This distinction governs the entire schedule-design methodology. Whenever a quantity is intrinsically one-time, we evaluate it deterministically from exact marginals. Only when a quantity is genuinely pathwise do we resort to stochastic trajectory-based estimation.

3.3. Deterministic Optimization Principle: Velocity-Gradient Sensitivity

Our first optimization principle is deterministic and is based on the sensitivity of the controlled dynamics to perturbations in state space. Following the viewpoint developed in [9] and related stochastic-interpolant work [10,11], the spatial variation in the velocity field provides a useful proxy for stability, stiffness, and numerical tractability. Schedules that reduce excessive local amplification are therefore desirable.

This motivates the **velocity-gradient-sensitivity** objective

$$J_{\mathrm{VGS}}(\beta) = \int_0^1 \mathbb{E}\left[\|\nabla_x u_t^{(*)}(x;\beta)\|_F^2\right] dt, \tag{11}$$

where $\|\cdot\|_F$ denotes the Frobenius norm. The role of J_{VGS} is to penalize excessively stiff or irregular control fields and thereby favor smoother, more stable transport dynamics.

The important practical point is that J_{VGS} is built from one-time exact-marginal quantities. In the Gaussian-mixture setting, these quantities can be evaluated explicitly at each time. As a result, optimization of the VGS objective is deterministic and does not require Monte Carlo sampling.

3.4. Stochastic Optimization Principle: Regularized Transition Sharpness

Our second optimization principle is stochastic and is based on the temporal-memory observable introduced in [6]:

$$\hat{\mathcal{A}}(t;\beta_{0\to1}) \doteq \frac{\mathbb{E}_{p^{(\mathrm{path})}}\left[\hat{y}(t;x(t);\beta_{0\to1})^\top x_1\right]}{\mathbb{E}_{p^{(\mathrm{path})}}[\|x_1\|^2]} = \frac{\sum_{m=1}^M \hat{y}\left(t;x^{(m)};\beta_{0\to1}\right)^\top x_1^{(m)}}{\sum_{m=1}^M \|x_1^{(m)}\|^2}. \tag{12}$$

This quantity measures how strongly the current state reflects the eventual terminal organization, and it may be interpreted as a pathwise temporal-memory observable. As observed in [6], $\hat{\mathcal{A}}(t;\beta_{0\to 1})$ typically increases with time and, in some regimes, can exhibit a comparatively sharp transition from unstructured to target-informed behavior.

These properties motivate the sharpness objective

$$\hat{S}(\beta) = \int_0^1 \left(1 - \left|2\hat{\mathcal{A}}(t;\beta) - 1\right|\right) dt. \tag{13}$$

This objective favors trajectories in which the target structure builds up in a temporally organized way.

Unlike J_{VGS}, the sharpness objective is intrinsically multi-time. It cannot be reduced to exact-marginal deterministic evaluation and must therefore be estimated stochastically from trajectory ensembles.

On its own, $\hat{S}(\beta)$ may favor extreme stiffness regimes. To bias the transition toward a prescribed temporal window, we introduce a regularized version:

$$\hat{S}^{(\mathrm{reg})}(\beta) = \hat{S}(\beta) + \lambda^{(\mathrm{trans})}\left|t^*(\beta) - t^{(\mathrm{trans})}\right|^2, \tag{14}$$

where $t^*(\beta)$ denotes the empirically defined transition time associated with $\hat{\mathcal{A}}(t;\beta)$, and $t^{(\mathrm{trans})}$ is the desired transition time. In the current implementation, the regularization is two-sided: both too-early and too-late transitions are penalized relative to the chosen target time.

3.5. Experimental Methodology

All experiments in this paper follow the same implementation philosophy: keep the protocol space low-dimensional and interpretable, evaluate what can be computed exactly at the density level, estimate genuinely pathwise quantities only when necessary, and organize optimization through a stable hierarchical refinement.

3.5.1. Benchmark Families

We use two complementary Gaussian-mixture benchmark families:

Model A consists of random diagonal-anisotropic Gaussian-mixture targets with controlled radius, scale, and condition number. This family is used primarily for the deterministic VGS experiments. Its role is to provide a generic benchmark class in which protocol design can be studied without relying on highly structured low-dimensional symmetry.

Model B consists of structured grid/hypercube-type Gaussian-mixture targets built from the lattice $\{-a, 0, a\}^d$. In $d = 2$, we use the regular 3×3 model with nine equally weighted isotropic Gaussian modes placed on $\{-a, 0, a\}^2$. In higher dimensions, we generalize this construction by selecting a prescribed number of modes uniformly and at random from the hypercube candidate set $\{-a, 0, a\}^d$. This family is used for the sharpness-based experiments and for deterministic hypercube-type scaling studies. Its main advantage is interpretability: it supports both visually compelling low-dimensional illustrations and moderate-dimensional studies with controlled geometric structure.

3.5.2. Optimization Workflow

In all experiments, we begin by scanning the one-parameter family of constant schedules $\beta(t) \equiv \beta$. This preliminary scan serves two purposes: it exposes the baseline objective landscape, and it identifies a stable warm start for subsequent refinement.

Starting from the best constant schedule, we refine the protocol hierarchically through

$$L = 1 \ \rightarrow \ 2 \ \rightarrow \ 4 \ \rightarrow \ 8,$$

where L denotes the number of temporal subintervals in the PWC representation. At each level, the solution from the previous coarser level is prolonged to the finer grid and then improved by a lightweight coordinate-descent outer loop. This continuation-style strategy is central to the practical robustness of the implementation: it avoids entering a high-dimensional protocol space too early, while still allowing nontrivial temporal structure to emerge.

The use of coordinate descent is pragmatic rather than fundamental. The present torch-based implementation makes gradient-based optimization a natural future direction, but the current lightweight gradient-free approach is sufficient for the present study and has the advantage of remaining transparent and stable across all reported experiments.

3.5.3. Deterministic and Stochastic Evaluation

The evaluation methodology follows directly from the one-time vs. two-time distinction above.

For deterministic objectives, such as the VGS functional, the relevant quantities can be expressed through the time-marginal density and are therefore evaluated directly from exact marginals on moderate time grids. This is what makes the deterministic experiments particularly transparent: the objective is computed without Monte Carlo noise, so the resulting optimization can be interpreted directly in terms of target geometry and protocol structure.

For stochastic temporal-memory objectives, such as the sharpness criteria, the relevant observables depend on two-time correlation structure and do not reduce to one-time exact-marginal quantities. These are therefore estimated from controlled Monte Carlo trajectory simulations, typically using common random numbers to reduce variance across schedule comparisons.

Thus, the overall workflow is as follows:

- Specify a Gaussian-mixture target and a low-dimensional PWC protocol family;
- Scan the constant-β family to expose the baseline objective landscape;
- Use the best constant schedule as a warm start for hierarchical refinement $1 \rightarrow 2 \rightarrow 4 \rightarrow 8$;
- Evaluate deterministic one-time diagnostics from exact marginals and stochastic two-time diagnostics from controlled simulation when needed;
- Optimize the PWC coefficients with a lightweight coordinate-descent routine;
- Interpret the resulting schedules through both scalar diagnostics and representative low-dimensional visualizations.

For the deterministic **Model A + VGS** experiments, the primary reported quantity is the dimension-normalized velocity-gradient-sensitivity objective (11). When useful, we also report temporal-memory summaries such as $t_{\mathrm{cross}}(\hat{\mathcal{A}} = 0.8)$, not as optimized quantities in this model family but as auxiliary descriptors of how the resulting protocol affects temporal organization.

For the stochastic **Model B + sharpness** experiments, the primary reported quantities are the sharpness (13) and its regularized version (14). These scalar diagnostics are supported by direct plots of $\hat{\mathcal{A}}(t)$ and $\beta(t)$, which are especially informative in the low-dimensional structured examples.

