TIME SERIES CONSTRAINED SAMPLING VIA REFINING TRAJECTORY-LEVEL CONTROL

Anonymous authors
Paper under double-blind review

ABSTRACT

Time series generation is central to applications such as finance, energy, and biomedical sensing, where realistic samples must respect partial observations or structural constraints. Existing training-free approaches based on guidance or projection enforce such constraints only locally at each diffusion step, which we show is insufficient to recover the conditional posterior in path space. In this work, we reformulate constrained sampling for time-series diffusion models as a trajectory-level stochastic optimal control problem, and unify guidance- and projection-based methods as instances of local control within this framework. We propose **Trajectory-level Time-series Control (TTC)**, a training-free, inference-time adjoint scheme that parameterizes the control in Fourier and wavelet subspaces aligned with global and localized temporal structure. A single backward pass already moves samples substantially closer to the constraint set at guidance-level cost, additional iterations tighten this further, and a terminal projection supplies exact feasibility. Across four datasets and multiple constraint types, TTC reaches exact constraint satisfaction with the best distributional quality among the compared methods, while running up to an order of magnitude faster than per-step projection.

1 INTRODUCTION

Time series generation plays a central role in applications such as financial scenario synthesis, energy demand simulation, and biomedical signal augmentation, where realistic samples must respect specifications the domain imposes, such as partial observations or structural constraints (Coletta et al., 2024; Narasimhan et al., 2025; Xiao et al., 2021). Such specifications are requirements rather than preferences: a grid planner stress-testing against a prescribed peak load learns nothing from a trace that misses it (Rizzato et al., 2022). Unlike images, where a constraint set is hard to define or verify, time series admit statistics that can be computed exactly, so satisfaction is both checkable and worth enforcing strictly (Narasimhan et al., 2025).

Among generative models, diffusion models have become the standard for time-series synthesis, performing strongly on unconditional generation (Yuan & Qiao, 2024) and on conditional tasks including forecasting (Rasul et al., 2021), imputation (Tashiro et al., 2021; Alcaraz & Strodthoff, 2023), and constrained generation (Coletta et al., 2024). Adapting them to conditional tasks follows two paradigms. Conditional training fits a diffusion model to each conditional task (Ho & Salimans, 2021; Singhal et al., 2025), which is impractical for time series, where conditions are diverse and often time series themselves, such as separate sequences or sparse past observations, so each new condition type requires retraining (Park et al., 2026). Training-free conditional sampling instead steers a pretrained unconditional model at inference time (Dhariwal & Nichol, 2021; Yuan & Qiao, 2024; Coletta et al., 2024; Narasimhan et al., 2025), adapting to arbitrary conditions without retraining.

Existing training-free approaches fall into **guidance**- and **projection**-based methods. Guidance steers each diffusion step with the gradient of a differentiable constraint loss (Yuan & Qiao, 2024; Coletta et al., 2024), but its point-estimate approximation biases the output (Chidambaram et al., 2024) and its weight has no principled relation to constraint violation (Song et al., 2023). Projection enforces the constraint by convex optimization at the final (Coletta et al., 2024) or every diffusion step (Narasimhan et al., 2025), at substantial computational cost and at the risk of distorting the conditional posterior. Both steer each step with a one-step surrogate of the constraint rather than its

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.

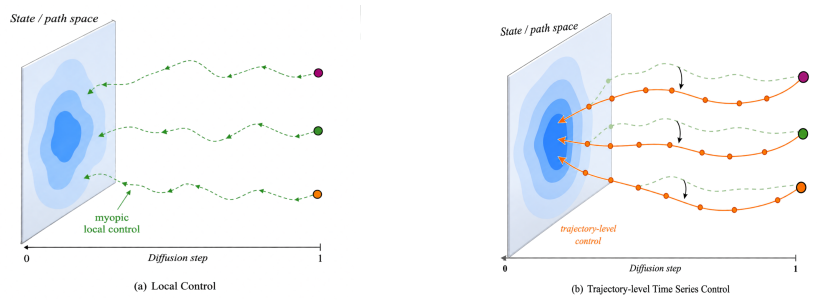

Figure 1: Overview of the proposed Trajectory-level Time-series Control (TTC). Blue clouds denote the data density on path space, and dark-blue ellipses denote constraint sets. (a) Existing methods, such as guidance and projection, correct each step with a one-step surrogate of the constraint, which can steer the trajectory away from the constraint set. (b) TTC differentiates the terminal constraint through the entire remaining trajectory and restricts the resulting control to subspaces matched to time-series structure.

exact path-wise expectation, which in general prevents them from recovering the optimal posterior path measure.

To mitigate this gap, we adopt a stochastic optimal control (SOC) (Bellman, 1966; Chen et al., 2016) perspective on training-free conditional sampling for time-series diffusion. SOC introduces an explicit control variable that perturbs the diffusion drift and formulates conditional sampling as a path-space optimization, providing the trajectory-level handle that local guidance and projection lack. Although SOC has been increasingly applied to diffusion-based generative modeling, prior work primarily targets training-based samplers[1] (Domingo-Enrich et al., 2025; Havens et al., 2025; Liu et al., 2025), and its application to time series remains limited (Park et al., 2024). Furthermore, we observe that the naive control space of SOC overlooks the fact that time-series signals exhibit prominent patterns, such as global trends and localized fluctuations.

In this paper, we reformulate conditional sampling as a trajectory-level SOC problem by interpreting the conditional posterior as a tilted distribution, and unify guidance- and projection-based methods as instances of local control within this framework. To refine the SOC control space for time series, we parameterize the control in Fourier and wavelet subspaces aligned with global and localized patterns of time-series signals, with a provable bias bound. Finally, we propose **Trajectory-level Time-series Control (TTC)**, an inference-time adjoint scheme that computes per-condition control without retraining (Figure 1). A single adjoint pass already outranks the guidance baselines at comparable cost, further iterations tighten constraint satisfaction with a stationarity guarantee, and a terminal projection supplies exact feasibility at little distortion. Across four datasets and multiple constraint types, TTC reaches exact constraint satisfaction with the best distributional quality among the compared methods, running up to an order of magnitude faster than per-step projection.

2 Related Work

2.1 Training-Free Time Series Conditional Sampling

Training-free conditional sampling for diffusion models has been addressed mainly through guidance and projection. Diffusion-TS (Yuan & Qiao, 2024) utilizes a guidance-based Langevin sampler for conditional generation based on Bayes' theorem, similar to diffusion posterior sampling (Chung et al., 2022). GuidedDiffTime (Coletta et al., 2024) steers sampling with gradients of differentiable constraint functions. While flexible, guidance-based methods rely on point-estimate approximations (Song et al., 2023) that deviate from the target posterior (Chidambaram et al., 2024), and violations vary non-monotonically with the guidance weight (Narasimhan et al., 2025).

Projection-based methods enforce constraints explicitly. The Constrained Optimization Problem (COP) (Coletta et al., 2024) projects a generated sample onto the constraint set via convex optimization.

[1]Here, a sampler draws samples from an energy function without access to data, as in molecule generation.

More recently, Constrained Posterior Sampling (CPS) (Narasimhan et al., 2025) performs a penalized projection on the Tweedie posterior mean (Efron, 2011) at every denoising step, mitigating off-manifold issues. While these methods achieve strong constraint satisfaction, they incur substantial optimization cost and converge to points in the constraint set rather than the conditional posterior.

While these two strategies represent the dominant paradigms for time series constrained sampling, neither directly addresses the path-space optimality of the sampling trajectory, motivating a stochastic optimal control perspective discussed next.

2.2 Optimal Control in Diffusion Sampling

Stochastic optimal control (SOC) provides a principled framework for steering stochastic dynamical systems toward desired objectives while minimizing path-space deviation from a reference process (Chen et al., 2016). The connection between Schrödinger bridges and entropy-regularized SOC is well established (Chen et al., 2016; De Bortoli et al., 2021) and increasingly applied to diffusion models, where the reverse sampling process is naturally interpreted as controlled stochastic dynamics (Berner et al., 2024; Holdijk et al., 2023; Chen et al., 2023).

Most prior work uses SOC to train or fine-tune diffusion generators toward reward or energy objectives (Domingo i Enrich et al., 2024; Uehara et al., 2024; Zhu et al., 2025; Ou et al., 2025). Others (Domingo-Enrich et al., 2025; Havens et al., 2025; Liu et al., 2025) amortize the optimal control by training a network on adjoint-derived signals, whereas we compute the adjoint at inference time for each condition without retraining. Noise-optimization methods (Wallace et al., 2023; Ben-Hamu et al., 2024; Tang et al., 2024) also backpropagate a terminal objective, but optimize a single realization rather than a control over the path measure.

Closer to our setting, several recent works apply SOC at inference time to pretrained diffusion models. Pandey et al. (2025) formulate conditioning as a variational control problem with gradient corrections, while Li et al. (2025) address hard-constrained sampling for flow-matching models. RB-Modulation (Rout et al., 2025) performs training-free stylization by framing the control as a proximal gradient procedure on the diffusion trajectory. Others restrict the control space to preserve manifold structure (He et al., 2024) or enforce geometric constraints (Yang et al., 2024), sharing our motivation but in image domains. In time series, Park et al. (2024) apply SOC to irregular series through training-based bridge construction. To our knowledge, inference-time SOC for constrained time series sampling has not been studied.

3 Conditional Sampling as Stochastic Optimal Control

In this section, we reformulate guidance- and projection-based conditional sampling within the SOC framework and show their theoretical limitation. Full derivations are in Appendix A.1.

3.1 SOC Reformulation of Conditional Sampling

Setup of Diffusion Sampling. The reverse process of a pretrained diffusion model (Song et al., 2021) is the SDE

$$\mathrm{d}X_t = \underbrace{\big(f(X_t,t) - \sigma(t)^2 s_\theta(X_t,t)\big)}_{:=\tilde{f}_t(X_t)}\,\mathrm{d}t + \sigma(t)\mathrm{d}W_t, \tag{1}$$

where s_θ is the learned score, f and σ are the drift and diffusion coefficients, W_t is a standard Wiener process, and x_t denotes a realization of X_t. Unconditional sampling induces a base path measure $\mathbb{P}_{\text{base}}$ with terminal data distribution p_{base}.

Conditional Target as Tilted Posterior. Following Song et al. (2023); Bhattacharya & Martin (2022), we encode the condition $\mathcal{C}$ through a terminal loss $L(x_0;\mathcal{C})$ and define the conditional **target** as the exponentially tilted distribution

$$p^\star(x \mid \mathcal{C}) = p_{\text{base}}(x)\exp\big(-\lambda L(x;\mathcal{C})\big)/Z, \tag{2}$$

with tilting strength $\lambda \geq 0$ and normalizing constant Z. This is a generalized Bayesian posterior with p_{base} as the *prior* and $\exp(-\lambda L)$ as the unnormalized *likelihood* (Bhattacharya & Martin, 2022).

Conditional Sampling as Path-Space Optimization. We now reformulate conditional sampling as an optimization over the entire reverse trajectory. Following SOC, we introduce a control vector field u that perturbs the drift in Equation 1:

$$\mathrm{d}X_t = \big[\tilde{f}_t(X_t) + \sigma(t)u_t(X_t)\big]\mathrm{d}t + \sigma(t)\mathrm{d}W_t. \tag{3}$$

The goal is a control u whose path measure $\mathbb{Q}_u$ has terminal marginal $p^\star$ while deviating minimally from $\mathbb{P}_{\text{base}}$, which is a **Schrödinger bridge** problem (Schrödinger, 1932) with the entropic objective

$$\min_{\mathbb{Q}_u} \; D_{\mathrm{KL}}(\mathbb{Q}_u \| \mathbb{P}_{\text{base}}) + \mathbb{E}_{\mathbb{Q}_u}[\lambda L(X_0; \mathcal{C})]. \tag{4}$$

Relation to Conditional Posterior. By Girsanov's theorem, $D_{\mathrm{KL}}(\mathbb{Q}_u \| \mathbb{P}_{\mathrm{base}})$ equals the expected control energy, reducing Equation 4 to the entropic optimal control problem

$$\min_{u} \; \mathbb{E}_{\mathbb{Q}_u}\Big[\int_0^1 \tfrac{1}{2}\|u_t(X_t)\|^2 dt + \lambda L(X_0; \mathcal{C})\Big]. \tag{5}$$

When the base process is memoryless, the first variation of this objective yields the optimal terminal marginal $p^\star(x_0)$ which exactly recovers the tilted posterior $p_{\text{base}}(x_0)\exp(-\lambda L(x_0; \mathcal{C}))$ in Equation 2.

3.2 Understanding Guidance and Projection in SOC

We now examine existing methods through the SOC lens, showing that the one-step surrogates they steer with cannot recover the optimal posterior path measure. Detailed proofs are in Appendix B.1.

Assumption 3.1. Define the unnormalized Doob's h-function at $t \in [0,1]$ as $h_t(x_t; \mathcal{C}) := \mathbb{E}_{\mathbb{P}_{\mathrm{base}}}[\exp(-\lambda L(X_0; \mathcal{C})) \;\big|\; X_t = x_t]$. Assume $h_t(x_t; \mathcal{C}) \in (0, \infty)$ and $\nabla_{x_t} \log h_t$ exists.

Guidance and Projection as Posterior Tilting. We first show that the entropic SOC formulation reveals a unified view of guidance- and projection-based methods. The control weight λ governs how strongly the posterior is tilted toward the constraint.

Proposition 3.2 (Posterior tilting as a control-energy trade-off)**.** *Under Assumption 3.1, the weight $\lambda \geq 0$ in Equation 5 governs the degree of posterior tilting, interpolating between the base distribution ($\lambda = 0$) and strict constraint enforcement ($\lambda \to \infty$).*

This unifies guidance- and projection-based methods within the entropic SOC framework: projection corresponds to the limit $\lambda \to \infty$, enforcing hard constraints at the terminal time, whereas guidance operates at finite λ, inducing a softer tilt through local surrogate losses (we return to the $\lambda \to \infty$ endpoint in Section 4.2).

Non-optimality in Path Space. Both paradigms steer with a locally defined surrogate rather than the exact h-function, which is generally insufficient to recover the optimal posterior path measure (Song et al., 2023; Chidambaram et al., 2024):

Theorem 3.3 (Local control is path-space suboptimal)**.** *Let $\tilde{L}_t(x_t)$ be a surrogate depending only on the current state, inducing $u_t^{\mathrm{local}}(x_t) = -\sigma(t)^\top \lambda \nabla_{x_t} \tilde{L}_t(x_t)$. Under Assumption 3.1, the excess objective value of u^{local} over the optimum equals*

$$D_{\mathrm{KL}}\left(\mathbb{Q}_{u^{\mathrm{local}}} \| \mathbb{Q}_{u^\star}\right) = \tfrac{1}{2}\mathbb{E}_{\mathbb{Q}_{u^{\mathrm{local}}}} \int_0^1 \left\| u_t^{\mathrm{local}} - u_t^\star \right\|^2 dt, \tag{6}$$

which vanishes if and only if $\lambda \tilde{L}_t(x_t) = -\log h_t(x_t; \mathcal{C}) + c(t)$ for a.e. (x_t, t) and some scalar $c(t)$.

Theorem 3.3 shows that neither paradigm meets this condition: guidance replaces the conditional expectation defining h_t with a one-step estimate $L(\hat{x}_0(x_t); \mathcal{C})$, which differs from $-\lambda^{-1}\log h_t$ whenever $X_0 \mid x_t$ is not degenerate, and projection enforces feasibility of the current state rather than following h_t at all. The gap is thus a property of the surrogate, not of tuning. For projection, the resulting distortion is governed by the displacement it must apply, which motivates using trajectory-level control to reduce it (Section 5.3).