In the earlier version of the manuscript [12] we discussed a broader quality-of-sampling suite, including path cross entropy, cost-to-go decompositions, drift–diffusion

balance, Langevin mismatch, speciation timing, and related quantities. We refer readers interested in such exploratory diagnostics to that earlier version.

3.5.4. Experimental Scope

The experiments reported in this paper compare only the protocol families that remain central to the revised narrative:

- Constant β schedules;
- Deterministic VGS-optimized PWC schedules;
- Sharpness- or regularized-sharpness-optimized PWC schedules.

Constant schedules are not included merely as trivial baselines. They are a core part of the methodology: they expose the basic objective landscape; reveal whether the relevant objective favors small-, intermediate-, or large-β regimes; and generate the warm starts for hierarchical refinement. The optimized PWC schedules then show how additional temporal structure can be exploited beyond the constant family.

All of the experiments are designed to remain computationally controlled and laptop-feasible, not because of any intrinsic hardware limitation but because this is part of the methodological philosophy of the paper. The central point is to preserve as much analytic and computational control as possible while still allowing meaningful target-dependent protocol design.

4. Results

4.1. Model A: Deterministic VGS on Random Anisotropic GMMs

Model A uses the random diagonal-anisotropic benchmark family with fixed parameters $K = 4$, radius 3.0, scale 0.6, and condition number 20.0. The main paper notebook studies $d = 2$ and $d = 8$, and the associated scaling notebook extends the deterministic analysis to $d = 4$ and $d = 16$. Constant-β scans are carried out over $[0.05, 20]$, and hierarchical refinement uses the levels $L = 1, 2, 4, 8$.

Figure 1 gives a more dynamical view of the deterministic protocol-design problem in the representative $d = 2$ Model A instance. The top row shows the evolution induced by the best constant schedule, while the bottom row shows the evolution under the best VGS-optimized PWC-8 protocol. In both cases, the target law is the same; what changes is the transient path through probability space. This is precisely the point of the protocol-design framework: even when the terminal target distribution is fixed, the intermediate marginal evolution can still be shaped in a nontrivial way. The comparison in Figure 1 makes this concrete. The optimized PWC protocol does not simply reproduce the best constant schedule with a slightly different scalar score; it induces a visibly different temporal organization of the marginal cloud as it evolves toward the target.

Figure 2 summarizes the deterministic protocol-design logic for Model A in its simplest nontrivial setting. The left panel shows that even within the constant family $\beta(t) \equiv \beta$, the VGS objective has a clear and non-flat dependence on β, so the choice of protocol stiffness already matters at the one-parameter level. This constant-β scan therefore plays a dual role: it provides an interpretable baseline, and it identifies the warm start for subsequent refinement. The right panel then shows what is gained by allowing temporal structure. As the hierarchy progresses from $L = 1$ to $L = 2, 4, 8$, the optimized schedule departs from the constant baseline and develops nontrivial time dependence. Thus, the main lesson of Figure 2 is that deterministic protocol design is already meaningful in low dimensions: even when the terminal target law is fixed, the transient control field can be improved by introducing a modest amount of temporal flexibility into $\beta(t)$.

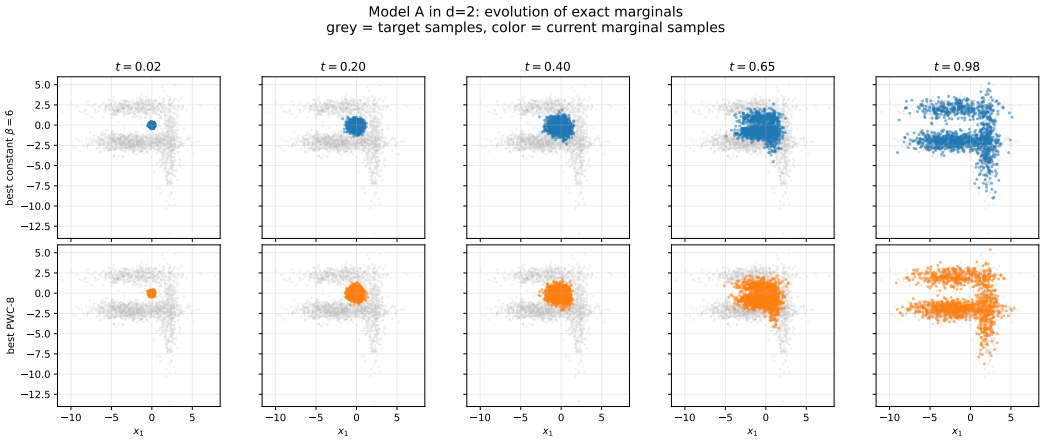


Figure 1. Model A in $d = 2$: evolution of exact marginals under the best constant schedule (**top row**) and the best optimized PWC-8 schedule (**bottom row**). In each panel, grey points show samples from the target density, while colored points show samples from the current marginal at the indicated time. The top row corresponds to the best constant-β baseline, and the bottom row to the best hierarchical VGS-optimized PWC protocol.

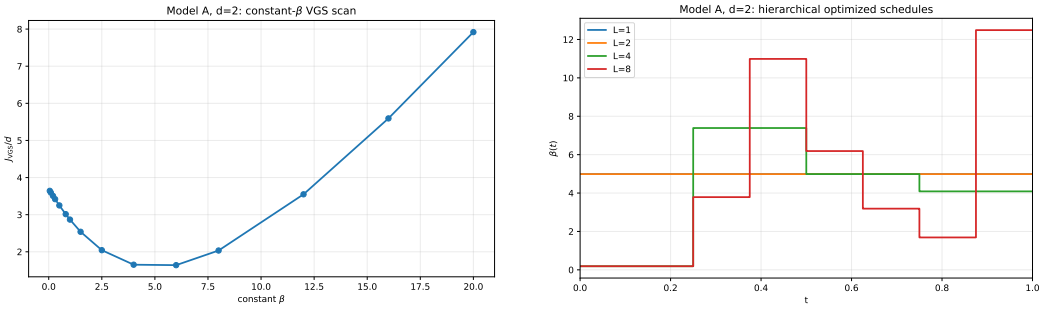


Figure 2. Model A in $d = 2$: (**left**), constant-β scan of the deterministic VGS objective; (**right**), hierarchical optimized step-function schedules $\beta(t)$ for $L = 1, 2, 4, 8$. The figure emphasizes that nontrivial optimized schedules emerge already in low dimensions.

The deterministic optimization logic remains meaningful beyond the low-dimensional visual example. For Model A, the $d = 8$ study confirms that the same workflow—constant-β scan, warm start from the best constant schedule, and hierarchical refinement through L=1,2,4,8—continues to produce interpretable protocol structure in moderate dimensions. Therefore, the point of the $d = 8$ experiment is not visual aesthetics but methodological robustness: the VGS-based protocol-design problem remains well posed, the exact-marginal deterministic evaluation remains available, and the optimized PWC schedule still improves on the best constant baseline.

Table 1 summarizes the main quantitative comparison for Model A. The $d = 2$ and $d = 8$ cases should be read together. The $d = 2$ example shows most clearly that nontrivial optimized schedules emerge already in low dimensions, while the $d = 8$ example shows that this is not merely a low-dimensional artifact. In both cases, the best PWC-8 schedule improves on the best constant schedule with respect to the deterministic VGS objective, while still remaining easy to interpret as a time-dependent refinement of the constant baseline. This is exactly the kind of behavior one would want from a protocol-design methodology: the optimization is strong enough to uncover nontrivial temporal structure, but controlled enough that the resulting schedules remain understandable rather than opaque.