4 PARAMETERIZED TRAJECTORY CONTROL FOR TIME SERIES

We present **Trajectory-level Time-series Control (TTC)**, which exploits the SOC formulation: since the control is an explicit variable, unlike in guidance or projection, it can be restricted to subspaces matched to time-series structure. We restrict the control to Fourier and wavelet subspaces (Section 4.1) and optimize it via an inference-time adjoint scheme (Section 4.2).

4.1 CONTROL PARAMETERIZATION FOR TIME SERIES

Time series exhibit two complementary structures, global stationary patterns and localized transient patterns (Alaa et al., 2021; Wu et al., 2025; Liu et al., 2024), and statistical constraints typically inherit one of them: mean constraints have gradients concentrated at low frequencies, whereas value-at-timestamp constraints have gradients concentrated in time. Aligning the control space with these structures via $\mathcal{R}_{\mathrm{FFT}}$ and $\mathcal{R}_{\mathrm{WAV}}$ controls more effectively than the full signal space.

Fourier-based Control. The Fourier basis sparsely captures global stationary patterns such as smooth trends and periodic components (Alaa et al., 2021; Liu et al., 2024), suiting constraints whose gradient has compact frequency support. We construct $\mathcal{R}_{\mathrm{FFT}}$ by applying the discrete Fourier transform (DFT) $\mathcal{F}_T$ to each channel c along the temporal axis and retaining a subset of modes, where

$$\mathcal{F}_T(u_t)[\omega, c] \coloneqq \sum_{\tau=0}^{T-1} u_t[\tau, c]\, e^{-2\pi i \omega \tau / T}. \tag{7}$$

Given a frequency budget $k \in \{1, \ldots, \lfloor T/2 \rfloor\}$, a symmetric low-pass mask $\mathbf{M}_k \in \{0,1\}^T$ retains the DC component and the k lowest positive and negative frequencies, fixed independently of the input and the diffusion step. With the inverse DFT $\mathcal{F}_T^{-1}$ and the Hadamard product $\odot$,

$$\mathcal{R}_{\mathrm{FFT}}(u_t) = \mathcal{F}_T^{-1}\Big(\mathbf{M}_k \mathbf{1}_C^\top \odot \mathcal{F}_T(u_t)\Big). \tag{8}$$

Wavelet-based Control. The wavelet basis represents nonstationary series through compactly supported functions (Wu et al., 2025; Liu et al., 2024): unlike global Fourier modes, a control restricted to coarse-scale coefficients can vanish outside a short interval, suiting constraints whose gradient is concentrated in time (e.g., endpoint or peak constraints). We define $\mathcal{R}_{\mathrm{WAV}}$ with an orthonormal discrete wavelet transform $\mathcal{W}$ applied per channel, so that $\mathcal{W}^{-1} = \mathcal{W}^\top$, which yields approximation coefficients through the scaling function φ and details through the mother wavelet ψ:

$$a_j(u_t)[\ell, c] = \sum_{\tau=0}^{T-1} u_t[\tau, c]\varphi_\ell^j[\tau], \qquad d_j(u_t)[\ell, c] = \sum_{\tau=0}^{T-1} u_t[\tau, c]\psi_\ell^j[\tau], \tag{9}$$

with $\varphi_\ell^j[\tau] = 2^{-j/2}\varphi(2^{-j}\tau - \ell)$, $\psi_\ell^j[\tau] = 2^{-j/2}\psi(2^{-j}\tau - \ell)$, and position index ℓ. We write $\mathcal{W}(u_t)$ for $(a_J, d_J, \ldots, d_1)$ in a fixed ordering. Given a level budget J, a mask $\mathbf{S}_J \in \{0,1\}^T$ retains a_J and zeros all detail coefficients, fixed independently of the input and the diffusion step, giving

$$\mathcal{R}_{\mathrm{WAV}}(u_t) = \mathcal{W}^{-1}\Big(\mathbf{S}_J \mathbf{1}_C^\top \odot \mathcal{W}(u_t)\Big). \tag{10}$$

Under $\mathcal{R} \in \{\mathcal{R}_{\mathrm{FFT}}, \mathcal{R}_{\mathrm{WAV}}\}$ and its range $V \coloneqq \mathrm{Range}(\mathcal{R})$, the final entropic SOC becomes:

$$\min_{v \in V} \mathbb{E}_{\mathbb{Q}_v}\Big[\int_0^1 \tfrac{1}{2}\|v_t(X_t)\|^2 dt + \lambda L(X_0; \mathcal{C})\Big]. \tag{11}$$

Theoretical Bound. Restricting the control to V may distort the sampling distribution, since the tilted posterior is attained only at the unconstrained optimum $u^\star$. The following bounds this distortion.

Theorem 4.1 (Approximation bias under control parameterization)**.** *Let $\mathcal{R} : L^2 \to L^2$ be a linear, self-adjoint, idempotent operator with range V, and let the base process be memoryless and $u^\star$ the optimal control of Equation 5. Define $v^\star = \mathcal{R}(u^\star)$ and let $\mathbb{Q}_0^{v^\star}$ be the terminal marginal of the diffusion controlled by $v^\star$. Then*

$$D_{\mathrm{KL}}(\mathbb{Q}_0^{v^\star} \| p^\star) \le \frac{1}{2}\,\|u^\star - v^\star\|_{L^2}^2, \tag{12}$$

where $p^\star$ is the target tilted distribution in Equation 2 and the norm is taken over $\mathbb{Q}^{v^\star} \otimes \mathrm{d}t$. The bound vanishes when $u^\star \in V$.

A detailed proof is in Appendix B.2.1. The bound concerns the representational cost of V, not the optimization error of the computed control. Since $u_t^\star$ inherits the structure of the constraint through h_t, the residual is small for constraints matched to the chosen subspace (low-frequency gradients for $\mathcal{R}_{\mathrm{FFT}}$, temporally concentrated gradients for $\mathcal{R}_{\mathrm{WAV}}$), so restricting the control to V reduces its dimension at little cost in distributional accuracy, as we confirm in Appendix C.2.

4.2 Adjoint-Based Trajectory-level Time-series Control

Exact feasibility is the $\lambda \to \infty$ end of Proposition 3.2, which no drift can reach, since Girsanov's theorem keeps the terminal law absolutely continuous with respect to the base law uniformly in λ (Appendix B.1.3). In that limit, Equation 11 reduces to the minimum-energy problem of reaching $\{L(X_0;\mathcal{C}) = 0\}$, approximated by a terminal projection whose distortion grows with the displacement it applies. TTC therefore proceeds in three stages (Algorithm 1): a single adjoint pass approximates the finite-λ control to first order, the regime Theorem 4.1 bounds; refinement within V raises the effective tilt, trading fidelity to the finite-λ posterior for proximity to the constraint set; and a terminal projection approximates the $\lambda \to \infty$ instance at little cost, since its displacement is already small.

Single adjoint pass. The optimal control involves the intractable Doob h-transform $\nabla_{x_t} \log h_t(x_t;\mathcal{C})$, which we bypass via Pontryagin's Maximum Principle (PMP) (Pavliotis, 2014; Pontryagin, 2018): the optimal control satisfies $u_t^\star(x_t) = -\sigma(t)^\top a_t$, where the adjoint state a_t solves the backward equation

$$\dot{a}_t = -\big(\nabla_x \tilde{f}_t(x_t)\big)^\top a_t, \tag{13}$$

with $a_0 = \lambda \nabla_{x_0} L(x_0;\mathcal{C})$. Full derivations are in Appendix A.2. Evaluating Equation 13 once along an uncontrolled trajectory and restricting the result to V yields a single-pass controller, a first-order approximation of $u^\star$ at guidance-level cost, where restricting to V costs only the residual bounded in Theorem 4.1.

Iterative refinement. Since the single pass is a first-order approximation subject to numerical error and stochasticity (Zhuang et al., 2020; Nadarajah & Jameson, 2000), it leaves a residual violation. We therefore refine the control within V by descending the terminal objective

$$J_V(v) = \mathbb{E}_{\mathbb{Q}^v}\big[\lambda L(X_0;\mathcal{C})\big], \qquad v \in V. \tag{14}$$

At each iteration k, backpropagating through the controlled trajectory under $v^{(k)}$ gives the adjoint $a^{(k)}$, which supplies the gradient of J_V along the sampled path, and we update

$$v^{(k+1)} = v^{(k)} + \gamma\, \mathcal{R}\big(-\sigma^\top a^{(k)}\big). \tag{15}$$

Each sample maintains its own control, updated with a fresh noise draw at every iteration, so that the update descends the path-space objective rather than fitting a single realization. Since J_V retains only the terminal cost of Equation 11, the control energy is not re-minimized and K acts as an effective tilt strength: in the linear-Gaussian case the iterates move the mean along $p^\star_{\lambda'}$ with λ' increasing in K (Remark B.10 and Appendix C.3.1).

Algorithm 1 Trajectory-level Time-series Control

Require: Base drift $\tilde{f}$, loss $L(\cdot;\mathcal{C})$, parameterization $\mathcal{R}$, weight λ, step size γ, iterations K

1: Initialize control $v^{(0)} \leftarrow 0$
2: **for** $k = 0$ **to** $K-1$ **do**
3: Draw $x_1^{(k)} \sim \mathcal{N}(0, I)$
4: Sample a trajectory $(x_{t_i}^{(k)})_{i=0}^N \sim \mathbb{Q}^{v^{(k)}}$
5: Compute terminal adjoint: $a_0^{(k)} \leftarrow \lambda \nabla_{x_0} L(x_0^{(k)};\mathcal{C})$
6: Backward adjoint propagation along the controlled trajectory: $\dot{a}_t^{(k)} = -\big(\nabla_x \tilde{f}_t\big)^\top a_t^{(k)}$
7: Update control: $v^{(k+1)} \leftarrow v^{(k)} + \gamma \mathcal{R}\big(-\sigma(t)^\top a^{(k)}\big)$
8: **end for**
9: Draw $x_0 \sim \mathbb{Q}^{v^{(K)}}$
10: Terminal projection ($\lambda \to \infty$ step): $x_0 \leftarrow \mathrm{Proj}_{\mathcal{C}}(x_0)$
11: **Output:** sample x_0

Proposition 4.2 (Stationarity under parameterization $\mathcal{R}$)**.** *Let $\mathcal{R}$ be linear, self-adjoint and idempotent, i.e., the orthogonal projection onto $V = \mathrm{Range}(\mathcal{R})$, and let $J_V(v) = \mathbb{E}_{\mathbb{Q}^v}[\lambda L(X_0;\mathcal{C})]$ for $v \in V$. Then the update $v^{(k+1)} = v^{(k)} + \gamma \mathcal{R}(-\sigma^\top a^{(k)})$ is a projected stochastic gradient step on J_V, and*

(i) *any fixed point $\bar{v}$ of its expected update satisfies $\mathcal{R}(\nabla J_V(\bar{v})) = 0$;*

(ii) *if $\sigma^\top a^{(k)}$ is an unbiased estimate of $\nabla J_V(v^{(k)})$ with $\mathbb{E}\|\mathcal{R}(\sigma^\top a^{(k)} - \nabla J_V(v^{(k)}))\|^2 \le s_V^2$, and J_V is β-smooth on V and bounded below by $J_V^\star$, then for $\gamma \le 1/\beta$,*

$$\min_{0 \le k < K} \mathbb{E}\big\|\mathcal{R}\big(\nabla J_V(v^{(k)})\big)\big\|^2 \le \frac{2\big(J_V(v^{(0)}) - J_V^\star\big)}{\gamma K} + \beta\gamma\, s_V^2, \tag{16}$$

Table 1: Quality comparison of guidance-based methods with final projection and projection-based methods. The best results are in **bold** and the second best are underlined. DownArrow (↓) indicates lower is better.

		Violation ↓			Context-FID ↓			Correlation ↓			Avg Rank ↓
Dataset	Method	Mean	ΔMean	Max	Mean	ΔMean	Max	Mean	ΔMean	Max	(Rank)
Stock	GuidedDiffTime	0.834	**0.000**	**0.000**	608.311	379.406	538.001	0.282	0.278	0.402	5.44 (8)
	Diffusion-TS	1.923	**0.000**	**0.000**	13.717	656.096	276.481	0.133	0.298	0.431	6.00 (9)
	TTC-FFT$_s$	**0.000**	**0.000**	**0.000**	0.192	283.406	29.460	0.020	0.276	**0.326**	2.56 (2)
	TTC-WAV$_s$	**0.000**	**0.000**	**0.000**	0.216	363.999	28.367	0.039	0.278	0.428	3.44 (4)
	COP	0.020	**0.000**	0.018	**0.149**	416.229	483.345	0.017	0.293	0.508	5.22 (7)
	COP-FT	0.020	**0.000**	0.018	0.183	316.462	261.026	**0.015**	0.304	0.501	4.89 (6)
	CPS	**0.000**	**0.000**	**0.000**	0.379	428.990	24.428	0.230	0.266	0.648	4.44 (5)
	TTC-FFT	**0.000**	**0.000**	**0.000**	0.177	**209.189**	**16.235**	0.034	**0.248**	0.472	**2.00 (1)**
	TTC-WAV	**0.000**	**0.000**	**0.000**	0.185	238.360	16.639	0.029	0.279	0.590	3.22 (3)
ETTh	GuidedDiffTime	0.006	**0.000**	0.003	12.497	12.216	11.378	0.319	0.330	0.364	6.11 (6)
	Diffusion-TS	0.013	**0.000**	0.002	9.538	12.498	12.571	0.199	0.327	0.392	6.11 (6)
	TTC-FFT$_s$	**0.000**	**0.000**	**0.000**	9.184	11.791	6.196	**0.167**	0.326	0.398	2.89 (4)
	TTC-WAV$_s$	**0.000**	**0.000**	**0.000**	9.330	12.620	7.084	0.171	0.316	0.375	3.67 (5)
	COP	0.029	0.002	0.020	9.267	12.454	8.307	0.199	0.341	0.363	6.44 (8)
	COP-FT	0.029	0.002	0.020	9.597	14.878	8.346	0.212	0.333	0.362	7.33 (9)
	CPS	**0.000**	**0.000**	**0.000**	**5.271**	13.143	7.094	0.189	0.313	**0.242**	2.78 (3)
	TTC-FFT	**0.000**	**0.000**	**0.000**	7.835	11.761	**5.092**	**0.167**	**0.310**	0.429	**2.22 (1)**
	TTC-WAV	**0.000**	**0.000**	**0.000**	7.501	**11.593**	6.832	0.168	0.313	0.391	**2.22 (1)**
Energy	GuidedDiffTime	0.459	0.001	0.087	19.164	21.674	18.676	8.076	8.226	8.221	8.22 (9)
	Diffusion-TS	1.636	0.001	0.084	5.312	19.273	11.427	2.911	7.813	8.472	6.89 (7)
	TTC-FFT$_s$	**0.000**	**0.000**	**0.000**	2.194	**17.079**	9.750	**1.025**	8.006	7.084	2.56 (3)
	TTC-WAV$_s$	**0.000**	**0.000**	**0.000**	2.312	18.180	10.418	1.099	7.870	7.484	3.11 (5)
	COP	0.106	0.015	0.109	2.927	21.515	12.942	1.438	7.967	8.146	7.00 (8)
	COP-FT	0.107	0.017	0.109	3.050	17.943	16.092	1.409	7.744	8.151	6.33 (6)
	CPS	**0.000**	**0.000**	**0.000**	2.353	20.290	**4.863**	1.436	7.743	**5.093**	2.78 (4)
	TTC-FFT	**0.000**	**0.000**	**0.000**	1.783	17.130	6.681	1.307	7.910	5.894	2.33 (2)
	TTC-WAV	**0.000**	**0.000**	**0.000**	**1.756**	18.771	9.975	1.223	**7.641**	6.318	**2.22 (1)**
fMRI	GuidedDiffTime	0.365	**0.000**	**0.000**	20.429	20.217	15.567	3.395	3.417	3.898	6.00 (7)
	Diffusion-TS	**0.000**	**0.000**	**0.000**	20.573	19.864	13.862	3.350	3.359	4.056	4.33 (6)
	TTC-FFT$_s$	**0.000**	**0.000**	**0.000**	19.172	19.440	3.641	3.356	3.325	4.970	3.56 (5)
	TTC-WAV$_s$	**0.000**	**0.000**	**0.000**	18.918	18.788	4.716	3.376	3.317	4.819	3.33 (4)
	COP	0.144	0.070	0.021	19.425	18.930	14.606	3.379	3.395	3.935	6.44 (8)
	COP-FT	0.144	0.071	0.021	19.425	19.116	14.528	3.362	3.435	3.977	6.67 (9)
	CPS	**0.000**	**0.000**	**0.000**	21.547	**16.650**	11.582	3.927	**3.173**	**3.834**	3.22 (3)
	TTC-FFT	**0.000**	**0.000**	**0.000**	**18.714**	18.692	**3.336**	**3.310**	3.340	4.749	**2.33 (1)**
	TTC-WAV	**0.000**	**0.000**	**0.000**	19.252	18.446	3.764	3.341	3.312	4.621	2.44 (2)

so $\gamma \propto 1/\sqrt{K}$ reaches an ε-stationary point of J_V within V in $O(1/\varepsilon^2)$ iterations.