Table 1. Model A comparison of the best constant schedule and the optimized PWC-8 schedule in $d = 2$ and $d = 8$. The optimized PWC protocol improves the deterministic VGS objective in both cases. Full optimized schedules are reported in the accompanying notebooks, available at: https://gihub.com/mchertkov/ShortAdaPID (accessed on 19 April 2026).

d	Best Const. β	Const. J_{VGS}/d	PWC-8 J_{VGS}/d	Improvement (%)	t_{cross} Shift
2	6.0	1.64027	1.51286	7.77	$0.470 \to 0.458$
8	2.5	0.87640	0.84375	3.72	$0.543 \to 0.550$

A broader deterministic scaling study further supports this interpretation. In that study, the same Model A family was examined over $d \in \{2, 4, 8, 16\}$, again using constant-β scans over $[0.05, 20]$ and the same hierarchical $1 \to 2 \to 4 \to 8$ refinement strategy. The purpose of that study was not to claim a universal scaling law but to demonstrate that the deterministic evaluation pipeline remains computationally manageable, and that constant baselines remain meaningful across dimensions. We therefore view the scaling evidence as methodological rather than asymptotic: the exact-marginal VGS framework continues to function in a controlled way as d increases over the tested range. Additional scaling experiments involving changes in target complexity, including scans over the number of modes K for hypercube-type targets, are included in the accompanying notebooks, available at: https://github.com/mchertkov/ShortAdaPID (accessed on 19 April 2026).

Taken together, the Model A experiments support three conclusions: First, deterministic VGS optimization is already informative in low dimensions and produces nontrivial optimized PWC protocols. Second, the same protocol-design logic remains meaningful in moderate dimensions, as demonstrated explicitly in $d = 8$. Third, the exact-marginal deterministic framework is computationally controlled enough to support systematic scans in dimension and in target complexity as well. This makes Model A the clean methodological backbone of this paper: it shows that protocol design can be posed, optimized, and interpreted at the density level without sacrificing computational tractability.

4.2. Model B: Sharpness-Based Optimization on Structured Grid/Hypercube GMMs

We now turn to Model B, where the emphasis shifts from deterministic control regularity to temporal organization, as measured by sharpness-based observables. In contrast to Model A, whose main role was methodological genericity, Model B is designed to expose how the geometry of the target distribution influences the optimization landscape itself. The experiments below therefore serve two related purposes: first, to show in a visually transparent setting that sharpness-based optimization can produce nontrivial time-dependent protocols; and second, to show that the resulting dependence on β is genuinely target-dependent and, hence, non-universal.

We begin with the regular 3×3 model in $d = 2$, which is the clearest visual demonstration of the sharpness-based protocol-design idea. This is the natural place to explain the full logic of the optimization: we first scan the constant family $\beta(t) \equiv \beta$, then use the best constant schedule as a warm start, and finally refine hierarchically through $L = 1, 2, 4, 8$. Because the target geometry is so transparent in this case, both the scalar objectives and the evolving densities admit a direct visual interpretation.

Figure 3 provides the most direct dynamical view of what the sharpness objective is optimizing. The top row shows the evolution induced by the best constant schedule, while the bottom row shows the evolution under the best optimized PWC-8 protocol. In both rows, the terminal target law is the same; what differs is the transient organization of the marginal cloud. This is precisely the aspect of the dynamics that the sharpness objective is meant to probe. Thus, the optimized protocol is not merely better in the scalar objective but visibly different in how probability mass is reorganized in time on the way to the same

final target. This figure is therefore best read as the qualitative, density-level counterpart of the scalar sharpness curves discussed next.

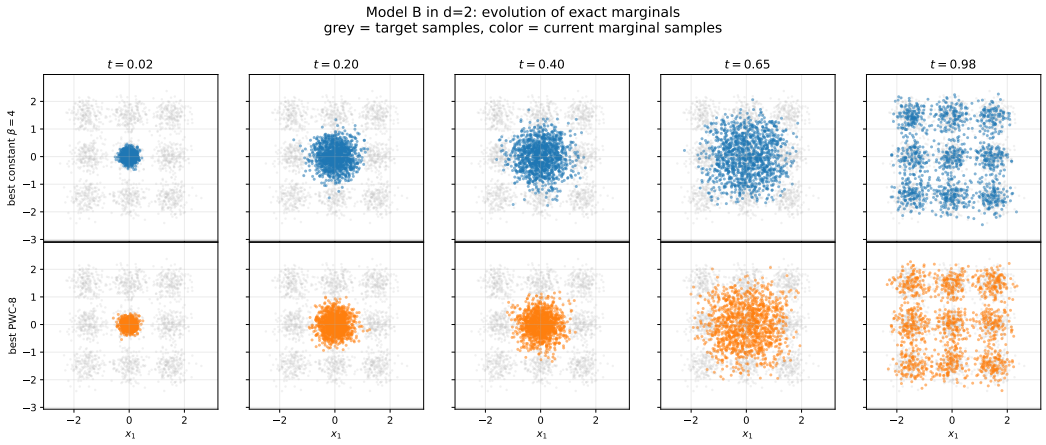


Figure 3. Model B in $d = 2$: evolution of exact marginals under the best constant schedule (**top row**) and the best optimized PWC-8 schedule (**bottom row**). In each panel, grey points show samples from the target density, while colored points show samples from the current marginal at the indicated time. The top row corresponds to the best constant-β baseline, and the bottom row to the best regularized-sharpness-optimized piecewise-constant protocol.

Figure 4 explains how that nontrivial protocol emerges. The left panel shows that even within the one-parameter constant family, the sharpness landscape is already structured: neither $\hat{S}(\beta)$ nor $\hat{S}^{(\mathrm{reg})}(\beta)$ is flat in β, and the regularized objective selects a meaningful operating region rather than a trivial boundary value. The right panel then shows that once temporal flexibility is introduced, the optimizer uses it: the protocol departs from the constant baseline and develops nontrivial time dependence under hierarchical refinement. The main lesson of Figure 4 is therefore the same as in the dynamical visualization, but now at the optimization level: the sharpness objective does not merely rank constant schedules; it supports genuinely time-structured protocol design.

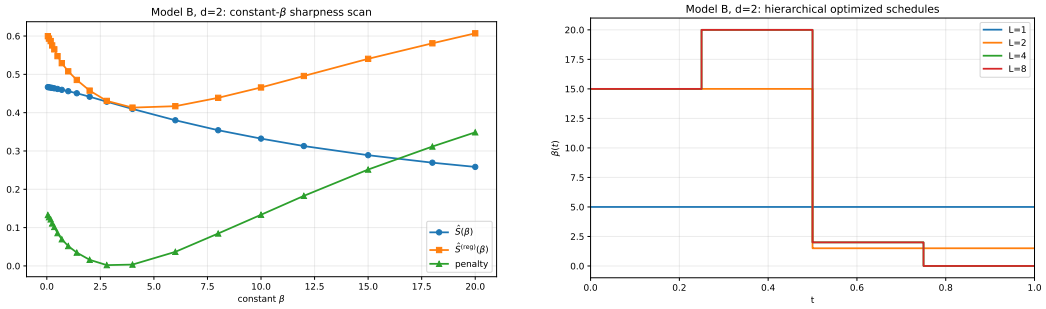


Figure 4. Model B in $d = 2$: (**left**), $\hat{S}(\beta)$ and $\hat{S}^{(\mathrm{reg})}(\beta)$ as functions of constant β; (**right**), hierarchical optimized step-function schedules $\beta(t)$ for $L = 1, 2, 4, 8$. The figure emphasizes that regularized sharpness yields a nontrivial optimized PWC protocol in the visually interpretable $d = 2$ setting.

Figure 5 makes the same point in the reduced one-dimensional temporal-memory observable. The optimized protocol does not simply lower $\hat{S}^{(\mathrm{reg})}$ numerically; it changes the way temporal correlation builds up. This is important because it confirms that the sharpness-based optimization is not exploiting a numerical artifact of the scalar objective. Rather, it is selecting a protocol with a genuinely different temporal organization, and $\hat{A}(t)$ provides a compact summary of that difference. Taken together, Figures 3–5 establish the main low-dimensional conclusion of this subsection: in the structured $d = 2$ benchmark, regularized sharpness produces a nontrivial optimized protocol that is interpretable both geometrically and temporally.

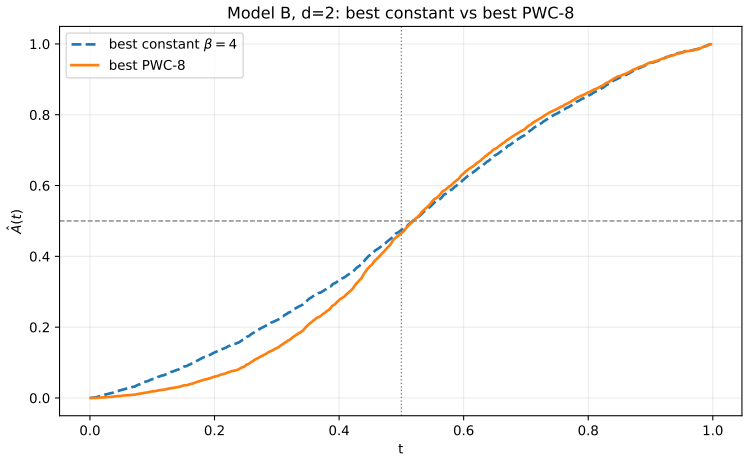


Figure 5. Model B in $d = 2$: temporal-memory curves $\hat{A}(t)$ for the best constant schedule and the best optimized PWC-8 schedule under regularized sharpness.

We next turn to the hypercube-random $K = 9$ case in $d = 8$. This example plays a different role. It is not primarily a visual showcase but a test of target-dependence and non-universality. Here, the key object is the constant-β scan itself, because it reveals the shape of the sharpness landscape before any hierarchical refinement is attempted.

Figure 6 shows that, in $d = 8$, the dependence of both $\hat{S}(\beta)$ and $\hat{S}^{(\mathrm{reg})}(\beta)$ on the constant schedule is nontrivial and can exhibit clear competition between small-β and large-β regimes. This is the essential evidence for our claim that the sharpness landscape is not universal. There is no single stiffness regime that dominates across targets. Instead, the shape of the objective depends substantially on the geometry of the target GMM. For the selected $d = 8$ instance, the constant scan already reveals this competition, and this is the main message that we need from the higher-dimensional sharpness experiment in this paper.

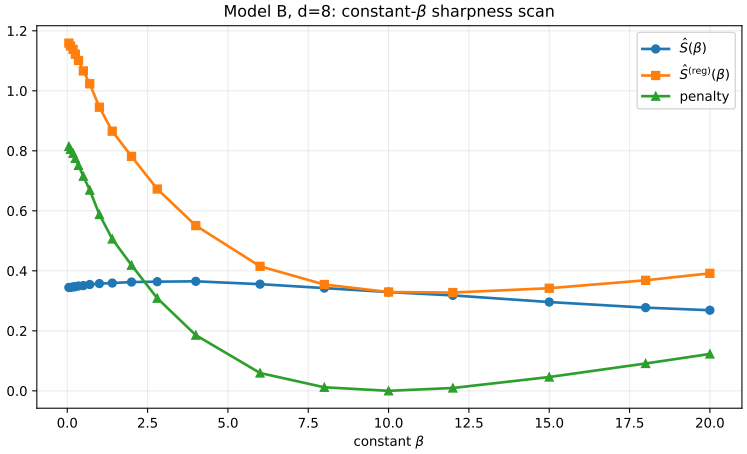


Figure 6. Model B in $d = 8$: constant-β sharpness scans for the selected hypercube-random $K = 9$ instance. This figure supports the claim that the β-dependence of sharpness is target-dependent and may involve competing small-β and large-β regimes.

As with Model A, we show only the most significant results. The full schedule plots, optimization traces, and auxiliary diagnostics are available at: https://github.com/mchertkov/ShortAdaPID (accessed on 19 April 2026).

The experiments discussed in this subsection therefore support two complementary conclusions: At the level of individual examples, the regular 3×3 $d = 2$ benchmark shows that regularized sharpness yields a genuinely nontrivial optimized PWC protocol, visible

both in the evolving densities and in the temporal-memory curve $\hat{A}(t)$. At the collective level, the comparison between the $d = 2$ and $d = 8$ cases shows that sharpness-based protocol design is inherently target-dependent: the optimization landscape may favor different stiffness regimes for different targets, and the resulting protocol structure is not universal. We view this non-universality not as a weakness but as one of the main empirical insights of the Model B study. It suggests that temporal-memory-based objectives are sufficiently expressive to register meaningful geometric differences between targets, and that they therefore deserve broader systematic study in future work.

4.3. Deterministic Scaling on Hypercube-Type Targets

This subsection extends the reports of the two preceding subsections with the deterministic VGS study to hypercube-type targets and plays a supporting but important role in the experimental narrative. Its purpose is not to introduce a new objective but to test how the deterministic protocol-design pipeline behaves when the target geometry becomes more combinatorial and when either the ambient dimension or mode count is increased. In this sense, this subsection complements both of the earlier parts of the Results: it uses the deterministic VGS objective of Model A but applies it to the more structured hypercube-type targets associated with Model B.

Two families of scans are considered here: The first is a **dimension scan**, where we fix the number of modes at $K = 9$ and vary the ambient dimension over $d \in \{2, 4, 8, 16\}$. The second is a **mode-count scan**, where we fix the ambient dimension at $d = 8$ and vary the number of modes over $K \in \{9, 18, 27, 36, 45\}$. In both families, the optimization workflow is the same as before: we first scan the constant family $\beta(t) \doteq \beta$, then use the best constant schedule as a warm start, and finally refine hierarchically through $L = 1, 2, 4, 8$ using coordinate descent. Thus, the only thing that changes across this subsection is the structure of the target family itself.

Figure 7 summarizes the main deterministic scaling evidence. The left panel shows the effect of increasing ambient dimension at fixed target complexity $K = 9$, while the right panel shows the effect of increasing target complexity at a fixed ambient dimension $d = 8$. In both panels, the comparison is between the best constant schedule and the best optimized PWC-8 schedule under the deterministic VGS objective. The key qualitative message is that the optimization pipeline remains computationally controlled and interpretable in both directions of variation. As either d or K increases, the constant-schedule baseline remains meaningful, the hierarchical optimization remains executable, and the refined PWC protocol continues to provide a coherent comparison point.

At the same time, the purpose of Figure 7 is deliberately modest. We do not read these experiments as evidence for a universal asymptotic scaling law. Rather, we read them as evidence of methodological robustness. The exact-marginal deterministic framework can still be run on moderate-sized problems of varying structure, and the resulting optimization output remains interpretable enough to support systematic comparison. That is the main conclusion that this subsection is intended to support.

To keep the main paper focused, Figure 7 presents only the compact summary comparison. The full families of constant-β scans, hierarchical schedule profiles $\beta(t)$, selected temporal-memory comparisons, and auxiliary tables for both scan families are not shown here but are available at: https://github.com/mchertkov/ShortAdaPID (accessed on 19 April 2026). Those Materials are useful for reproducibility and for readers interested in the detailed dependence of the optimized schedules on d and K, but they are not needed for the central narrative of this paper.