The proof is in Appendix B.2.2. The scheme is also anytime: each iteration yields a valid sample under $v^{(k)}$, so K can be chosen to fit the available budget.

Terminal projection. The final stage applies a single penalty projection at $t = 0$, a deterministic approximation of the $\lambda \to \infty$ instance, which changes samples little since the preceding stages have already moved them near the constraint set. Algorithm 1 summarizes the full procedure.

5 Experiments

5.1 Experimental Setups

Tasks and Metrics. Following Coletta et al. (2024); Narasimhan et al. (2025), we consider three constraints: the temporal mean (**Mean**), the mean consecutive change (**ΔMean**), and the value at the argmax (**Max**). Violation is the average magnitude of constraint infeasibility beyond the tolerance $\tau = 0.01$ of Narasimhan et al. (2025), so zero violation denotes satisfaction within τ. For distributional quality, we use Context-FID (Jeha et al., 2022) on TS2Vec (Yue et al., 2022) embeddings and the correlational score (Liao et al., 2020), the cross-correlation error between real and synthetic series; results are averaged over five runs.

Table 2: Computational time (seconds per batch) across constraints.

Method	Stock			ETTh			Energy			fMRI		
	Mean	ΔMean	Max	Mean	ΔMean	Max	Mean	ΔMean	Max	Mean	ΔMean	Max
GuidedDiffTime	24	23	23	24	24	25	38	36	35	45	44	44
Diffusion-TS	30	31	32	33	35	34	50	51	51	63	63	61
TTC-FFT$_s$	26	28	29	29	32	31	49	52	59	71	73	77
TTC-WAV$_s$	31	32	32	33	34	34	55	58	61	75	76	80
COP	60	54	266	65	69	410	199	274	2400	432	520	3619
COP-FT	56	47	174	63	55	179	160	159	812	266	267	1098
CPS	429	452	643	504	533	721	1554	1812	2527	2909	3448	4562
TTC-FFT	139	132	134	146	148	150	210	206	237	291	295	291
TTC-WAV	167	162	167	166	172	177	280	272	275	345	345	363

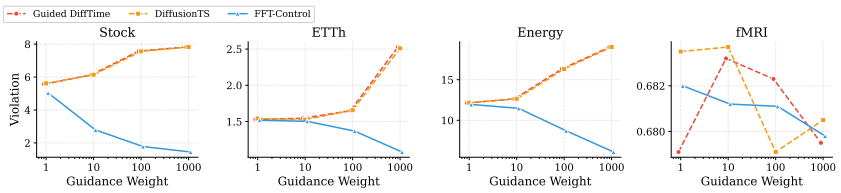

Figure 2: Constraint violation under varying control weight λ (or guidance weight for baselines).

Datasets and Baselines. We use four datasets: **Stock** (Google stock prices and volumes, 6 features), **ETTh** (hourly electricity transformer temperature, 7 features), **Energy** (UCI appliance energy usage in a low-energy building, 28 features), and **fMRI** (blood-oxygen-level-dependent signals, 50 features). We compare against guidance-based methods (Diffusion-TS (Yuan & Qiao, 2024), GuidedDiffTime (Coletta et al., 2024)) and projection-based methods (COP, COP-FT (Coletta et al., 2024), CPS (Narasimhan et al., 2025)), following their original implementations. All methods share the same unconditional Diffusion-TS backbone with sequence length 48, sampled by DDIM with 200 steps.

5.2 Main Results

Table 1 reports constrained generation with the final projection applied to all methods, so that they are compared at a common feasibility target; runtimes are in Table 2. TTC-FFT and TTC-WAV denote our method under the Fourier and wavelet parameterizations, and TTC-FFT$_s$ and TTC-WAV$_s$ their single-pass variants ($K = 1$), whose cost is comparable to that of the guidance baselines.

With the terminal penalty projection at their selected weights, guidance-based methods retain residual Mean violations (e.g., 0.834 and 1.923 on Stock), as do COP and COP-FT despite their own terminal projection (0.144 on fMRI, 0.106 on Energy). Only CPS and TTC reach exact satisfaction on all four datasets, through opposite mechanisms: CPS projects at every sampling step, whereas TTC steers the trajectory so that one terminal projection suffices.

The comparison therefore reduces to what feasibility costs in quality and runtime. On quality, TTC variants hold the top two ranks on all four datasets, ahead of CPS, with the largest margin on Stock (2.00 and 2.56 against 4.44). On runtime, per-step convex optimization makes CPS over $30\times$ slower than the guidance baselines on Energy and fMRI, while TTC is roughly $10\times$ faster than CPS on fMRI. Even the single-pass variants, at guidance-level time cost (26s against 24s on Stock), reach zero violation and rank second and fourth on Stock, ahead of every projection-based method, and TTC-FFT$_s$ also ranks ahead of CPS on Energy; additional iterations further improve quality.

5.3 Ablation Study

Effect of Posterior Tilting via Control Weight. To verify Proposition 3.2, we vary the control weight λ and find that increasing λ consistently reduces constraint violation for our methods (Figure 2). In contrast, increasing the guidance weight does not reduce violation for the gradient-based baselines and often increases it, consistent with Theorem 3.3: a larger weight amplifies the one-step surrogate rather than correcting its bias. This supports the SOC view of λ as trading control energy for constraint satisfaction.

Table 3: Violation comparison without the final single projection. The best results are in **bold** and the second best are underlined. Lower is better.

	Stock			ETTh			Energy			fMRI		
Method	Mean	ΔMean	Max	Mean	ΔMean	Max	Mean	ΔMean	Max	Mean	ΔMean	Max
GuidedDiffTime	5.552	**0.000**	4.976	1.524	0.009	**1.646**	12.017	0.110	14.818	0.681	0.252	32.498
Diffusion-TS	7.569	**0.000**	6.588	1.645	0.009	5.300	16.384	0.109	21.585	0.682	0.253	36.636
TTC-FFT (No proj)	0.653	**0.000**	**3.949**	**1.088**	0.008	1.948	**4.743**	**0.094**	**14.046**	**0.649**	**0.247**	14.966
TTC-WAV (No proj)	**0.650**	**0.000**	5.387	1.096	**0.005**	2.095	4.749	0.095	14.828	0.650	0.249	**14.889**

Effect of Trajectory Control. To isolate trajectory-level control, we evaluate our method without the final projection step in Table 3. TTC-FFT and TTC-WAV reduce mean violation relative to the stronger guidance-based baseline by 88% on Stock, 61% on Energy and 29% on ETTh, with fMRI already close to satisfied under the unconstrained prior. In the linear-Gaussian setting of Appendix C.3.1, a single adjoint pass halves the KL to the closed-form tilted posterior relative to the uncontrolled prior (0.081 against 0.170), whereas guidance ends further from the target than no control at all. For Mean-type constraints, the control thus does most of the work of approaching the constraint set, leaving the final projection only a small correction. Appendix C.2.1 reports the full comparison without projection across all metrics.

Effect of Control Parameterization. We compare against direct control without any parameterization (TTC-Direct). As summarized in Table 4, TTC-Direct reaches an average rank of 3.20, close to the 2.92 of CPS, the strongest baseline. A basis parameterization reverses the gap: TTC variants improve on TTC-Direct on every dataset and rank ahead of CPS in all eight dataset-method comparisons, reaching 2.03 and 2.22 on average, so the parameterization adds an inductive bias complementary to the SOC formulation (details in Appendix C.2.2).

Table 4: Average ranks over the methods of Table 9. Best results in **bold**.

Dataset	CPS	TTC-Direct	TTC-FFT	TTC-WAV
Stock	4.11	3.00	**1.78**	2.89
ETTh	2.22	3.44	2.11	**1.89**
Energy	2.44	2.78	2.00	**1.89**
fMRI	2.89	3.56	**2.22**	**2.22**
Avg	2.92	3.20	**2.03**	2.22

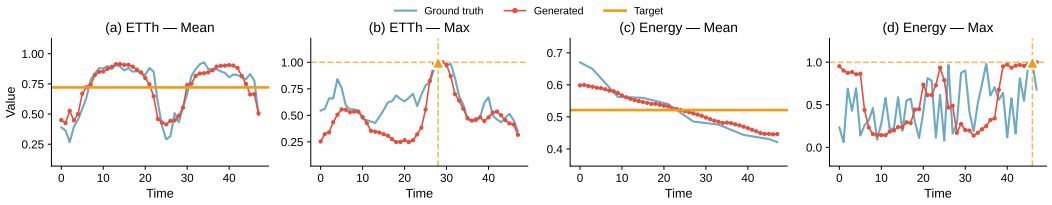

Figure 3: Constrained generation on ETTh and Energy under Mean and Max constraints.

Visualization. Figure 3 shows samples on ETTh and Energy under the Mean and Max constraints: the generated trajectories (red) follow the temporal patterns of the ground truth (blue) while meeting the constraint targets (orange), with more visualizations in Appendix C.4.

6 Conclusion

In this paper, we present a training-free constrained generation framework for time-series diffusion models, formulating constrained generation as a trajectory-level stochastic optimal control problem. To exploit the structure of time series, we introduced Fourier- and wavelet-based control parameterizations and an adjoint-based iterative refinement, completed by a terminal projection that the control makes inexpensive. Several limitations remain. First, our iterative refinement incurs additional computational cost compared to single-pass guidance methods, although it remains substantially faster than per-step projection approaches on high-dimensional datasets. Second, the parameterization is chosen per constraint type rather than selected automatically; extending it to data-driven basis selection is a promising future direction.

AI USE STATEMENT

Generative AI tools were used in this work only to edit the manuscript: checking grammar, correcting typos, and verifying that cross-references and citations are not broken. They were not used to come up with research ideas, derive the theoretical results, or run or analyze any experiment, and we did not use retrieval-augmented literature search. All AI-assisted output was reviewed by the authors, and we take full responsibility for the final content of this paper, including any text, claims or artifacts produced with the help of generative AI.

ETHICS STATEMENT

This work proposes a training-free method for constrained sampling from a pretrained time series diffusion model and does not involve human subjects or personally identifiable data. All experiments use publicly available benchmark datasets (Stock, ETTh, Energy and fMRI) that are standard in time series generation. We are not aware of ethical concerns specific to this work beyond those that apply to generative models in general.

REPRODUCIBILITY STATEMENT

Section 4 describes the method in full, and Algorithm 1 gives it as pseudocode. Section 5.1 and Appendix C.1 specify the experimental setup. Complete derivations and proofs of the theoretical results are given in Appendices A.1, A.2, B.1, B.2.1 and B.2.2. We will release the code upon publication.

REFERENCES

Ahmed Alaa, Alex James Chan, and Mihaela van der Schaar. Generative time-series modeling with fourier flows. In *International Conference on Learning Representations*, 2021.

Juan Lopez Alcaraz and Nils Strodthoff. Diffusion-based time series imputation and forecasting with structured state space models. *Transactions on Machine Learning Research*, 2023.

Richard Bellman. Dynamic programming. *science*, 153(3731):34–37, 1966.

Heli Ben-Hamu, Omri Puny, Itai Gat, Brian Karrer, Uriel Singer, and Yaron Lipman. D-flow: Differentiating through flows for controlled generation. *arXiv preprint arXiv:2402.14017*, 2024.

Julius Berner, Lorenz Richter, and Karen Ullrich. An optimal control perspective on diffusion-based generative modeling. *Transactions on Machine Learning Research*, 2024.

Indrabati Bhattacharya and Ryan Martin. Gibbs posterior inference on multivariate quantiles. *Journal of Statistical Planning and Inference*, 218:106–121, 2022.

Tianrong Chen, Jiatao Gu, Laurent Dinh, Evangelos Theodorou, Joshua M Susskind, and Shuangfei Zhai. Generative modeling with phase stochastic bridge. In *International Conference on Learning Representations*, 2023.

Yongxin Chen, Tryphon T Georgiou, and Michele Pavon. On the relation between optimal transport and schrödinger bridges: A stochastic control viewpoint. *Journal of Optimization Theory and Applications*, 169(2):671–691, 2016.

Muthu Chidambaram, Khashayar Gatmiry, Sitan Chen, Holden Lee, and Jianfeng Lu. What does guidance do? a fine-grained analysis in a simple setting. *Advances in Neural Information Processing Systems*, 37:84968–85005, 2024.

Hyungjin Chung, Jeongsol Kim, Michael T Mccann, Marc L Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. *arXiv preprint arXiv:2209.14687*, 2022.

Andrea Coletta, Sriram Gopalakrishnan, Daniel Borrajo, and Svitlana Vyetrenko. On the constrained time-series generation problem. *Advances in Neural Information Processing Systems*, 36, 2024.

Talgat Daulbaev, Alexandr Katrutsa, Larisa Markeeva, Julia Gusak, Andrzej Cichocki, and Ivan Oseledets. Interpolation technique to speed up gradients propagation in neural odes. *Advances in Neural Information Processing Systems*, 33:16689–16700, 2020.

Valentin De Bortoli, James Thornton, Jeremy Heng, and Arnaud Doucet. Diffusion schrödinger bridge with applications to score-based generative modeling. *Advances in Neural Information Processing Systems*, 34:17695–17709, 2021.

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021.

Steven Diamond and Stephen Boyd. Cvxpy: A python-embedded modeling language for convex optimization. *Journal of Machine Learning Research*, 17(83):1–5, 2016.

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

Carles Domingo i Enrich, Jiequn Han, Brandon Amos, Joan Bruna, and Ricky TQ Chen. Stochastic optimal control matching. *Advances in Neural Information Processing Systems*, 37:112459–112504, 2024.

Bradley Efron. Tweedie's formula and selection bias. *Journal of the American Statistical Association*, 106(496):1602–1614, 2011.

Aaron Havens, Benjamin Kurt Miller, Bing Yan, Carles Domingo-Enrich, Anuroop Sriram, Brandon Wood, Daniel Levine, Bin Hu, Brandon Amos, Brian Karrer, et al. Adjoint sampling: Highly scalable diffusion samplers via adjoint matching. *arXiv preprint arXiv:2504.11713*, 2025.

Yutong He, Naoki Murata, Chieh-Hsin Lai, Yuhta Takida, Toshimitsu Uesaka, Dongjun Kim, Wei-Hsiang Liao, Yuki Mitsufuji, J Zico Kolter, Ruslan Salakhutdinov, et al. Manifold preserving guided diffusion. In *International Conference on Learning Representations*, 2024.

Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017.

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. In *NeurIPS 2021 Workshop on Deep Generative Models and Downstream Applications*, 2021.

Lars Holdijk, Yuanqi Du, Ferry Hooft, Priyank Jaini, Berend Ensing, and Max Welling. Stochastic optimal control for collective variable free sampling of molecular transition paths. *Advances in Neural Information Processing Systems*, 36:79540–79556, 2023.