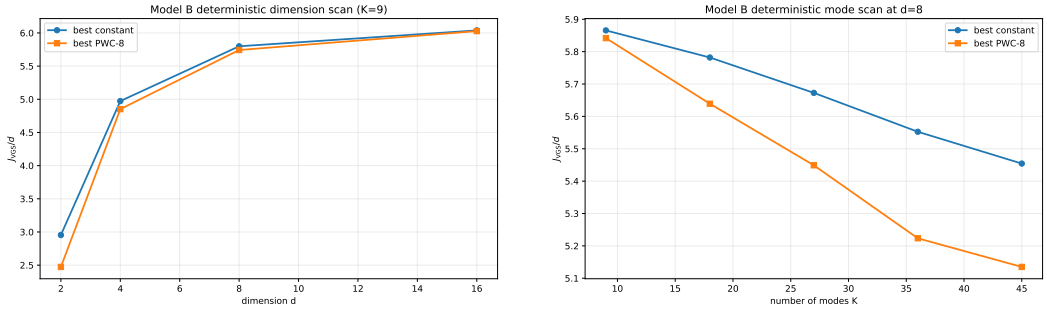


Figure 7. Deterministic VGS scaling on hypercube-type targets. (**Left**): best constant versus best PWC-8 for the dimension scan at fixed $K = 9$. (**Right**): best constant versus best PWC-8 for the mode-count scan at fixed $d = 8$. Together, the panels summarize the computationally controlled behavior of the deterministic protocol-design pipeline as either d or K increases.

The collective conclusion of this subsection is therefore clear. The deterministic VGS framework is not confined to a single low-dimensional illustrative example. It remains computationally manageable under both increasing ambient dimension and increasing target complexity, at least over the experimentally tested range. This strengthens the broader conclusion already suggested by Model A: the proposed protocol-design methodology is not only interpretable in simple examples but also sufficiently robust to support structured scaling studies on more complex Gaussian-mixture targets.

5. Conclusions, Positioning, and Outlook

Harmonic PID provides a tractable and expressive framework for schedule design beyond terminal-law matching. The central contribution of this paper is to turn that tractable structure into a practical methodology for protocol optimization. Rather than treating the stiffness protocol $\beta(t)$ as a fixed background ingredient, we treat it as a design variable that shapes the transient evolution of probability mass while preserving the terminal target law. This is the main conceptual shift of this paper.

Our methodological formulation relies on a clean distinction between two classes of observables. One-time quantities, such as velocity-gradient sensitivity, can be evaluated deterministically from exact time marginals in the Gaussian-mixture setting. In contrast, two-time quantities, such as temporal-memory observables derived from $\hat{\mathcal{A}}(t)$, remain genuinely pathwise and must be estimated stochastically from trajectories. This separation is not merely technical; it is what allows us to combine exact-marginal deterministic optimization with controlled stochastic optimization in a single coherent schedule-design framework.

Our experiments support three main conclusions:

1. Nontrivial protocol design is already meaningful in the analytically controlled setting studied here. For Model A, deterministic VGS optimization produces non-constant optimized PWC schedules already in $d = 2$, where the effect can be seen directly in the evolving exact marginals, and remains meaningful in $d = 8$, where the main point is methodological robustness rather than visual intuition. This makes the exact-marginal VGS framework the clean backbone of this paper: it shows that protocol design can be posed, optimized, and interpreted at the density level without sacrificing computational tractability.
2. Sharpness-based objectives reveal a different and complementary aspect of the problem. In the structured $d = 2$ Model B benchmark, regularized sharpness yields a genuinely nontrivial optimized protocol that is interpretable both geometrically and temporally. The evolving densities, the optimized $\beta(t)$, and the corresponding $\hat{\mathcal{A}}(t)$ curves all show that the protocol is not merely lowering a scalar objective numerically but is changing the temporal organization of the path itself. At the same time,

the $d = 8$ hypercube-random experiments show that the dependence on β is not universal: different target geometries may favor substantially different stiffness regimes, and in some cases there is visible competition between small-β and large-β behavior. We view this non-universality as one of the main empirical insights of this paper. It suggests that temporal-memory-based objectives are sensitive enough to register meaningful geometric differences between targets and, therefore, deserve broader systematic study.

3. The hierarchical constant-scan $\rightarrow$ warm-start $\rightarrow 1 \rightarrow 2 \rightarrow 4 \rightarrow 8$ refinement strategy is effective and scalable enough for the present study. Across all reported notebooks, the same optimization logic is used successfully: first, scan the constant family, and then refine temporally through a small hierarchy of PWC schedules. This makes the optimization path interpretable and stable, and it provides a natural bridge between simple baselines and genuinely time-dependent protocols. The accompanying deterministic scaling study on hypercube-type targets further shows that this framework remains computationally manageable when either the ambient dimension or the target complexity is increased over the tested range.

In this sense, this paper is best positioned not as a claim about a single universally optimal schedule objective but as a methodology paper. Different objectives expose different aspects of the transient dynamics, and the appropriate notion of schedule quality is application-dependent. The exact-marginal VGS objective emphasizes control regularity and deterministic tractability. The sharpness-based objective emphasizes temporal organization and pathwise memory. The key point is that Harmonic PID is rich enough to support both viewpoints within a common analytically transparent framework.

Several directions for future work are immediate.

Gradient-Based Outer Optimization:

The present implementation uses hierarchical warm-started coordinate descent, which proved stable and sufficiently expressive for all experiments reported here. However, the codebase is already Torch-based, and a natural next step is to refactor the protocol-construction layer so that gradients with respect to the PWC coefficients propagate through the full evaluation pipeline.

Richer Protocol Families:

We restricted attention to low-dimensional PWC schedules because they provide a clean compromise between expressiveness and interpretability. A natural extension is to consider smoother or higher-resolution parameterizations of $\beta(t)$, adaptive interval placement, or protocol families with explicit regularity constraints.

Broader Objective Design:

The present paper focuses on two complementary principles: deterministic velocity-gradient sensitivity, and stochastic temporal sharpness. A broader future program would develop additional objectives that connect more directly to specific application goals—for example, accelerated target assembly, robustness to multimodality, or controlled preservation/forgetting of temporal information.

More Systematic Scaling and Geometry Studies:

The current experiments already show that the target geometry matters, and that the resulting protocol landscape is not universal. A natural next step is therefore a more systematic empirical and theoretical study of how optimized schedules depend on the dimension, number of modes, anisotropy, and geometric arrangement of the target mixture components.

Beyond Gaussian Mixtures:

Gaussian-mixture targets are especially valuable because they preserve exact-marginal tractability, which is essential for the present methodological study. At the same time, one of the long-term goals is to transfer these schedule-design principles to broader classes of target distributions.

Overall, the main conclusion of this paper is that **the stiffness protocol** $\beta(t)$ should be viewed as a meaningful design variable, not merely as a passive scalar regularization parameter. Even when the terminal target distribution is fixed, different protocol objectives lead to different optimized schedules and reveal different notions of what constitutes a "good" path to the same endpoint. Harmonic PID provides a rare setting in which such questions can be studied with both theoretical transparency and computational control, and this is what makes it a useful laboratory for schedule design in diffusion bridge-based generative modeling.

Author Contributions: M.C. developed the methodology, conducted the experiments, and prepared the manuscript. H.B. contributed to methodological development, evaluated the experiments, helped down-select the techniques, and improved the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially supported by M.C.'s start-up funds at the University of Arizona. A substantial portion of the research was conducted while M.C. was partially supported by the Mini-Sabbatical Faculty Fellow Program at Lawrence Livermore National Laboratory (LLNL).

Data Availability Statement: In the interest of reproducibility, all code used to generate the figures and experiments reported in this paper is available at: https://github.com/mchertkov/ShortAdaPID (accessed on 19 April 2026).

Acknowledgments: M.C. is thankful to José Blanchet and Youssef Marzouk for stimulating discussions and for their guidance on the broader development of PID and related methodologies. M.C. gratefully acknowledges the scientific engagement, and encouragement of colleagues in the Center for Applied Scientific Computing (CASC) at LLNL. The authors used large language models to assist with editing, code refactoring, and the organization of computational experiments. All mathematical derivations, scientific claims, and final code were independently checked, validated, and approved by the authors.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Path Integral Diffusion (PID)

This technical appendix recaps the linearly solvable Path Integral Diffusion (PID) framework [6]. Appendix A.1 clarifies how accounting for the time-dependent quadratic potential modifies the resulting dynamics.

From Sampling to Stochastic Optimal Transport

Let $x_t \in \mathbb{R}^d$ evolve according to the controlled SDE (2), and denote by p_t its marginal density. We prescribe $p_0(x) = \delta(x)$ and $p_1(x) \propto \exp[-E(x)]$, and we pose the question of inferring the score field $u_{0\to 1}(\cdot)$. PID casts the search for u as the **stochastic optimal transport** (SOT) problem Equation (1), subject to the dynamics (2) and the terminal constraint $p_1 \propto e^{-E}$. Throughout this section, we focus on the low-integrability setting of [6], where the gauge and drift vector fields vanish ($A = f = 0$) and the potential is isotropic and quadratic, $V_t(x) = \beta x^2/2$ with $\beta > 0$.

Linearly Solvable Structure

Problem (1) becomes tractable because the optimal score field can be expressed in closed form once two linear PDEs are solved:

$$-\partial_t G_t^{(-)} + V_t(x)\, G_t^{(-)} = \tfrac{1}{2}\Delta_x G_t^{(-)}, \quad G_1^{(-)}(x;y) = \delta(x-y), \tag{A1}$$

$$\partial_t G_t^{(+)} + V_t(x)\, G_t^{(+)} = \tfrac{1}{2}\Delta_x G_t^{(+)}, \quad G_0^{(+)}(x;y) = \delta(x-y). \tag{A2}$$

With $G^{(\pm)}$ at hand, we arrive at Equation (6). Because the control is recovered from linear PDEs, the problem is said to be linearly solvable [8].

Constant Stiffness: Analytic Benchmark

For $V_t(x) = \frac{1}{2}\beta x^2$ with non-negative constant β, the Green functions are Gaussian, and Equations (A1) and (A2) yield an explicit analytic score [6] (see Appendix A.1 for details):

$$u_t^*(x;\beta) = b_t^{(-)}\hat{y}(t;x;\beta) - a_t^{(-)}x = -a_t^{(-)}(x - \tilde{y}(t;x;\beta)), \tag{A3}$$

$$\hat{y}(t;x;\beta) = \mathbb{E}_{y\sim p^{(\mathrm{probe})}(\cdot;x;\beta)}[y], \quad p^{(\mathrm{probe})}(y;x;\beta) \propto \exp(-E(y) - \Delta(t;x;y;\beta)), \tag{A4}$$

$$\Delta(t;x;y;\beta) = \frac{\mathcal{K}_t}{2}\left\| y - \frac{b_t^{(-)}}{\mathcal{K}_t}x \right\|^2 - \frac{d}{2}\log\left(\frac{\mathcal{K}_t}{2\pi}\right), \quad \text{where} \quad \mathcal{K}_t \doteq c_t^{(-)} - a_1^{(+)}, \tag{A5}$$

$$a_t^{(+)} = \sqrt{\beta}\coth(t\sqrt{\beta}),\ a_t^{(-)} = c_t^{(-)} = \sqrt{\beta}\coth\big((1-t)\sqrt{\beta}\big),\ b_t^{(-)} = \frac{\sqrt{\beta}}{\sinh((1-t)\sqrt{\beta})}. \tag{A6}$$

Note that, at $\beta \to 0$, Equation (A6) becomes

$$a_t^{(+)} = \frac{1}{t}, \quad a_t^{(-)} = c_t^{(-)} = b_t^{(-)} = \frac{1}{1-t}. \tag{A7}$$

Why Time–Dependent Stiffness?

A fixed β cannot simultaneously encourage exploration at early times and enforce contraction near $t = 1$. Allowing β_t to vary introduces exactly this flexibility. Appendix A.1 derives the Gaussian Green functions for a general schedule $\beta_{0\to 1}$.

Appendix A.1. Time-Dependent Quadratic Potential

Here, we continue to discuss the case with $A = f = 0$ and isotropic quadratic/harmonic but **time-dependent** potential (4), where $\beta_t \geq 0$. This case is more general than the "low"-level integrability setting discussed in [6] and also briefly reviewed above. This case is of interest to study because of the intuition gained on the benefit of different but time-independent β. The intuition—applied to the goal of building the target distribution discussed in [6]—suggests that smaller β at the beginning and larger β closer to the end may be beneficial for building the target distribution gradually. Indeed, as we start at $x_0 = 0$, small β at early stages will encourage faster spread—this is the stage of exploration. On the contrary, when you are closer to the end, you want to limit exploration, thus increasing β and allowing the model to fine-tune—in this regime, the terminal term bias, guaranteeing that we are getting to the target distribution, will be even stronger. Actually, if true, this logic suggests that the largest t does not matter much, but we need to increase β at some intermediate times—to stabilize memorization—locking to the final image at the point of phase transition, which should probably happen somewhere in the middle of the time interval.

In this regime of an isotropic but time-dependent quadratic potential, the differential relations from Appendix B of [6] still apply, except that the constant stiffness β is now

replaced by the schedule β_t. We therefore obtain (in the following, and when not confusing we drop dependence on β to simplify notations)

$$G_t^{(-)}(x;y) \propto \exp\left(-\tfrac{a_t^{(-)}}{2}x^2 + b_t^{(-)}x^\top y - \tfrac{c_t^{(-)}}{2}y^2\right), \quad G_t^{(+)}(y;0) \propto \exp\left(-\tfrac{a_t^{(+)}}{2}y^2\right), \tag{A8}$$

$$\mp\dot{a}_t^{(\pm)} + \beta_t = \left(a_t^{(\pm)}\right)^2,\ \dot{b}_t^{(-)} = a_t^{(-)}\, b_t^{(-)},\ \dot{c}_t^{(-)} = (b_t^{(-)})^2, \tag{A9}$$

where $\propto$ means that we dropped the t-dependent multiplier.

Expressions for $u_t^*(x;\beta_{0\to1})$, $\hat{y}(t;x;\beta_{0\to1})$, and $\Delta(t;x;y;\beta_{0\to1})$—originally introduced in Equations (A3)–(A5) for the case of the constant β—stay exactly the same in the general case of the time-dependent $\beta_{0\to1}$, provided that we substitute original explicit expressions for $a_t^{(\pm)}, b_t^{(-)}$, and $c_t^{(-)}$ set in Equation (A6) by a generally implicit representation via Equation (A9).

Recall that the Green functions must satisfy the boundary conditions—$G_1^{(-)}(x;y) = \delta(x-y)$ and $G_0^{(+)}(y;0) = \delta(y)$, resulting in the following asymptotic requirements:

$$t \to 1^-: \quad a_t^{(-)}, b_t^{(-)}, c_t^{(-)} \to \frac{1}{1-t}, \tag{A10}$$

$$t \to 0^+: \quad a_t^{(+)} \to \frac{1}{t}. \tag{A11}$$

Note that combining Equation (A3) with the $t \to 0$ asymptotics (A10) and accounting for the fact that $x_0 = 0$, we estimate that at sufficiently small t, $u_t^*(x;\beta) \approx b_0^{(-)}\bar{y}$, where $\bar{y} = \mathbb{E}p_1[y]$ is the first moment of the target distribution (which we may assume to be known).