Jan Christian Hückelheim and Jens-Dominik Müller. Checkpointing with time gaps for unsteady adjoint cfd. In *Advances in Evolutionary and Deterministic Methods for Design, Optimization and Control in Engineering and Sciences*, pp. 117–130. Springer, 2018.

Paul Jeha, Michael Bohlke-Schneider, Pedro Mercado, Shubham Kapoor, Rajbir Singh Nirwan, Valentin Flunkert, Jan Gasthaus, and Tim Januschowski. Psa-gan: Progressive self attention gans for synthetic time series. In *International Conference on Learning Representations*, 2022.

Ioannis Karatzas and Steven Shreve. *Brownian motion and stochastic calculus*. springer, 2014.

Zeyang Li, Kaveh Alim, and Navid Azizan. Hardflow: Hard-constrained sampling for flow-matching models via trajectory optimization. *arXiv preprint arXiv:2511.08425*, 2025.

Shujian Liao, Hao Ni, Lukasz Szpruch, Magnus Wiese, Marc Sabate-Vidales, and Baoren Xiao. Conditional sig-wasserstein gans for time series generation. *arXiv preprint arXiv:2006.05421*, 2020.

Guan-Horng Liu, Jaemoo Choi, Yongxin Chen, Benjamin Kurt Miller, and Ricky T. Q. Chen. Adjoint schrödinger bridge sampler. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

Peiyuan Liu, Beiliang Wu, Naiqi Li, Tao Dai, Fengmao Lei, Jigang Bao, Yong Jiang, and Shu-Tao Xia. Wftnet: Exploiting global and local periodicity in long-term time series forecasting. In *Icassp 2024-2024 ieee international conference on acoustics, speech and signal processing (icassp)*, pp. 5960–5964. IEEE, 2024.

Siva Nadarajah and Antony Jameson. A comparison of the continuous and discrete adjoint approach to automatic aerodynamic optimization. In *38th Aerospace sciences meeting and exhibit*, pp. 667, 2000.

Sai Shankar Narasimhan, Shubhankar Agarwal, Litu Rout, Sanjay Shakkottai, and Sandeep P. Chinchali. Constrained posterior sampling: Time series generation with hard constraints. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

Zijing Ou, Ruixiang Zhang, and Yingzhen Li. Discrete neural flow samplers with locally equivariant transformer. *arXiv preprint arXiv:2505.17741*, 2025.

Kushagra Pandey, Farrin Marouf Sofian, Felix Draxler, Theofanis Karaletsos, and Stephan Mandt. Variational control for guidance in diffusion models. In *International Conference on Machine Learning*, 2025.

Byoungwoo Park, Hyungi Lee, and Juho Lee. Efficient modeling of irregular time-series with stochastic optimal control. In *NeurIPS 2024 Workshop on Bayesian Decision-making and Uncertainty*, 2024.

Jinseong Park, Seungyun Lee, Woojin Jeong, Yujin Choi, and Jaewook Lee. Timebridge: Better diffusion prior design with bridge models for time series generation. In *Proceedings of the 32th ACM SIGKDD conference on knowledge discovery & data mining*, 2026.

Grigorios A Pavliotis. Stochastic processes and applications. *Texts in applied mathematics*, 60, 2014.

Lev Semenovich Pontryagin. *Mathematical theory of optimal processes*. Routledge, 2018.

Kashif Rasul, Calvin Seward, Ingmar Schuster, and Roland Vollgraf. Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. In *International Conference on Machine Learning*, pp. 8857–8868. PMLR, 2021.

Matteo Rizzato, Nicolas Morizet, William Maréchal, and Christophe Geissler. Stress testing electrical grids: Generative adversarial networks for load scenario generation. *Energy and AI*, 9:100177, 2022.

Litu Rout, Yujia Chen, Nataniel Ruiz, Abhishek Kumar, Constantine Caramanis, Sanjay Shakkottai, and Wen-Sheng Chu. Rb-modulation: Training-free stylization using reference-based modulation. In *International Conference on Learning Representations*, 2025.

Erwin Schrödinger. Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique. In *Annales de l'institut Henri Poincaré*, volume 2, pp. 269–310, 1932.

Raghav Singhal, Zachary Horvitz, Ryan Teehan, Mengye Ren, Zhou Yu, Kathleen McKeown, and Rajesh Ranganath. A general framework for inference-time scaling and steering of diffusion models. In *Forty-second International Conference on Machine Learning*, 2025. URL `https://openreview.net/forum?id=Jp988ELppQ`.

Jiaming Song, Qinsheng Zhang, Hongxu Yin, Morteza Mardani, Ming-Yu Liu, Jan Kautz, Yongxin Chen, and Arash Vahdat. Loss-guided diffusion models for plug-and-play controllable generation. In *International Conference on Machine Learning*, pp. 32483–32498. PMLR, 2023.

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

Zhiwei Tang, Jiangweizhi Peng, Jiasheng Tang, Mingyi Hong, Fan Wang, and Tsung-Hui Chang. Inference-time alignment of diffusion models with direct noise optimization. *arXiv preprint arXiv:2405.18881*, 2024.

Yusuke Tashiro, Jiaming Song, Yang Song, and Stefano Ermon. Csdi: Conditional score-based diffusion models for probabilistic time series imputation. *Advances in Neural Information Processing Systems*, 34:24804–24816, 2021.

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

Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2.

Bram Wallace, Akash Gokul, Stefano Ermon, and Nikhil Naik. End-to-end diffusion latent optimization improves classifier guidance. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 7246–7256. IEEE, 2023.

Haibo Wu, Marina I Knight, Keiland W Cooper, Norbert J Fortin, and Hernando Ombao. Wavelet canonical coherence for nonstationary signals. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

Zhisheng Xiao, Karsten Kreis, and Arash Vahdat. Tackling the generative learning trilemma with denoising diffusion gans. In *International Conference on Learning Representations*, 2021.

Lingxiao Yang, Shutong Ding, Yifan Cai, Jingyi Yu, Jingya Wang, and Ye Shi. Guidance with spherical gaussian constraint for conditional diffusion. In *Proceedings of the 41st International Conference on Machine Learning*, pp. 56071–56095, 2024.

Xinyu Yuan and Yan Qiao. Diffusion-TS: Interpretable diffusion for general time series generation. In *The Twelfth International Conference on Learning Representations*, 2024. URL `https://openreview.net/forum?id=4h1apFjO99`.

Zhihan Yue, Yujing Wang, Juanyong Duan, Tianmeng Yang, Congrui Huang, Yunhai Tong, and Bixiong Xu. Ts2vec: Towards universal representation of time series. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pp. 8980–8987, 2022.

Yuchen Zhu, Wei Guo, Jaemoo Choi, Guan-Horng Liu, Yongxin Chen, and Molei Tao. Mdns: Masked diffusion neural sampler via stochastic optimal control. *arXiv preprint arXiv:2508.10684*, 2025.

Juntang Zhuang, Nicha Dvornek, Xiaoxiao Li, Sekhar Tatikonda, Xenophon Papademetris, and James Duncan. Adaptive checkpoint adjoint method for gradient estimation in neural ode. In *International Conference on Machine Learning*, pp. 11639–11649. PMLR, 2020.

A PRELIMINARIES

Problem Setup and Notation. In Appendices A and B, time runs backward from noise at $t = 1$ to data at $t = 0$, as in the main text. Let $\mathbb{P}_{\text{base}}$ be the path measure on trajectories $X_{0:1} = \{X_t\}_{t\in[0,1]}$ induced by the uncontrolled diffusion

$$\mathrm{d}X_t = f_t(X_t)\mathrm{d}t + \sigma(t)\mathrm{d}W_t, \tag{17}$$

and let $\mathbb{Q}_u$ be the path measure of the controlled diffusion

$$\mathrm{d}X_t = \big(f_t(X_t) + \sigma(t)u_t(X_t)\big)\mathrm{d}t + \sigma(t)\mathrm{d}W_t, \tag{18}$$

started from the same noise marginal at $t = 1$. We assume u is progressively measurable and satisfies Novikov's condition, so that $\mathbb{Q}_u \ll \mathbb{P}_{\text{base}}$. With a terminal cost $g(X_0)$, we study the entropic stochastic optimal control (SOC) objective

$$\min_u \ \mathbb{E}_{\mathbb{Q}_u}\left[\int_0^1 \frac{1}{2}\|u_t(X_t)\|^2\mathrm{d}t + g(X_0)\right]. \tag{19}$$

A.1 Details of Derivations for the SOC Interpretation and Terminal Tilting

This appendix provides a self-contained derivation of the entropic stochastic optimal control (SOC) objective and its equivalent path-space variational formulation. Our goal is to make explicit the connection between drift-controlled diffusions, quadratic control energy, and relative entropy on trajectory space. In particular, we show that the SOC objective can be written as a Kullback–Leibler divergence between path measures, and that the resulting optimization admits a closed-form solution via exponential tilting of a reference diffusion.

A.1.1 Entropic SOC Objective and Path-space Formulation

Girsanov Theorem and KL Decomposition. Because Equation 17 and Equation 18 share the same diffusion coefficient, Girsanov's theorem gives the likelihood ratio on path space:

$$\frac{\mathrm{d}\mathbb{Q}_u}{\mathrm{d}\mathbb{P}_{\mathrm{base}}}(X_{0:1}) = \exp\left(\int_0^1 u_t(X_t)^\top \mathrm{d}W_t - \frac{1}{2}\int_0^1 \|u_t(X_t)\|^2 \mathrm{d}t\right), \tag{20}$$

where the stochastic integral is understood under $\mathbb{P}_{\mathrm{base}}$.

By definition,

$$D_{\mathrm{KL}}(\mathbb{Q}_u\|\mathbb{P}_{\mathrm{base}}) = \mathbb{E}_{\mathbb{Q}_u}\left[\log \frac{\mathrm{d}\mathbb{Q}_u}{\mathrm{d}\mathbb{P}_{\mathrm{base}}}\right]. \tag{21}$$

A standard consequence of Equation 20 is

$$D_{\mathrm{KL}}(\mathbb{Q}_u\|\mathbb{P}_{\mathrm{base}}) = \mathbb{E}_{\mathbb{Q}_u}\left[\frac{1}{2}\int_0^1 \|u_t(X_t)\|^2 \mathrm{d}t\right] + D_{\mathrm{KL}}(q_1\|p_1^{\mathrm{base}}), \tag{22}$$

where q_1 and p_1^{base} are the marginals of X_1 under $\mathbb{Q}_u$ and $\mathbb{P}_{\mathrm{base}}$. Equation 22 follows from the chain rule of the KL divergence after disintegrating both path measures with respect to X_1, together with the conditional Girsanov theorem. In our setting, the controlled sampling starts from the same prior/noise marginal as the reference process, i.e., $q_1 = p_1^{\mathrm{base}}$. Then the boundary term vanishes and we obtain

$$D_{\mathrm{KL}}(\mathbb{Q}_u\|\mathbb{P}_{\mathrm{base}}) = \mathbb{E}_{\mathbb{Q}_u}\left[\frac{1}{2}\int_0^1 \|u_t(X_t)\|^2 \mathrm{d}t\right]. \tag{23}$$

Consider the entropic stochastic optimal control objective in Equation 19. Using Equation 23, this is equivalent to the path-space variational problem

$$\min_{\mathbb{Q}} \; D_{\mathrm{KL}}(\mathbb{Q}\|\mathbb{P}_{\mathrm{base}}) + \mathbb{E}_{\mathbb{Q}}[g(X_0)], \qquad \text{s.t. } \mathbb{Q}(X_1) = \mathbb{P}_{\mathrm{base}}(X_1). \tag{24}$$

Indeed, any admissible u induces some $\mathbb{Q}_u$, and conversely any $\mathbb{Q}$ absolutely continuous w.r.t. $\mathbb{P}_{\mathrm{base}}$ with the same X_1 marginal admits a drift perturbation representation of the form Equation 18 (under standard regularity assumptions).

A.1.2 Closed-form Optimal Path Measure and Terminal Tilting

We next show that the minimizer of Equation 24 has a simple closed form.

Variational derivation. Let $\rho(X_{0:1}) := \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}_{\mathrm{base}}}(X_{0:1})$. Then $\mathbb{Q}$ is feasible iff $\rho \geq 0$ and $\mathbb{E}_{\mathbb{P}_{\mathrm{base}}}[\rho] = 1$, together with the marginal constraint $\mathbb{Q}(X_1) = \mathbb{P}_{\mathrm{base}}(X_1)$. Ignoring the marginal constraint for a moment (it will be handled below), the unconstrained objective reads

$$D_{\mathrm{KL}}(\mathbb{Q}\|\mathbb{P}_{\mathrm{base}}) + \mathbb{E}_{\mathbb{Q}}[g(X_0)] = \mathbb{E}_{\mathbb{P}_{\mathrm{base}}}[\rho\log\rho] + \mathbb{E}_{\mathbb{P}_{\mathrm{base}}}[\rho g(X_0)]. \tag{25}$$

Introduce a Lagrange multiplier η for normalization $\mathbb{E}_{\mathbb{P}_{\mathrm{base}}}[\rho] = 1$:

$$\mathcal{L}(\rho) = \mathbb{E}_{\mathbb{P}_{\mathrm{base}}}[\rho\log\rho + \rho g(X_0)] + \eta\left(\mathbb{E}_{\mathbb{P}_{\mathrm{base}}}[\rho] - 1\right). \tag{26}$$

Taking the pointwise first variation yields the stationary condition

$$\log\rho(X_{0:1}) + 1 + g(X_0) + \eta = 0, \tag{27}$$

so the optimizer must satisfy

$$\rho^\star(X_{0:1}) = \frac{1}{Z}\exp\big(-g(X_0)\big), \qquad Z := \mathbb{E}_{\mathbb{P}_{\mathrm{base}}}\big[\exp\big(-g(X_0)\big)\big]. \tag{28}$$

Thus, the optimal path measure is the exponentially tilted measure

$$\mathrm{d}\mathbb{Q}^\star(X_{0:1}) = \frac{1}{Z}\exp\big(-g(X_0)\big)\mathrm{d}\mathbb{P}_{\mathrm{base}}(X_{0:1}). \tag{29}$$

Terminal state tilt. Marginalizing Equation 29 over $X_{(0,1]}$ gives the induced law at $t = 0$,

$$q^\star(x_0) = \frac{1}{Z}\, p_{\text{base}}(x_0) \exp\big(-g(x_0)\big), \tag{30}$$

which is the tilted posterior of the main text with $g = \lambda L(\cdot; \mathcal{C})$.

Handling the fixed X_1 marginal. The derivation above yields the unique minimizer among all absolutely continuous measures without the X_1 constraint. When we impose $\mathbb{Q}(X_1) = \mathbb{P}_{\text{base}}(X_1)$, the optimal solution corresponds to the Doob h-transform of $\mathbb{P}_{\text{base}}$ with

$$h_t(x) := \mathbb{E}_{\mathbb{P}_{\text{base}}}\left[\exp\big(-g(X_0)\big) \mid X_t = x\right]. \tag{31}$$

Then the optimal controlled process under $\mathbb{Q}^\star$ is Markov and admits the drift correction

$$\mathrm{d}X_t = \Big(f_t(X_t) + \sigma(t)^2 \nabla_x \log h_t(X_t)\Big)\mathrm{d}t + \sigma(t)\mathrm{d}W_t, \tag{32}$$

which is the standard Schrödinger bridge / entropic control solution. (When $\sigma(t)$ is scalar; for matrix-valued diffusion, replace $\sigma(t)^2$ with $\sigma(t)\sigma(t)^\top$.)

A.2 Details of Derivations for Adjoint

This appendix derives adjoint optimality conditions in a *standard diffusion* setup (forward SDE on $t \in [0, 1]$ with terminal cost at $t = 1$). The main text uses reverse-time sampling with terminal cost at $t = 0$; we conclude by stating the direct mapping from the standard setup to our reverse-time parameterization.