Piece-Wise-Constant β_t: Analytic Solution

This case is advantageous, as it allows for a fully analytic solution. Let us split the time domain into K equally spaced time intervals and then use the piecewise-constant (PWC) profile:

$$t \in [(k-1)/K, k/K],\ k = 1, \cdots, K: \quad \beta_t = \beta_k. \tag{A12}$$

Then, assuming continuity in time, and taking into account the asymptotic behavior of $a_t^{(-)}, b_t^{(-)}, c_t^{(-)}$ at $t \to 1$ documented in Equation (A10) and of $a_t^{(+)}$ at $t \to 0$ documented in Equation (A11), we derive

$$t \in [1-1/K, 1]: \ a_t^{(-)} = c_t^{(-)} = \sqrt{\beta_K}\coth\left((1-t)\sqrt{\beta_K}\right),\ b_t^{(-)} = \frac{\sqrt{\beta_K}}{\sinh\left((1-t)\sqrt{\beta_K}\right)}, \tag{A13}$$

$$t \in \left[(k-1)/K,\, k/K\right],\ k = K-1, \ldots, 1: \tag{A14}$$

$$a_t^{(-)} = \sqrt{\beta_k}\,\frac{a_{k/K}^{(-)} + \sqrt{\beta_k}\tanh\left(\sqrt{\beta_k}(k/K-t)\right)}{\sqrt{\beta_k} + a_{k/K}^{(-)}\tanh\left(\sqrt{\beta_k}(k/K-t)\right)},$$

$$b_t^{(-)} = b_{k/K}^{(-)}\sqrt{\frac{\beta_k - \left(a_t^{(-)}\right)^2}{\beta_k - \left(a_{k/K}^{(-)}\right)^2}},$$

$$c_t^{(-)} = c_{k/K}^{(-)} + \frac{\left(b_{k/K}^{(-)}\right)^2}{\beta_k - \left(a_{k/K}^{(-)}\right)^2}\left(a_{k/K}^{(-)} - a_t^{(-)}\right),$$

$$t \in [0, 1/K]: \quad a_t^{(+)} = \sqrt{\beta_1}\coth\left(t\sqrt{\beta_1}\right), \tag{A15}$$

$$t \in \left[k/K, (k+1)/K\right],\ k = 1, \ldots, K-1: \tag{A16}$$

$$a_t^{(+)} = \sqrt{\beta_{k+1}}\,\frac{1+\rho_{k+1}(t)}{1-\rho_{k+1}(t)}, \qquad \rho_{k+1}(t) = \exp\left(-2\sqrt{\beta_{k+1}}\,(t-k/K)\right)\frac{a_{k/K}^{(+)} - \sqrt{\beta_{k+1}}}{a_{k/K}^{(+)} + \sqrt{\beta_{k+1}}}.$$

Remark on how the formulae are derived and stated. For the "minus" branch $\dot{a} = a^2 - \beta$, the fractional-linear variable $u = (a - \sqrt{\beta})/(a + \sqrt{\beta})$ satisfies $\dot{u} = +2\sqrt{\beta}\, u$, which yields the tanh/Möbius form used in (A14). For the "plus" branch $\dot{a} = \beta - a^2$, the same u obeys $\dot{u} = -2\sqrt{\beta}\, u$, giving the exponential Möbius map in (A16).

Appendix B. Representations of Probability Densities

To set up a proper metric for this paper's main goal—**choosing optimal** $\beta_{0\to 1}$—we utilize and juxtapose to each other the following (probability) densities, all defined for a given (fixed) $\beta_{0\to 1}$:

- The **target** density itself—$p^{(\mathrm{tar})}(y)$, defined according to Equation (3) up to the normalization factor (partition function) by the explicitly given energy function $E(\cdot)$.
- Probability density of a **path** $p^{(\mathrm{path})}(x_{0\to 1}|u^*_{0\to 1}(\cdot;\beta_{0\to 1}))$ generated according to the SODE (2), with $u_{0\to 1}(x_{0\to 1}) \doteq (u_t(x_t)|t\in[0,1])$ substituted by the optimal protocol (conditioned to $\beta_{0\to 1}$) and defined in Equation (6). We will also be working with the **marginal** density of the state x_t observed at the moment of time t and, thus, defined according to the following path integral:

$$p^{(\mathrm{mar})}(x_t|t;u^*_{0\to t}(\cdot;\beta_{0\to 1})) = \int p^{(\mathrm{path})}(x_{0\to 1}|u^*_{0\to 1}(\cdot;\beta_{0\to 1}))\mathcal{D}x_{0^+\to t^-}, \tag{A17}$$

where we utilize the standard (in theoretical physics) notations for the path integral differential:

$$\mathcal{D}x_{0^+\to t^-} = \lim_{N\to\infty}\prod_{n=1}^{N-1} dx_{t_n}, \quad t_n = t\frac{n}{N}.$$

Given dependence of the definitions of the path density and its marginals on the SODE (2), the densities will be represented empirically via generated sample paths.

- The density of the final state as **probe** at the time t, given that the system is observed in the state x, is another important object of the H-PID theory:

$$p^{(\mathrm{probe})}(\cdot|t;x;\beta_{0\to 1}) = \frac{q^{(\mathrm{probe})}(\cdot|t;x;\beta_{0\to 1})}{Z^{(\mathrm{probe})}(t;x;\beta_{0\to 1})}. \tag{A18}$$

Note that, according to Equations (8) and (9),

$$\lim_{t\to 1} p^{(\mathrm{probe})}(y|t;x;\beta_{0\to 1}) = \delta(x-y), \quad \lim_{t\to 0} p^{(\mathrm{probe})}(y|t;x;\beta_{0\to 1}) = p^{(\mathrm{tar})}(y).$$

In other words this density is certainly not a good approximation for the target density at larger t, and especially at $t\to 1$. This suggests that we should also consider **integrated probe** density

$$p^{(\text{i-probe})}(\cdot|t;\beta_{0\to 1}) = \int p^{(\mathrm{mar})}(x|t;u^*_{0\to t}(\cdot;\beta_{0\to 1}))\frac{q^{(\mathrm{probe})}(\cdot|t;x;\beta_{0\to 1})}{Z^{(\mathrm{probe})}(t;x;\beta_{0\to 1})}dx, \tag{A19}$$

for which the asymptotics become

$$\lim_{t\to 1} p^{(\text{i-probe})}(y|t;\beta_{0\to 1}) = \lim_{t\to 0} p^{(\text{i-probe})}(y|t;\beta_{0\to 1}) = p^{(\mathrm{tar})}(y).$$

- Let us define the **expectation** (average value) of the final state (conditioned to the observation of x at time t):

$$\hat{y}(t;x;\beta_{0\to 1}) \doteq \mathbb{E}_{y\sim p^{(\mathrm{probe})}(\cdot|t;x;\beta_{0\to 1})}[y] = \int_{\mathbb{R}^d} y p^{(\mathrm{probe})}(y|t;x;\beta_{0\to 1})dy. \tag{A20}$$

The expected final state was called an "order-parameter" in [6] because, in analysis of auto-correlations, it shows a sharper (than for the original x) transition from the de-correlated case to what seems to be relatively close structurally to the finite generated sample.

Next, it is natural to introduce what we call **predicted** density of the **expected** state:

$$p^{(\text{pred-e})}(y|t;\beta_{0\to 1}) = \int \delta(y-\hat{y}(t;x;\beta_{0\to 1}))p^{(\text{mar})}(x|t;u^{*}_{0\to t}(\cdot;\beta_{0\to 1}))dx \tag{A21}$$

We will see below that these densities will be useful for understanding the quality of the final state as we **grow** it in time with H-PID. Let us now discuss how we can sample from $p^{(\text{pred-e})}(y|t;\beta_{0\to 1})$ (or from $p^{(\text{pred-r-e})}(y|t;\beta_{0\to 1})$). The sampling is done in two steps: first, sample a path and record the state x where it arrives at the moment of time t; and second, evaluate the function $\hat{y}(t;x;\beta_{0\to 1})$ (or $\hat{y}(t;x;\beta_{0\to 1})$) at these t and x.