A.2.1 HJB Derivation and Optimal Feedback Control

Define the value function

$$V(x, t) := \inf_u \mathbb{E}\left[\int_t^1 \frac{1}{2}\|u_s(X_s)\|^2 \mathrm{d}s + g(X_1) \;\middle|\; X_t = x\right]. \tag{33}$$

Then V satisfies the Hamilton–Jacobi–Bellman (HJB) equation (Pavliotis, 2014)

$$-\partial_t V(x, t) = \inf_{u \in \mathbb{R}^d} \left\{\frac{1}{2}\|u\|^2 + \big\langle \nabla_x V(x, t), f(x, t) + \sigma(t)u\big\rangle + \frac{1}{2}\mathrm{Tr}\big(\sigma(t)\sigma(t)^\top \nabla_x^2 V(x, t)\big)\right\}, \tag{34}$$

with $V(x, 1) = g(x)$. The minimizer of the quadratic expression in u is obtained by stationarity:

$$0 = \nabla_u\Big(\tfrac{1}{2}\|u\|^2 + \langle \nabla V, \sigma u\rangle\Big) = u + \sigma(t)^\top \nabla_x V(x, t), \tag{35}$$

hence the optimal feedback control is

$$u_t^\star(x) = -\sigma(t)^\top \nabla_x V(x, t). \tag{36}$$

In the scalar isotropic case $\sigma(t) = \sigma(t)I$, Equation 36 reduces to $u_t^\star(x) = -\sigma(t)\nabla_x V(x, t)$.

A.2.2 Pontryagin Maximum Principle and the Adjoint (Costate)

For computational implementation in samplers, we also derive the adjoint variable along discrete trajectories using Pontryagin's Maximum Principle (Pontryagin, 2018), which will be used in the iterative refinement procedure.

Hamiltonian. To apply PMP, we introduce the *adjoint* (or costate) variable $a_t \in \mathbb{R}^d$, which tracks the sensitivity of the optimal cost with respect to the state x_t. We then define the Hamiltonian as

$$\mathcal{H}(x, u, a, t) := \frac{1}{2}\|u\|^2 + \langle a, f(x, t) + \sigma(t)u\rangle. \tag{37}$$

The optimal control will later be characterized as the one that minimizes $\mathcal{H}$ along the trajectory.

Stationarity condition. Optimality requires $\nabla_u \mathcal{H} = 0$, i.e.,

$$\nabla_u \mathcal{H} = u + \sigma(t)^\top a = 0 \qquad \Longrightarrow \qquad u_t^\star = -\sigma(t)^\top a_t. \tag{38}$$

Thus the optimal control is determined by the adjoint.

Adjoint dynamics. In stochastic control, $a_t = \nabla_x V(X_t, t)$ along the optimal process satisfies a backward SDE; when $\sigma(t)$ is state-independent (as in diffusion models), it reads

$$\mathrm{d}a_t = -\big(\nabla_x f(X_t, t)\big)^\top a_t \mathrm{d}t + \nabla_x^2 V(X_t, t)\sigma(t) dW_t, \qquad a_1 = \nabla g(X_1). \tag{39}$$

For our purposes, we use the *pathwise sensitivity* form that corresponds to backpropagating through the sampler's deterministic transition map (mean update); this yields the deterministic adjoint ODE:

$$\dot{a}_t = -\big(\nabla_x f(X_t, t)\big)^\top a_t, \qquad a_1 = \nabla g(X_1). \tag{40}$$

Equation 40 is exactly the continuous-time analogue of reverse-mode differentiation through the flow induced by the drift.

A.2.3 Discrete-time Adjoint

In diffusion samplers we have a discrete trajectory $\{x_k\}_{k=0}^K$ generated by update maps $x_{k+1} = F_k(x_k)$ (deterministic part of the per-step update map F_k, ignoring additive noise). For a terminal loss $\ell(x_K) = g(x_K)$, define

$$a_K := \nabla_{x_K} \ell(x_K), \tag{41}$$

and propagate adjoints by vector–Jacobian products (VJPs):

$$a_k = \big(\nabla_x F_k(x_k)\big)^\top a_{k+1}, \qquad k = K-1, \dots, 0. \tag{42}$$

This is precisely what automatic differentiation computes when we evaluate $\mathrm{VJP}(F_k, x_k, a_{k+1})$.

A.2.4 How This Converts to Our Reverse-time Setup

The main text uses reverse-time sampling with terminal cost at $t = 0$ and a fixed pretrained reverse drift $\tilde{f}_t(x)$. Concretely, we write

$$\mathrm{d}X_t = \big(\tilde{f}_t(X_t) + \sigma(t) u_t(X_t)\big)\mathrm{d}t + \sigma(t)\mathrm{d}W_t, \qquad t \in [0, 1], \tag{43}$$

and the terminal cost is $L(x_0; \mathcal{C})$ at $t = 0$.

To map the standard derivation above to Equation 43, take the standard setup Equation 18–Equation 19 and apply the time-reversal change of variables $\tau = 1 - t$, $Y_\tau = X_{1-\tau}$:

- The forward drift f is replaced by the reverse drift $\tilde{f}$ (fixed by the pretrained diffusion model).
- The terminal condition $a_1 = \nabla g(X_1)$ becomes an adjoint initialized at $t = 0$ from the gradient of the terminal cost $g = \lambda L(\cdot; \mathcal{C})$:
$$a_0 = \lambda \nabla_{x_0} L(x_0; \mathcal{C}). \tag{44}$$
- The adjoint matching rule keeps the same *scaling*:
$$u_t^\star(x_t) = -\sigma(t)^\top a_t, \tag{45}$$
i.e., the control in Equation 43 is proportional to the adjoint with a factor $\sigma(t)^\top$.

Finally, in discrete-time samplers (DDPM/DDIM), we propagate a backward along the trajectory $\{x_k\}$ using VJPs. This yields the discrete adjoint used in our iterative refinement procedure.

B Proof

Setup and Assumption.

Assumption B.1. Let $X_{0:1}$ be a trajectory governed by the base reverse diffusion with path measure $\mathbb{P}_{\mathrm{base}}$. Define the unnormalized Doob's h-function $h_t(x_t; \mathcal{C})$ at time $t \in [0, 1]$ as

$$h_t(x_t; \mathcal{C}) \coloneqq \mathbb{E}_{\mathbb{P}_{\mathrm{base}}}\Big[\exp(-\lambda L(X_0; \mathcal{C})) \;\Big|\; X_t = x_t\Big]. \tag{46}$$

Assume $h_t(x_t; \mathcal{C}) \in (0, \infty)$ and is smooth enough so that $\nabla_{x_t} \log h_t$ exists.

B.1 Understanding Guidance and Projection in SOC

B.1.1 Posterior tilting as a control-energy trade-off

Proposition B.2. *Consider the following entropic SOC problem*

$$\min_u \ \mathbb{E}_{\mathbb{Q}_u}\left[\int_0^1 \frac{1}{2}\|u_t(X_t)\|^2 dt + \lambda L(X_0;\mathcal{C})\right], \tag{47}$$

where $\lambda \geq 0$ is a weighting parameter. Under Assumption B.1, the control weight λ directly governs the degree of posterior tilting, interpolating between the base distribution ($\lambda = 0$) and strict constraint enforcement ($\lambda \to \infty$).

Proof. When $\lambda = 0$, the terminal cost vanishes, so $u^\star = 0$ and $\mathbb{Q}_{u^\star} = \mathbb{P}_{\text{base}}$. For $\lambda > 0$, the optimal terminal law is $p^\star_\lambda(x_0) \propto p_{\text{base}}(x_0)\exp(-\lambda L(x_0;\mathcal{C}))$ (Appendix A.1.2). Assume $L \geq 0$ with $\mathcal{C} = \{L = 0\}$, and $\mathbb{P}_{\text{base}}(L(X_0;\mathcal{C}) \leq \delta) > 0$ for every $\delta > 0$ (the constraint set is approachable under the base law, which holds even when $\mathcal{C}$ has measure zero). Fix $\varepsilon > 0$ and choose $\delta \in (0, \varepsilon)$. Then

$$p^\star_\lambda\big(L(X_0;\mathcal{C}) > \varepsilon\big) = \frac{\mathbb{E}_{\mathbb{P}_{\text{base}}}\left[e^{-\lambda L}\mathbf{1}_{L>\varepsilon}\right]}{\mathbb{E}_{\mathbb{P}_{\text{base}}}\left[e^{-\lambda L}\right]} \leq \frac{e^{-\lambda\varepsilon}}{e^{-\lambda\delta}\,\mathbb{P}_{\text{base}}(L \leq \delta)} = \frac{e^{-\lambda(\varepsilon-\delta)}}{\mathbb{P}_{\text{base}}(L \leq \delta)} \xrightarrow{\lambda\to\infty} 0. \tag{48}$$

Hence $p^\star_\lambda$ concentrates on $\mathcal{C}$ as $\lambda \to \infty$, so λ interpolates between the base distribution and strict constraint enforcement. If moreover $\mathbb{P}_{\text{base}}(X_0 \in \mathcal{C}) > 0$, the same argument gives weak convergence of $p^\star_\lambda$ to $p_{\text{base}}(\cdot \mid X_0 \in \mathcal{C})$. □

B.1.2 Simple Example of Projection-Induced Distortion

Consider $X \sim \mathcal{N}(0,1)$ and the condition $X > 0$. The correct conditional distribution is the truncated normal

$$p(x \mid x > 0) = 2\phi(x), \quad x > 0,$$

where $\phi(x)$ denotes the probability density function of the standard normal distribution. In contrast, a projection-based procedure that maps samples as $x \mapsto \max(x, 0)$ collapses all probability mass from $x < 0$ onto the boundary at $x = 0$, producing an artificial atom at the constraint boundary and failing to recover the true conditional distribution.

B.1.3 No exact feasibility from drift control

Proposition B.3. *Let $\mathbb{P}_{\text{base}}$ be the path measure of the uncontrolled reverse diffusion in Equation 1, and let $\mathbb{Q}^u$ be the path measure induced by an admissible control u through Equation 3. Assume (i) $\sigma(t)\sigma(t)^\top \succ 0$ for a.e. $t \in [0,1]$, (ii) u satisfies Novikov's condition, $\mathbb{E}_{\mathbb{P}_{\text{base}}}\exp\left(\frac{1}{2}\int_0^1 \|u_t(X_t)\|^2 dt\right) < \infty$, and (iii) the terminal law $\mathbb{P}_{\text{base},0}$ of the uncontrolled process admits a density with respect to Lebesgue measure on $\mathbb{R}^{T\times C}$. Let $c : \mathbb{R}^{T\times C} \to \mathbb{R}^m$ be a constraint map and $S = \{x : c(x) = b\}$ the corresponding feasible set. If $\text{Leb}(S) = 0$, then the terminal law $\mathbb{Q}^u_0$ satisfies $\mathbb{Q}^u_0(S) = 0$.*

Proof. By (ii) and Girsanov's theorem, the process $W^u_t = W_t - \int_0^t u_s(X_s)\,ds$ is a Brownian motion under the measure $\mathbb{Q}^u$ defined by

$$\frac{d\mathbb{Q}^u}{d\mathbb{P}_{\text{base}}} = \exp\Big(\int_0^1 u_t(X_t)^\top dW_t - \tfrac{1}{2}\int_0^1 \|u_t(X_t)\|^2 dt\Big), \tag{49}$$

and X solves Equation 3 under $\mathbb{Q}^u$. The density is strictly positive and $\mathbb{P}_{\text{base}}$-integrable, so $\mathbb{Q}^u$ and $\mathbb{P}_{\text{base}}$ are mutually absolutely continuous on path space. In particular, for any measurable A in path space, $\mathbb{P}_{\text{base}}(A) = 0 \Rightarrow \mathbb{Q}^u(A) = 0$.

Let π_0 denote the evaluation map $X \mapsto X_0$, so that $\mathbb{Q}^u_0 = \mathbb{Q}^u \circ \pi_0^{-1}$ and $\mathbb{P}_{\text{base},0} = \mathbb{P}_{\text{base}} \circ \pi_0^{-1}$. Taking $A = \pi_0^{-1}(B)$ for a Borel set $B \subseteq \mathbb{R}^{T\times C}$ gives

$$\mathbb{P}_{\text{base},0}(B) = 0 \Longrightarrow \mathbb{Q}^u_0(B) = 0, \tag{50}$$

that is, $\mathbb{Q}_0^u \ll \mathbb{P}_{\text{base},0}$.

By (iii), $\mathbb{P}_{\text{base},0} \ll \text{Leb}$, and absolute continuity is transitive, so $\mathbb{Q}_0^u \ll \text{Leb}$. Since $\text{Leb}(S) = 0$ by assumption, it follows that $\mathbb{Q}_0^u(S) = 0$. □

B.1.4 LOCAL CONTROL IS PATH-SPACE SUBOPTIMAL

Theorem B.4 (Local control is path-space suboptimal)**.** *Let $\tilde{L}_t(x_t)$ be a surrogate depending only on the current state, inducing $u_t^{\text{local}}(x_t) = -\sigma(t)^\top \lambda \nabla_{x_t} \tilde{L}_t(x_t)$. Under Assumption B.1, the excess objective value of u^{local} over the optimum equals*

$$D_{\text{KL}}\left(\mathbb{Q}_{u^{\text{local}}} \| \mathbb{Q}_{u^\star}\right) = \tfrac{1}{2}\,\mathbb{E}_{\mathbb{Q}_{u^{\text{local}}}} \int_0^1 \left\| u_t^{\text{local}} - u_t^\star \right\|^2 dt, \tag{51}$$

which vanishes if and only if $\lambda \tilde{L}_t(x_t) = -\log h_t(x_t; \mathcal{C}) + c(t)$ for a.e. (x_t, t) and some scalar $c(t)$.

Proof. **Step 1 (Optimal Path Measure).** By Girsanov's theorem and the shared initial law at $t = 1$, $D_{\text{KL}}(\mathbb{Q}_u \| \mathbb{P}_{\text{base}}) = \frac{1}{2}\mathbb{E}_{\mathbb{Q}_u} \int_0^1 \|u_t(X_t)\|^2 dt$, so the objective in Equation 5 equals $J(u) = D_{\text{KL}}(\mathbb{Q}_u \| \mathbb{P}_{\text{base}}) + \mathbb{E}_{\mathbb{Q}_u}[\lambda L(X_0; \mathcal{C})]$. Define $\mathbb{Q}^\star$ by

$$\frac{\mathrm{d}\mathbb{Q}^\star}{\mathrm{d}\mathbb{P}_{\text{base}}}(X_{0:1}) = \frac{\exp(-\lambda L(X_0; \mathcal{C}))}{h_1(X_1; \mathcal{C})}, \tag{52}$$

which is well defined by Assumption B.1 and keeps the initial law at $t = 1$ unchanged. Then for any admissible u,

$$D_{\text{KL}}(\mathbb{Q}_u \| \mathbb{Q}^\star) = \mathbb{E}_{\mathbb{Q}_u}\left[\log \tfrac{\mathrm{d}\mathbb{Q}_u}{\mathrm{d}\mathbb{P}_{\text{base}}} + \lambda L(X_0; \mathcal{C})\right] + \mathbb{E}_{p_1}\left[\log h_1(X_1; \mathcal{C})\right] = J(u) + \text{const}, \tag{53}$$

so minimizing J is equivalent to minimizing $D_{\text{KL}}(\mathbb{Q}_u \| \mathbb{Q}^\star)$, whose minimum zero is attained exactly at $\mathbb{Q}_u = \mathbb{Q}^\star$. This is the Gibbs variational principle (Bhattacharya & Martin, 2022); for a memoryless base process $h_1 \equiv Z$ and $\mathbb{Q}^\star$ reduces to the tilt of Appendix A.1.2.