- Gaussian density built from $\Delta(t;x;y;\beta_{0\to 1})$—which is a quadratic form in x and y with coefficients dependent on time, defined in Equation (9)—becomes

$$\begin{aligned}\rho(y\,|\,t;x;\beta_{0\to 1}) &= \exp\big(-\Delta(t;x;y;\beta_{0\to 1})\big) = \mathcal{N}\Big(y\,\Big|\nu_t(x) \doteq \frac{b_t^{(-)}}{\mathcal{K}_t}x, \frac{1}{\mathcal{K}_t}\mathrm{I}_{d\times d}\Big) \\ &= \exp\left(-\frac{\mathcal{K}_t}{2}\left\|y-\frac{b_t^{(-)}}{\mathcal{K}_t}x\right\|^2 + \frac{d}{2}\log\left(\frac{\mathcal{K}_t}{2\pi}\right)\right), \quad \mathcal{K}_t \doteq c_t^{(-)} - a_1^{(+)}.\end{aligned} \tag{A22}$$

Here, $a^{(+)}$, $a^{(-)}$, $b^{(-)}$, and $c^{(-)}$ are the time (and $\beta_{0\to 1}$)-dependent coefficients obtained by solving the coupled Riccati-type ODEs for the harmonic path integral (cf. Appendix A). As discussed in Appendix A.1, in the case of piecewise-constant $\beta_{0\to 1}$ these coefficients are computed exactly in each interval and matched continuously at the interval boundaries.

We call the auxiliary Gaussian density $\rho(\cdot\,|\,\cdot;\cdot;\cdot)$ the **re-weighting** density because of its role played in Equation (8)—re-weighting from the target to probe densities.

Appendix C. Gaussian Mixtures

We consider the special case when the target distribution is a Gaussian mixture with component–specific full covariances:

$$p^{(\text{tar})}(y) \;\to\; p^{(\text{tar-G-Mix})}(y) \;\doteq\; \sum_{n=1}^{N} \varrho_n\, \mathcal{N}\big(y;\mu_n,\Sigma_n\big), \qquad \sum_{n=1}^{N}\varrho_n = 1, \tag{A23}$$

where each $\Sigma_n \in \mathbb{R}^{d\times d}$ is symmetric positive definite.

The **re-weighting density** $\rho(y \mid t;x)$, defined in Equation (A22) (restated here for convenience),

$$\rho(y\,|\,t;x) = \mathcal{N}\Big(y\,\Big|\,\frac{b_t^{(-)}}{K_t}x,\; \frac{1}{K_t}I_d\Big),$$

is isotropic with scalar gain $b_t^{(-)}$ and precision $K_t = c_t^{(-)} - a_1^{(+)} > 0$ determined by the chosen $\beta_{0\to 1}$–protocol, with $a_t^{(\pm)}, b_t^{(-)}, c_t^{(-)}$ solving ODEs (A9) subject to the boundary conditions Equations (A10) and (A11).

https://doi.org/10.3390/e28050492

Probe Density as a Product of Experts

The **probe** (a product of experts: target × re-weighting) is defined by Equations (8) and (A18) and restated as

$$p^{(\text{probe})}(y \mid t;x) = \frac{q^{(\text{probe})}(y \mid t;x)}{Z^{(\text{probe})}(t;x)} = \frac{p^{(\text{tar})}(y)\,\rho(y \mid t;x)}{\int p^{(\text{tar})}(y')\,\rho(y' \mid t;x)\,dy'}.$$

Plugging in (A23) and ρ yields the (unnormalized) mixture–Gaussian product

$$\begin{aligned} q^{(\text{probe-G-Mix})}(y \mid t;x) &\propto \mathcal{N}\Big(y \;\Big|\; \nu(t;x) = \frac{b_t^{(-)}}{K_t}x,\ \frac{1}{K_t}I_d\Big) \sum_{n=1}^{N} \varrho_n\,\mathcal{N}(y;\mu_n,\Sigma_n), \\ &\propto \sum_{n=1}^{N} \varrho_n\,w_n(t;x)\,\mathcal{N}\big(y;\tilde{\mu}_n(t;x),\tilde{\Sigma}_n(t)\big), \end{aligned} \tag{A24}$$

where the newly introduced variance, mean, and mixing weights are

$$\tilde{\Sigma}_n(t) \doteq \Big(\Sigma_n^{-1} + K_t\,I_d\Big)^{-1}, \quad \tilde{\mu}_n(t;x) \doteq \tilde{\Sigma}_n(t)\Big(\Sigma_n^{-1}\mu_n + b_t^{(-)}\,x\Big), \tag{A25}$$

$$\begin{aligned} w_n(t;x) &\doteq \mathcal{N}\Big(\nu(t;x);\,\mu_n,\,\Sigma_n + \tfrac{1}{K_t}I_d\Big) \\ &= \frac{\exp\Big(-\frac{1}{2}\big(\nu(t;x)-\mu_n\big)^{\top}\big(\Sigma_n + \frac{1}{K_t}I_d\big)^{-1}\big(\nu(t;x)-\mu_n\big)\Big)}{\sqrt{(2\pi)^d\,\det\big(\Sigma_n + \frac{1}{K_t}I_d\big)}}. \end{aligned} \tag{A26}$$

(Equation (A24) is a an example of the Gaussian product identity in action.)
Then **predicted (expected) state map**, defined in (A20), also becomes a Gaussian mixture

$$\hat{y}(t;x) = \frac{\int y\,q^{(\text{probe-G-Mix})}(y \mid t;x)\,dy}{\int q^{(\text{probe-G-Mix})}(y \mid t;x)\,dy} = \frac{\sum_{n=1}^{N}\varrho_n\,w_n(t;x)\,\tilde{\mu}_n(t;x)}{\sum_{n=1}^{N}\varrho_n\,w_n(t;x)}. \tag{A27}$$

References

1. Sohl-Dickstein, J.; Weiss, E.A.; Maheswaranathan, N.; Ganguli, S. Deep Unsupervised Learning using Nonequilibrium Thermodynamics. *arXiv* **2015**, arXiv:1503.03585. [CrossRef]
2. Ho, J.; Jain, A.; Abbeel, P. Denoising Diffusion Probabilistic Models. *arXiv* **2020**, arXiv:2006.11239. [CrossRef]
3. Song, Y.; Sohl-Dickstein, J.; Kingma, D.P.; Kumar, A.; Ermon, S.; Poole, B. Score-Based Generative Modeling through Stochastic Differential Equations. *arXiv* **2021**, arXiv:2011.13456. [CrossRef]
4. De Bortoli, V.; Thornton, J.; Heng, J.; Doucet, A. Diffusion Schrödinger Bridge with Applications to Score-Based Generative Modeling. In *Proceedings of the Advances in Neural Information Processing Systems*; Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P.S., Vaughan, J.W., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2021; Volume 34, pp. 17695–17709.
5. Lipman, Y.; Chen, R.T.Q.; Ben-Hamu, H.; Nickel, M.; Le, M. Flow Matching for Generative Modeling. *arXiv* **2023**, arXiv:2210.02747. [CrossRef]
6. Behjoo, H.; Chertkov, M. Harmonic Path Integral Diffusion. *IEEE Access* **2025**, *13*, 42196–42213. [CrossRef]
7. Kappen, H.J. Path integrals and symmetry breaking for optimal control theory. *J. Stat. Mech. Theory Exp.* **2005**, *2005*, P11011. [CrossRef]
8. Todorov, E. Linearly-solvable Markov decision problems. In *Proceedings of the Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2007; Volume 19, pp. 1369–1376.
9. Tsimpos, P.; Ren, Z.; Zech, J.; Marzouk, Y. Optimal Scheduling of Dynamic Transport. *arXiv* **2025**, arXiv:2504.14425. [CrossRef]
10. Chen, Y.; Vanden-Eijnden, E.; Xu, J. Lipschitz-Guided Design of Interpolation Schedules in Generative Models. *arXiv* **2025**, arXiv:2509.01629. [CrossRef]
11. Kunkel, L. Distribution estimation via Flow Matching with Lipschitz guarantees. *arXiv* **2025**, arXiv:2509.02337. [CrossRef]
12. Chertkov, M.; Behjoo, H. Adaptive Path Integral Diffusion: AdaPID. *arXiv* **2025**, arXiv:2512.11858. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.