Step 2 (Optimal Control). $\mathbb{Q}^\star$ is the Doob h-transform of $\mathbb{P}_{\text{base}}$ with

$$h_t(x) = \mathbb{E}_{\mathbb{P}_{\text{base}}}[\exp(-\lambda L(X_0; \mathcal{C})) \mid X_t = x],$$

whose controlled SDE has control

$$u_t^\star(x) = \sigma(t)^\top \nabla_x \log h_t(x). \tag{54}$$

Step 3 (Gap Identity). Both $\mathbb{Q}_u$ and $\mathbb{Q}^\star$ solve SDEs with the same diffusion coefficient $\sigma(t)$, so Girsanov gives

$$D_{\text{KL}}(\mathbb{Q}_u \| \mathbb{Q}^\star) = \tfrac{1}{2}\mathbb{E}_{\mathbb{Q}_u} \int_0^1 \|u_t(X_t) - u_t^\star(X_t)\|^2 dt. \tag{55}$$

Combining with Equation 53, the excess objective value of any u over the optimum is exactly this quantity, and it vanishes if and only if $u_t = u_t^\star$ $\mathbb{Q}_u$-a.e.

Step 4 (Local Surrogates). Let $u_t^{\text{local}}(x) = -\sigma(t)^\top \lambda \nabla_x \tilde{L}_t(x)$ for a surrogate $\tilde{L}_t$ depending on x_t only. By Step 3, $\mathbb{Q}_{u^{\text{local}}} = \mathbb{Q}^\star$ iff $-\sigma(t)^\top \lambda \nabla_x \tilde{L}_t(x) = \sigma(t)^\top \nabla_x \log h_t(x)$ a.e. Since $\sigma(t)^\top$ has full column rank, this reduces to $\lambda \nabla_x \tilde{L}_t(x) = -\nabla_x \log h_t(x)$, and as $\mathbb{R}^d$ is connected,

$$\lambda \tilde{L}_t(x) = -\log h_t(x) + c(t) \tag{56}$$

for some scalar function $c(t)$.

Step 5 (Why this fails in practice). Guidance surrogates take the form $\lambda \tilde{L}_t(x_t) = \lambda L(\hat{x}_0(x_t); \mathcal{C})$ with a one-step denoiser $\hat{x}_0$, replacing the conditional expectation that defines $-\log h_t$ by a point estimate; by Jensen's inequality the two differ whenever $L(X_0; \mathcal{C})$ is non-degenerate given $X_t = x_t$. Projection surrogates enforce feasibility of the current state and do not follow h_t at all. Neither satisfies the condition of Step 4 in general, so both incur a positive path-space gap.

□

Remark B.5 (Terminal marginal vs. path measure). Equation 55 characterizes optimality *in path space*. The map from controls to terminal marginals is not injective, so a suboptimal control may still induce $p^\star$ at $t = 0$. For instance, in the linear-Gaussian case any state-independent control $u_t(x) = a(t)$ with the same value of $\int_0^1 \Phi(0, s)\sigma(s)a(s)\, ds$, where Φ is the state-transition matrix of the linear base drift, yields an identical terminal law while incurring different control energy. Local control therefore does not *guarantee* the correct terminal posterior, and the resulting discrepancy is measured directly in Appendix C.3.1.

B.2 Control Parameterization for Time Series

B.2.1 Approximation bias under control parameterization

Theorem B.6. *Let $\mathcal{R} : L^2 \to L^2$ be a linear, self-adjoint, idempotent operator with range $V := \mathrm{Range}(\mathcal{R})$, and let the base process be memoryless and $u^\star$ the optimal control of Equation 5. Define $v^\star = \mathcal{R}(u^\star)$ and let $\mathbb{Q}_0^{v^\star}$ be the terminal marginal of the diffusion controlled by $v^\star$. Then*

$$D_{\mathrm{KL}}(\mathbb{Q}_0^{v^\star} \| p^\star) \le \frac{1}{2} \|u^\star - v^\star\|_{L^2}^2, \tag{57}$$

where $p^\star$ is the target tilted distribution in Equation 2 and the norm is taken over $\mathbb{Q}^{v^\star} \otimes \mathrm{d}t$. The bound vanishes when $u^\star \in V$.

Proof. The proof proceeds in three steps: (i) compute the path-space KL divergence between $\mathbb{Q}^{v^\star}$ and $\mathbb{Q}^{u^\star}$ via Girsanov's theorem, (ii) apply the data-processing inequality to bound the terminal marginal KL, and (iii) use the fact that the unrestricted optimal control recovers the target posterior.

Step 1: Path-space KL via Girsanov's theorem. The controlled diffusion under any control u takes the form

$$\mathrm{d}X_t = (\tilde{f}_t(X_t) + \sigma(t)u_t(X_t))\mathrm{d}t + \sigma(t)\mathrm{d}W_t, \tag{58}$$

with $X_1 \sim \mathcal{N}(0, I)$ shared across all controls. Since $\mathbb{Q}^{u^\star}$ and $\mathbb{Q}^{v^\star}$ are induced by diffusions sharing the same diffusion coefficient $\sigma(t)$ and the same initial marginal at $t = 1$, Girsanov's theorem (Karatzas & Shreve, 2014) yields the Radon–Nikodym derivative

$$\frac{\mathrm{d}\mathbb{Q}^{v^\star}}{\mathrm{d}\mathbb{Q}^{u^\star}}(X_{0:1}) = \exp\Big(\int_0^1 (v_t^\star - u_t^\star)^\top \mathrm{d}\widetilde{W}_t - \frac{1}{2}\int_0^1 \|v_t^\star - u_t^\star\|^2 \mathrm{d}t \Big), \tag{59}$$

where $\widetilde{W}_t$ is the Wiener process under $\mathbb{Q}^{u^\star}$. Under $\mathbb{Q}^{v^\star}$, the Girsanov drift correction gives $\mathrm{d}\widetilde{W}_t = \mathrm{d}W_t^{v^\star} + (v_t^\star - u_t^\star)\mathrm{d}t$, where $W^{v^\star}$ is the Wiener process under $\mathbb{Q}^{v^\star}$. Substituting, and using that the stochastic integral against $W^{v^\star}$ has zero mean under $\mathbb{Q}^{v^\star}$,

$$\begin{aligned} D_{\mathrm{KL}}(\mathbb{Q}^{v^\star} \| \mathbb{Q}^{u^\star}) &= \mathbb{E}_{\mathbb{Q}^{v^\star}}\left[\int_0^1 (v_t^\star - u_t^\star)^\top \mathrm{d}\widetilde{W}_t - \frac{1}{2}\int_0^1 \|v_t^\star - u_t^\star\|^2 \mathrm{d}t \right] \\ &= \mathbb{E}_{\mathbb{Q}^{v^\star}}\left[\int_0^1 \|v_t^\star - u_t^\star\|^2 \mathrm{d}t - \frac{1}{2}\int_0^1 \|v_t^\star - u_t^\star\|^2 \mathrm{d}t \right] \\ &= \frac{1}{2}\mathbb{E}_{\mathbb{Q}^{v^\star}}\left[\int_0^1 \|v_t^\star - u_t^\star\|^2 \mathrm{d}t \right] = \frac{1}{2}\|u^\star - v^\star\|_{L^2}^2. \end{aligned} \tag{60}$$

Step 2: Data-processing inequality on the terminal marginal. The evaluation map $\pi_0 : X_{0:1} \mapsto X_0$ is measurable, and the terminal marginal of a path measure $\mathbb{Q}$ is its pushforward $\mathbb{Q}_0 = \mathbb{Q} \circ \pi_0^{-1}$. Since KL divergence cannot increase under a measurable map,

$$D_{\mathrm{KL}}(\mathbb{Q}_0^{v^\star} \| \mathbb{Q}_0^{u^\star}) \le D_{\mathrm{KL}}(\mathbb{Q}^{v^\star} \| \mathbb{Q}^{u^\star}). \tag{61}$$

Step 3: Optimal control recovers the target posterior. By Appendix A.1.2, under a memoryless base process the optimal control $u^\star$ of Equation 5 induces the target tilted distribution at $t = 0$:

$$\mathbb{Q}_0^{u^\star} = p^\star. \tag{62}$$

Substituting Equation 60 and Equation 62 into Equation 61 gives

$$D_{\mathrm{KL}}(\mathbb{Q}_0^{v^\star}\|p^\star) \le \frac{1}{2}\|u^\star - v^\star\|_{L^2}^2. \tag{63}$$

Vanishing condition. If $u^\star \in V$, then $\mathcal{R}(u^\star) = u^\star$ since $\mathcal{R}$ is the orthogonal projection onto V, so $\|u^\star - v^\star\|_{L^2} = 0$. By non-negativity of the KL divergence, $D_{\mathrm{KL}}(\mathbb{Q}_0^{v^\star}\|p^\star) = 0$, and the controlled marginal coincides with $p^\star$. □

Remark B.7 (Assumptions). The proof relies on three standard assumptions: (i) the difference $v^\star - u^\star$ satisfies Novikov's condition $\mathbb{E}\big[\exp\big(\frac{1}{2}\int_0^1 \|v_t^\star - u_t^\star\|^2 \mathrm{d}t\big)\big] < \infty$, so that the Radon–Nikodym derivative is a true martingale and $\mathbb{Q}^{v^\star} \ll \mathbb{Q}^{u^\star}$; (ii) $\mathcal{R}$ is linear, self-adjoint, and idempotent, so that $\mathcal{R}(u^\star)$ is the orthogonal projection of $u^\star$ onto V in L^2; and (iii) the base process is memoryless, so that $Q_0^{u^\star} = p^\star$ in Step 3.

B.2.2 Stationarity under parameterization $\mathcal{R}$

Proposition B.8. *Let $\mathcal{R}$ be linear, self-adjoint and idempotent, i.e., the orthogonal projection onto $V = \mathrm{Range}(\mathcal{R})$, and let $J_V(v) = \mathbb{E}_{\mathbb{Q}^v}[\lambda L(X_0;\mathcal{C})]$ for $v \in V$. Then the update $v^{(k+1)} = v^{(k)} + \gamma\mathcal{R}(-\sigma^\top a^{(k)})$ is a projected stochastic gradient step on J_V, and*

(i) *any fixed point $\bar{v}$ of its expected update satisfies $\mathcal{R}(\nabla J_V(\bar{v})) = 0$;*

(ii) *if $\sigma^\top a^{(k)}$ is an unbiased estimate of $\nabla J_V(v^{(k)})$ with $\mathbb{E}\|\mathcal{R}(\sigma^\top a^{(k)} - \nabla J_V(v^{(k)}))\|^2 \le s_V^2$, and J_V is β-smooth on V and bounded below by $J_V^\star$, then for $\gamma \le 1/\beta$,*

$$\min_{0\le k<K} \mathbb{E}\big\|\mathcal{R}\big(\nabla J_V(v^{(k)})\big)\big\|^2 \le \frac{2\big(J_V(v^{(0)}) - J_V^\star\big)}{\gamma K} + \beta\gamma\, s_V^2, \tag{64}$$

so $\gamma \propto 1/\sqrt{K}$ reaches an ε-stationary point of J_V within V in $O(1/\varepsilon^2)$ iterations.

Proof. **Part (i).** The adjoint propagated backward along the trajectory is the reverse-mode derivative of the terminal cost through the sampler, initialized at $a_0 = \lambda\nabla_{x_0}L(x_0;\mathcal{C})$, so

$$\nabla J_V(v) = \mathbb{E}_{\mathbb{Q}^v}\big[\sigma^\top a(v)\big]. \tag{65}$$

The update therefore reads

$$v^{(k+1)} = v^{(k)} - \gamma\,\mathcal{R}\big(\nabla J_V(v^{(k)})\big), \tag{66}$$

a gradient step with step size γ whose increment is projected onto V. Since $v^{(0)} = 0 \in V$ and every increment lies in $\mathrm{Range}(\mathcal{R}) = V$, all iterates stay in V.

A fixed point satisfies $\mathcal{R}\big(-\sigma^\top a(\bar{v})\big) = 0$, which by Equation 65 is $\mathcal{R}\big(\nabla J_V(\bar{v})\big) = 0$. Since $\mathcal{R}$ is linear and self-adjoint, $\mathrm{null}(\mathcal{R}) = \mathrm{Range}(\mathcal{R})^\perp = V^\perp$, so this is equivalent to

$$\big\langle \nabla J_V(\bar{v}), w\big\rangle_{L^2} = 0 \qquad \text{for all } w \in V, \tag{67}$$

the first-order stationarity condition of J_V restricted to V.

Part (ii). Write $\hat{g}_k \coloneqq \mathcal{R}(\sigma^\top a^{(k)})$ and $g_k \coloneqq \mathcal{R}\big(\nabla J_V(v^{(k)})\big)$, so that $v^{(k+1)} = v^{(k)} - \gamma\hat{g}_k$ and, by unbiasedness, $\mathbb{E}[\hat{g}_k \mid v^{(k)}] = g_k$. All iterates lie in V by part (i), so β-smoothness on V gives the descent lemma

$$J_V(v^{(k+1)}) \le J_V(v^{(k)}) - \gamma\big\langle \nabla J_V(v^{(k)}), \hat{g}_k\big\rangle + \frac{\beta\gamma^2}{2}\|\hat{g}_k\|^2. \tag{68}$$

Since $\mathcal{R}$ is self-adjoint and idempotent, $\mathbb{E}\big[\langle \nabla J_V(v^{(k)}), \hat{g}_k\rangle \mid v^{(k)}\big] = \langle \nabla J_V(v^{(k)}), \mathcal{R}\nabla J_V(v^{(k)})\rangle = \|g_k\|^2$, and the bias–variance decomposition gives $\mathbb{E}[\|\hat{g}_k\|^2 \mid v^{(k)}] \le \|g_k\|^2 + s_V^2$. Taking conditional expectations and using $\gamma \le 1/\beta$,

$$\mathbb{E}\big[J_V(v^{(k+1)}) \mid v^{(k)}\big] \le J_V(v^{(k)}) - \gamma\Big(1 - \frac{\beta\gamma}{2}\Big)\|g_k\|^2 + \frac{\beta\gamma^2}{2}s_V^2 \le J_V(v^{(k)}) - \frac{\gamma}{2}\|g_k\|^2 + \frac{\beta\gamma^2}{2}s_V^2. \tag{69}$$

Taking total expectations, summing over $k = 0, \dots, K-1$ and using $J_V \geq J_V^\star$,

$$\frac{\gamma}{2} \sum_{k=0}^{K-1} \mathbb{E}\|g_k\|^2 \leq J_V(v^{(0)}) - J_V^\star + \frac{K\beta\gamma^2}{2} s_V^2. \tag{70}$$

Bounding the minimum by the average and dividing by $\gamma K/2$ gives the stated bound. With $\gamma = \min\{1/\beta,\, c/\sqrt{K}\}$, the right-hand side is $O(1/\sqrt{K})$, so $\min_{k<K} \mathbb{E}\|g_k\|^2 \leq \varepsilon$ after $O(1/\varepsilon^2)$ iterations. □

Remark B.9 (Scope of the rate). Part (ii) requires $\gamma \leq 1/\beta$; since $a_0 = \lambda \nabla_{x_0} L$, the smoothness constant of J_V scales with λ, so large λ may leave this regime and the iterates then act as an increasing tilt rather than as descent on a fixed objective. The variance term is where the parameterization enters: since $\mathcal{R}$ is an orthogonal projection, $s_V^2 \leq \mathbb{E}\|\sigma^\top a^{(k)} - \nabla J_V(v^{(k)})\|^2$, so restricting the control to V can only reduce the noise floor $\beta\gamma s_V^2$, consistent with the injected-noise experiment of Table 13. Finally, the adjoint interpolation introduces a small bias that the unbiasedness assumption does not cover.

Remark B.10 (The iteration count as tilt: linear-Gaussian case). Consider a memoryless linear-Gaussian base whose terminal state is $x_0 = m + Gw + Gs$, where $w \sim \mathcal{N}(0, I)$ stacks the whitened injected noise and $s = \sqrt{\Delta t}\, v$ stacks the control entering through the same channels, so that $\Sigma = GG^\top$ is the base covariance. Let the constraint be $Ax_0 = b$ for a single row A, with $L = \frac{1}{2}(Ax_0 - b)^2$, $d^2 = A\Sigma A^\top$, and assume $G^\top A^\top \in V$ (e.g., Mean constraints). In the idealized setting with exact adjoints and a fresh noise draw at every iteration, the expected update is gradient descent on $\frac{\lambda}{2}(AGs - (b - Am))^2$. Starting from $v^{(0)} = 0$, after K iterations the mean of x_0 equals

$$m + \Sigma A^\top \frac{1 - \rho^K}{d^2}\,(b - Am), \qquad \rho = 1 - \gamma\lambda d^2 \in (0, 1),$$

which is exactly the mean of the tilted posterior $p^\star_{\lambda'}$ with $(1 + \lambda' d^2)^{-1} = \rho^K$, so λ' increases with K. The covariance, however, remains Σ, whereas that of $p^\star_{\lambda'}$ is contracted along $\Sigma A^\top$. The iterates therefore move along the tilted family in mean but do not contract the constrained direction; the terminal projection supplies this remaining contraction.

C Experiments

C.1 Experimental Setups

C.1.1 Task and Metrics

Constraint Types. We consider three constraint types following Coletta et al. (2024); Narasimhan et al. (2025), matching the temporal mean (**Mean**), the mean consecutive change (**ΔMean**), and the value at the argmax (**Max**). We evaluate constraint satisfaction following Narasimhan et al. (2025), using a tolerance of $\tau = 0.01$.

Let $\mathbf{x} \in \mathbb{R}^{T \times C}$ be a generated sample and $\mathbf{y} \in \mathbb{R}^{T \times C}$ the corresponding target, where $x_{t,c}$ denotes the element at timestep t and channel c. For each constraint type, we define a non-negative violation $V(\mathbf{x}) \geq 0$ that equals zero if and only if the constraint is satisfied within tolerance τ. We report the violation averaged over all N generated samples: $\bar{V} = \frac{1}{N} \sum_{i=1}^{N} V(\mathbf{x}^{(i)})$.

Mean. This constraint requires that the temporal mean of the generated sample matches that of the target. The temporal mean for channel c is $\bar{x}_c = \frac{1}{T} \sum_{t=1}^{T} x_{t,c}$, and analogously $\bar{y}_c$ for the target. The per-channel deviation is $d_c = |\bar{x}_c - \bar{y}_c|$. A sample satisfies the mean constraint if $d_c \leq \tau$ for all $c \in \{1, \dots, C\}$, and the violation is:

$$V_{\text{Mean}} = \sum_{c=1}^{C} \max(d_c - \tau,\ 0)$$

Δ Mean. This constraint targets the average rate of temporal change. Let $\Delta x_{t,c} = x_{t+1,c} - x_{t,c}$ be the first-order difference, and let $\overline{\Delta x}_c = \frac{1}{T-1}\sum_{t=1}^{T-1} \Delta x_{t,c}$ be its temporal mean. The per-channel deviation is $d_c = |\overline{\Delta x}_c - \overline{\Delta y}_c|$. A sample satisfies this constraint if $d_c \le \tau$ for all channels, and the violation follows the same form:

$$V_{\Delta\text{Mean}} = \sum_{c=1}^{C} \max(d_c - \tau,\ 0)$$

Max. This constraint enforces that the generated sample preserves both the location and the value of the temporal maximum from the target. Let $k_c^* = \arg\max_t\ y_{t,c}$ and $M_c^* = y_{k_c^*,c}$ denote the target's argmax index and maximum value for channel c. The argmax violation checks whether $x_{k_c^*,c}$ is indeed the maximum: $a_c = \max_t(x_{t,c} - x_{k_c^*,c})$, which equals zero if and only if no timestep exceeds the value at the target argmax position. The max value violation checks whether the maximum value matches the target: $m_c = |\max_t\ x_{t,c} - M_c^*|$. A sample satisfies this constraint if $a_c \le \tau$ and $m_c \le \tau$ for all channels, and the violation is:

$$V_{\text{Max}} = \sum_{c=1}^{C} [\max(a_c - \tau, 0) + \max(m_c - \tau, 0)]$$

Context-FID. Context-FID (Jeha et al., 2022) adapts the Fréchet Inception Distance (Heusel et al., 2017) to time series by replacing the Inception network with TS2Vec (Yue et al., 2022), a representation learning model trained on time-series data. We encode both real and synthetic series with a pretrained TS2Vec and compute the standard FID on the resulting representations. Lower values indicate that the synthetic distribution lies closer to the real distribution in the representation space.

Correlation Score. The correlation score (Liao et al., 2020) measures how well the cross-channel dependency structure of the synthetic data matches the real data. Following Yuan & Qiao (2024), we standardize each channel over samples and time, compute for every channel pair $i \ge j$ the lag-0 cross-correlation $\hat{\rho}_{ij}$ averaged over samples, and report

$$\frac{1}{10}\sum_{i\ge j}\left|\hat{\rho}_{ij}^{o} - \hat{\rho}_{ij}^{s}\right| \tag{71}$$

between the real (o) and synthetic (s) data, averaged over 5 random subsets. Since the score sums over channel pairs, its scale grows with the number of channels.

C.1.2 Implementation Details

Our experiments are conducted using PyTorch on four NVIDIA GeForce RTX 4090 GPUs with an Intel Xeon Gold 6226R CPU (2.90GHz). For the projection methods, we use SciPy (Virtanen et al., 2020) and CVXPY (Diamond & Boyd, 2016) on CPU, following the original code of COP and CPS.

In all experiments, we use a batch size of 256 and DDIM with 200 sampling steps and $\eta = 0.1$. Our analysis assumes a memoryless base process, which the DDPM reverse SDE ($\eta = 1$) approximately satisfies. We use $\eta = 0.1$, which gave better sample quality than $\eta = 1$ in preliminary experiments; the theory then describes the idealized objective, while feasibility is guaranteed by the terminal projection regardless. For CPS and COP, we adopt the hyperparameter settings of Narasimhan et al. (2025). For Diffusion-TS and GuidedDiffTime, we search the guidance weight over $\{1, 10, 100, 1000\}$ and select the best for each dataset and constraint; for Diffusion-TS we set the Langevin step size to 0.05, following its original setting.

For our methods, the step size enters the update only through the product $\gamma\lambda$ (since $a_0 = \lambda\nabla_{x_0}L$), so we fix $\gamma = 1$ and absorb the step scale into λ, which is tuned per setting (Table 5); for a fixed budget K, a constant step is also the choice that minimizes the bound of Proposition 4.2(ii). The control weight λ for each experiment is reported in Table 5. We use $K = 10$ refinement iterations unless stated otherwise ($K = 1$ for the single-pass variants). $\mathcal{R}_{\text{FFT}}$ uses a frequency budget of $k = 16$, and $\mathcal{R}_{\text{WAV}}$ uses the orthonormal Haar transform with level budget $J = 2$. The control is updated by accumulation, so that the drift perturbation is $\sigma(t)u_t = -\sigma(t)^2 a_t$. Since computing the adjoint at

every diffusion step is expensive in high dimension, we follow incomplete checkpointing strategies that reconstruct intermediate states by interpolation (Hückelheim & Müller, 2018; Daulbaev et al., 2020): we evaluate the adjoint every 20 steps and linearly interpolate the remainder, which reduces per-iteration cost without notable degradation in sample quality.

Table 5: Control weight λ for each experiment.

Method	Stock			ETTh			Energy			fMRI		
	Mean	ΔMean	Max	Mean	ΔMean	Max	Mean	ΔMean	Max	Mean	ΔMean	Max
TTC-FFT	1000	100	1000	1000	100	1000	1000	10000	1000	1000	1000	100
TTC-WAV	1000	10000	10000	1000	10000	1000	1000	1000	1000	1000	100	100

C.2 Ablation Studies

C.2.1 Ablation Studies without Final Projection

Trajectory control at matched compute. We provide a detailed comparison without applying the final projection step in Table 6. Here TTC-FFT$_s$ and TTC-WAV$_s$ denote single iteration variants, whose sampling cost is comparable to that of the guidance-based baselines. TTC-FFT and TTC-WAV use the full iteration budget. This lets us separate the benefit of trajectory-level control from the effect of simply spending more compute. Under a matched budget, a single iteration already suffices. Averaged over the 36 settings (4 datasets $\times$ 3 metric families $\times$ 3 statistics), the single iteration variants attain mean ranks of 3.39 and 3.47 against 4.28 for GuidedDiffTime and 4.58 for Diffusion-TS, and both of them beat both baselines in 18 settings (20 counting ties), with at least one doing so in 24. On mean constraint violation the single iteration variants already improve over the stronger baseline on all four datasets, most visibly on Stock ($5.552 \rightarrow 1.658$) and Energy ($12.017 \rightarrow 5.913$). At equal cost, therefore, the advantage stems from where the control is applied rather than from how much computation is spent.

Effect of the iteration budget. Increasing the iteration budget then trades compute for constraint satisfaction. Moving from one iteration to the full budget reduces violation in 11 of the 12 columns for each parameterization, the remaining column being already zero. Sample quality generally follows: the same move improves 27 of the 36 columns for each parameterization, and the full variants reach mean ranks of 2.14 and 2.47. The exceptions are local rather than systematic. On Stock, Context-FID under ΔMean and Max is lowest for a single iteration, and on ETTh-Max both the violation and Context-FID of the best baseline remain competitive; correlation on Stock-Mean and fMRI-Max also remains best for Diffusion-TS, indicating that the residual gap there is not a constraint-satisfaction issue. Taken together, the improvements are not an artifact of extra compute, but arise from trajectory-level control that aligns samples with the target constraints throughout generation.

Projection alone does not deliver feasibility. To further examine the effect of projection, we compare Table 6 against the main results in Table 1, in which a single penalty-based projection is applied at the final step. Because the two tables report the same samplers with and without that step, they form a paired ablation in which every method is observed immediately before and immediately after the same corrective operator. At a finite penalty, a single projection leaves a residual that grows with the violation of the incoming sample, so samplers that have not approached the feasible region retain violation after it. After projection, GuidedDiffTime retains a nonzero mean violation on all four datasets (0.834 on Stock, 0.459 on Energy and 0.365 on fMRI, 0.006 on ETTh) and Diffusion-TS on three (1.923, 1.636 and 0.013), whereas all four TTC variants reach exact satisfaction on every dataset and every violation statistic. The failure is concentrated on constraints defined through global or aggregate statistics, which a single local correction cannot repair when the incoming sample is far from the constraint set.

Control and projection are complementary. The pairing also indicates why proximity matters. Projection is an optimization over a high-dimensional space whose reliability degrades as the required displacement grows. For the TTC variants, whose samples already lie close to the tilted target, projection improves distributional fidelity alongside feasibility: mean Context-FID on Stock falls from 1.854 to 0.177 for TTC-FFT and from 9.711 to 0.192 for TTC-FFT$_s$, and on Energy from

6.806 to 1.783. For GuidedDiffTime, the same operator degrades quality while still failing to reach feasibility, with mean Context-FID rising from 465.188 to 608.311 on Stock, from 17.974 to 19.164 on Energy and from 19.181 to 20.429 on fMRI. Within each dataset the ordering is consistent with this reading: on Stock the TTC variants enter projection with mean violations between 0.650 and 1.749 and leave it exactly feasible, whereas GuidedDiffTime and Diffusion-TS enter at 5.552 and 7.569 and retain residuals. Across datasets the magnitude alone is not decisive, since a sample entering at 4.743 on Energy projects exactly while GuidedDiffTime entering at 0.681 on fMRI does not; what matters is whether the constraint set is reachable from the sample without leaving the data manifold. The control moves the sampling distribution close to the tilted target during generation, which makes a single terminal projection sufficient and inexpensive, and the projection converts that proximity into exact satisfaction, which control alone does not guarantee.

Table 6: All performance without projection. The best results are in **bold** and the second best are underlined. DownArrow (↓) indicates lower is better.

		Stock			ETTh			Energy			fMRI		
Metric	Method	Mean	ΔMean	Max	Mean	ΔMean	Max	Mean	ΔMean	Max	Mean	ΔMean	Max
Violation ↓	GuidedDiffTime	5.552	**0.000**	4.976	1.524	0.009	**1.646**	12.017	0.110	14.818	0.681	0.252	32.498
	Diffusion-TS	7.569	**0.000**	6.588	1.645	0.009	5.300	16.384	0.109	21.585	0.682	0.253	36.636
	TTC-FFT$_s$ (No proj)	1.749	**0.000**	4.885	1.467	0.009	4.243	5.914	0.106	15.108	0.677	0.253	15.337
	TTC-WAV$_s$ (No proj)	1.658	**0.000**	5.843	1.467	0.008	4.021	5.913	0.108	15.176	0.678	0.252	16.615
	TTC-FFT (No proj)	0.653	**0.000**	**3.949**	**1.088**	0.008	1.948	**4.743**	**0.094**	**14.046**	**0.649**	**0.247**	14.966
	TTC-WAV (No proj)	**0.650**	**0.000**	5.387	1.096	**0.005**	2.095	4.749	0.095	14.828	0.650	0.249	**14.889**
Context-FID ↓	GuidedDiffTime	465.188	367.728	295.443	12.041	12.592	10.811	17.974	17.604	18.513	19.181	19.235	14.212
	Diffusion-TS	773.924	470.734	389.098	13.550	14.372	**6.609**	46.035	20.226	15.403	**18.716**	19.361	18.140
	TTC-FFT$_s$ (No proj)	9.711	219.968	**43.942**	10.679	12.316	10.612	6.737	20.882	15.135	19.496	**18.854**	**2.386**
	TTC-WAV$_s$ (No proj)	10.981	363.200	48.738	10.103	12.887	9.029	6.895	**16.368**	**13.142**	19.068	19.667	2.814
	TTC-FFT (No proj)	1.854	252.430	69.799	**7.458**	**11.089**	8.610	6.806	18.048	15.609	18.836	19.590	2.472
	TTC-WAV (No proj)	**1.720**	**194.939**	238.138	7.664	11.756	8.769	**6.055**	20.852	15.002	19.172	19.251	2.643
Correlation ↓	GuidedDiffTime	0.262	0.371	0.290	0.317	0.330	0.319	8.024	**7.923**	7.713	3.455	3.419	3.404
	Diffusion-TS	**0.077**	0.279	0.292	0.340	0.315	0.383	8.274	7.994	8.552	**3.361**	3.422	**3.181**
	TTC-FFT$_s$ (No proj)	0.190	0.271	0.279	0.299	0.324	0.484	7.669	7.978	6.912	3.432	3.413	4.739
	TTC-WAV$_s$ (No proj)	0.187	**0.265**	0.394	0.288	0.324	0.442	7.729	8.109	7.205	**3.361**	3.417	4.574
	TTC-FFT (No proj)	0.221	0.269	0.120	0.128	**0.312**	0.273	5.846	7.939	7.424	3.415	**3.401**	4.561
	TTC-WAV (No proj)	0.230	0.286	**0.062**	**0.110**	0.318	**0.239**	**5.717**	7.997	**6.852**	3.405	3.414	4.592

Table 7: Effect of a single final projection on constraint violation. Each method is shown immediately before and immediately after the same projection operator. DownArrow (↓) indicates lower is better.

		Stock			ETTh			Energy			fMRI		
Metric	Method	Mean	ΔMean	Max	Mean	ΔMean	Max	Mean	ΔMean	Max	Mean	ΔMean	Max
Violation ↓	GuidedDiffTime	5.552	0.000	4.976	1.524	0.009	1.646	12.017	0.110	14.818	0.681	0.252	32.498
	+ projection	0.834	0.000	0.000	0.006	0.000	0.003	0.459	0.001	0.087	0.365	0.000	0.000
	Diffusion-TS	7.569	0.000	6.588	1.645	0.009	5.300	16.384	0.109	21.585	0.682	0.253	36.636
	+ projection	1.923	0.000	0.000	0.013	0.000	0.002	1.636	0.001	0.084	0.000	0.000	0.000
	TTC-FFT	0.653	0.000	3.949	1.088	0.008	1.948	4.743	0.094	14.046	0.649	0.247	14.966
	+ projection	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	TTC-WAV	0.650	0.000	5.387	1.096	0.005	2.095	4.749	0.095	14.828	0.650	0.249	14.889
	+ projection	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

C.2.2 Ablation Studies for Control Parameterization

We provide detailed per-metric and per-constraint comparisons that complement the average-rank summary in Section 5.3.

Table 8 reports violation, Context-FID, and correlation across all four datasets and three constraint types. All three variants achieve zero violation, so performance differences appear in distributional quality (Context-FID) and temporal consistency (correlation).

Compared to TTC-Direct, TTC-FFT and TTC-WAV achieve lower Context-FID in 14 of the 16 Mean and ΔMean comparisons, consistent with these constraints having gradients concentrated at low frequencies. For Max constraints, TTC-Direct remains competitive on Stock and Energy, suggesting that the parameterization contributes most where the constraint gradient is well aligned with the retained subspace. Correlation results show similar trends with dataset-dependent variations.

Table 9 reports the average rank of every method with a final projection applied. Our method without a basis transformation already outperforms all guidance and terminal-projection baselines on all

four datasets, by 1.22 to 2.66 rank positions on Stock, ETTh and Energy, and by a narrower 0.33 on fMRI. Its remaining competitor is CPS, which projects at every sampling step and which TTC-Direct trails on three of the four datasets (3.44 against 2.22 on ETTh, 2.78 against 2.44 on Energy, 3.56 against 2.89 on fMRI). Adding a basis parameterization closes that gap and reverses it: TTC-FFT and TTC-WAV improve on TTC-Direct on every dataset and rank ahead of CPS in all eight dataset-method comparisons, reaching average ranks of 2.03 and 2.22 against 2.92 for CPS. The pattern aligns with the discussion in Section 5.3, in that the SOC formulation alone supplies the foundation while the basis parameterization contributes a complementary inductive bias that a per-step projection schedule does not recover.

Table 8: Per-metric and per-constraint comparison of control parameterizations. All three variants achieve zero violation. The best results and the second best are in **bold** and underlined, respectively. DownArrow (↓) indicates lower is better.

		Violation ↓			Context-FID ↓			Correlation ↓			AvgRk ↓
Dataset	Method	Mean	ΔMean	Max	Mean	ΔMean	Max	Mean	ΔMean	Max	(Rank)
Stock	TTC-Direct	**0.000**	**0.000**	**0.000**	0.201	321.151	**11.367**	0.035	0.282	0.479	2.00 (3)
	TTC-FFT	**0.000**	**0.000**	**0.000**	**0.177**	**209.189**	16.235	0.034	**0.248**	**0.472**	**1.22 (1)**
	TTC-WAV	**0.000**	**0.000**	**0.000**	0.185	238.360	16.639	**0.029**	0.279	0.590	1.78 (2)
ETTh	TTC-Direct	**0.000**	**0.000**	**0.000**	7.743	13.454	7.078	0.189	0.326	0.437	2.22 (3)
	TTC-FFT	**0.000**	**0.000**	**0.000**	7.835	11.761	**5.092**	**0.167**	**0.310**	0.429	1.44 (2)
	TTC-WAV	**0.000**	**0.000**	**0.000**	**7.501**	**11.593**	6.832	0.168	0.313	**0.391**	**1.33 (1)**
Energy	TTC-Direct	**0.000**	**0.000**	**0.000**	1.905	18.584	8.470	1.272	8.027	6.655	2.00 (3)
	TTC-FFT	**0.000**	**0.000**	**0.000**	1.783	**17.130**	**6.681**	1.307	7.910	**5.894**	**1.44 (1)**
	TTC-WAV	**0.000**	**0.000**	**0.000**	**1.756**	18.771	9.975	**1.223**	**7.641**	6.318	1.56 (2)
fMRI	TTC-Direct	**0.000**	**0.000**	**0.000**	19.291	19.334	6.087	3.400	**3.270**	6.028	2.11 (3)
	TTC-FFT	**0.000**	**0.000**	**0.000**	**18.714**	18.692	**3.336**	**3.310**	3.340	4.749	**1.44 (1)**
	TTC-WAV	**0.000**	**0.000**	**0.000**	19.252	**18.446**	3.764	3.341	3.312	**4.621**	**1.44 (1)**

C.2.3 Spectral Structure of Constraint Gradients

To empirically support the spectral compactness analysis in Section 4.1, we visualize the spectral energy distribution of the raw adjoint $u^\star$ and the TTC-FFT $\mathcal{R}_{\text{FFT}}(u^\star)$ across constraints and datasets (Figure 4).

For Mean constraints, the raw adjoint concentrates almost entirely at the DC component, confirming that mean-type constraints have gradient support concentrated at frequency 0. TTC-FFT coincides with the raw adjoint, indicating that the approximation gap $\|u^\star - \mathcal{R}_{\text{FFT}}(u^\star)\|$ and hence the bias bound of Theorem 4.1 are essentially zero.

For ΔMean constraints, the constraint reduces to an endpoint difference $(x_{T,c} - x_{1,c})/(T-1)$ via telescoping, so its gradient consists of two boundary impulses. The resulting spectral energy spreads across frequencies due to the broadband nature of localized impulses in the Fourier domain. The TTC-FFT captures the dominant low-frequency components within the cutoff $k = 16$.

For Max constraints, the raw adjoint exhibits the spectral spread characteristic of time-localized signals, indicating that the Fourier basis is less aligned with this constraint type and motivating the wavelet parameterization; empirically, however, TTC-FFT remains competitive on Max (Section 5.2), since the retained $k = 16$ modes still capture most of the adjoint energy.

C.2.4 Effect of Iterative Refinement

We examine the iterative refinement procedure (Proposition 4.2) by varying the number of refinement iterations $K \in \{1, 5, 10\}$, with the terminal projection applied (Table 10). Here, TTC-FFT (K) and TTC-WAV (K) denote our method with K refinement iterations.

With the terminal projection, violation is zero from $K = 1$ in every setting, so a single adjoint pass already brings samples close enough for one projection to reach exact feasibility. Increasing K improves Context-FID in most settings: $K = 10$ is best in 20 of the 24 cases. The exceptions are ΔMean on Energy, where the constraint gradient spreads across frequencies rather than concentrating in the retained subspace, and two fMRI cases where $K = 5$ is marginally better.

Table 9: Average rank of all methods with a final projection, computed across violation, Context-FID, and correlation. The best results are in **bold**.

Dataset	GuidedDiffTime	Diffusion-TS	COP	COP-FT	CPS	TTC-Direct	TTC-FFT	TTC-WAV
Stock	4.67	5.11	4.56	4.22	4.11	3.00	**1.78**	2.89
ETTh	5.33	5.22	5.67	6.44	2.22	3.44	2.11	**1.89**
Energy	7.22	6.00	6.00	5.44	2.44	2.78	2.00	**1.89**
fMRI	5.22	3.89	5.44	5.78	2.89	3.56	**2.22**	**2.22**
Avg	5.64	5.08	5.42	5.47	2.92	3.20	**2.03**	2.22

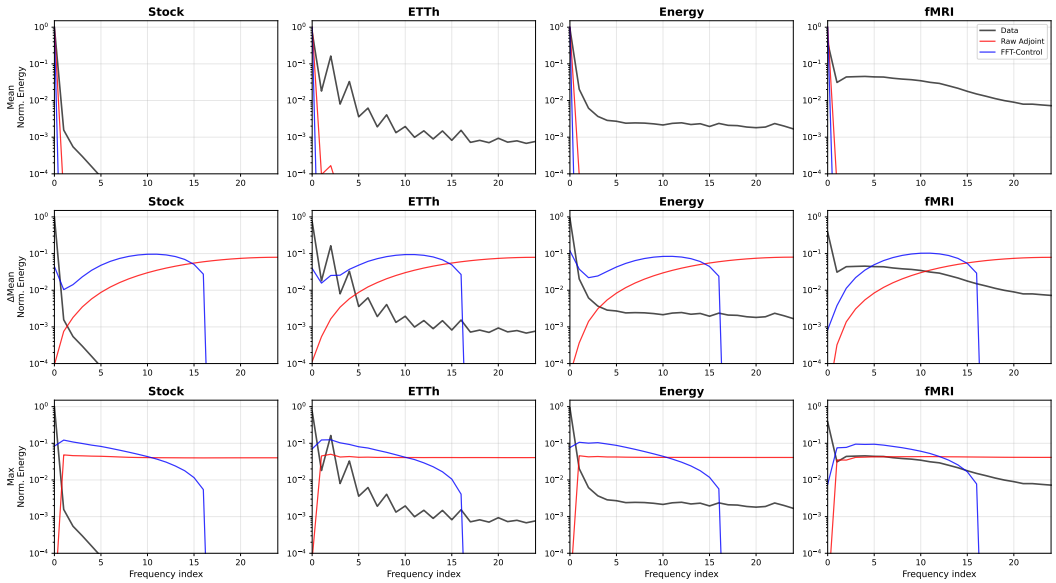

Figure 4: Spectral energy distribution of the raw adjoint $u^\star$ and the FFT-projected control $\mathcal{R}_{\text{FFT}}(u^\star)$ across datasets and constraint types. The adjoint concentrates at low frequencies for Mean constraints and spreads progressively for ΔMean and Max constraints, empirically supporting the constraint-inheritance analysis in Section 4.1.

These results support the anytime property: each K yields a valid sample from the controlled diffusion, and additional iterations move the control toward a stationary point of J_V within the parameterized subspace (Proposition 4.2).

C.3 Additional Experiments

C.3.1 Linear-Gaussian Validation

The experiments in Section 5 measure constraint violation and distributional quality, but not distance to the tilted posterior itself, since $p^\star_\lambda$ is unavailable in closed form on real data. We therefore construct a setting in which it is available exactly.

Setup. The prior is a Gaussian process with a Matérn-$3/2$ kernel (lengthscale 4, variance 0.09) on $d = 12$ unit-spaced points, so the data law is $x_0 \sim \mathcal{N}(0, \Sigma_0)$. The condition is a linear constraint on the temporal mean, $A = \frac{1}{d}\mathbf{1}^\top$ with target $b = 0.5$. Under this choice the denoiser, the tilted posterior $p^\star_\lambda$, the Doob h-function and the optimal control $u^\star$ are all available in closed form, and the $\lambda \to \infty$ limit is the hard conditional. We run our samplers and constraint code without modification, replacing only the denoiser by its closed form, so that any gap is attributable to the sampler rather than to a learned score. Tables 11–12 use $\lambda = 2$, and the reported $\mathrm{KL}(\cdot \| p^\star_\lambda)$ is the exact Gaussian KL between the empirical law $\mathcal{N}(\hat{\mu}, \hat{\Sigma})$ of $B = 8000$ draws ($T = 500$ diffusion steps) and the closed-form $p^\star_\lambda$.

Table 10: Performance under different iteration K. The best results and the second best for each group are in **bold** and underlined, respectively. DownArrow (↓) indicates lower is better.

		Violation ↓			Context-FID ↓		
Data	Method	Mean	ΔMean	Max	Mean	ΔMean	Max
Stock	TTC-FFT (1)	**0.000**	**0.000**	**0.000**	0.192	283.406	29.460
	TTC-FFT (5)	**0.000**	**0.000**	**0.000**	0.214	273.698	16.638
	TTC-FFT (10)	**0.000**	**0.000**	**0.000**	**0.177**	**209.189**	**16.235**
	TTC-WAV (1)	**0.000**	**0.000**	**0.000**	0.216	363.999	28.367
	TTC-WAV (5)	**0.000**	**0.000**	**0.000**	0.190	279.141	20.071
	TTC-WAV (10)	**0.000**	**0.000**	**0.000**	**0.185**	**238.360**	**16.639**
ETTh	TTC-FFT (1)	**0.000**	**0.000**	**0.000**	9.184	11.791	6.196
	TTC-FFT (5)	**0.000**	**0.000**	**0.000**	8.661	12.294	5.903
	TTC-FFT (10)	**0.000**	**0.000**	**0.000**	**7.835**	**11.761**	**5.092**
	TTC-WAV (1)	**0.000**	**0.000**	**0.000**	9.330	12.620	7.084
	TTC-WAV (5)	**0.000**	**0.000**	**0.000**	8.239	13.784	6.891
	TTC-WAV (10)	**0.000**	**0.000**	**0.000**	**7.501**	**11.593**	**6.832**
Energy	TTC-FFT (1)	**0.000**	**0.000**	**0.000**	2.194	**17.079**	9.750
	TTC-FFT (5)	**0.000**	**0.000**	**0.000**	2.015	18.168	8.853
	TTC-FFT (10)	**0.000**	**0.000**	**0.000**	**1.783**	17.130	**6.681**
	TTC-WAV (1)	**0.000**	**0.000**	**0.000**	2.312	**18.180**	10.418
	TTC-WAV (5)	**0.000**	**0.000**	**0.000**	1.924	20.065	10.608
	TTC-WAV (10)	**0.000**	**0.000**	**0.000**	**1.756**	18.771	**9.975**
fMRI	TTC-FFT (1)	**0.000**	**0.000**	**0.000**	19.172	19.440	3.641
	TTC-FFT (5)	**0.000**	**0.000**	**0.000**	19.305	18.809	**3.121**
	TTC-FFT (10)	**0.000**	**0.000**	**0.000**	**18.714**	**18.692**	3.336
	TTC-WAV (1)	**0.000**	**0.000**	**0.000**	18.918	18.788	4.716
	TTC-WAV (5)	**0.000**	**0.000**	**0.000**	**18.601**	18.680	3.844
	TTC-WAV (10)	**0.000**	**0.000**	**0.000**	19.252	**18.446**	**3.764**

For each baseline we sweep the guidance weight over $\{0.5, 1, 2, 4, 8\} \cdot B$ and report the lowest-KL run; Diffusion-TS selects the smallest weight ($\lambda_{\text{eff}} = 0.5$).

Table 11: Fidelity to the tilted posterior $p^\star_\lambda$ at $K = 1$. Lower KL is better. The best KL is in **bold**.

Method	$\mathrm{KL}(\cdot \| p^\star_\lambda) \downarrow$	Violation
Base (uncontrolled)	0.170	0.300
Diffusion-TS	0.226	0.327
COP	6.56	~ 0
CPS	5.73	~ 0
TTC-Direct	**0.081**	0.216
TTC-FFT	0.086	0.216
TTC-WAV	0.097	0.210

Posterior correctness. Only our method attains a KL below that of the uncontrolled prior, roughly halving it in a single adjoint pass (0.081 against 0.170). Diffusion-TS, even at its best swept weight, ends further from the target than no control (KL $0.226 > 0.170$) while also leaving more violation ($0.327 > 0.300$), worse on both axes than the unconstrained prior. This is consistent with Theorem 3.3: guidance substitutes a one-step estimate for the Doob h-function, and the resulting bias is not removed by tuning the weight. The projection methods target the $\lambda \to \infty$ conditional rather than the soft posterior, so their distance from $p^\star_\lambda$ is expected rather than a defect, and quantifies the trade-off of Proposition 3.2.

Role of the iteration count. Violation falls monotonically with K while the KL to $p^\star_\lambda$ rises, so K traverses the same soft-to-hard axis that λ parameterizes in Proposition 3.2. Each iteration raises the effective tilt, moving samples toward the feasible set at an increasing cost in distance from the

Table 12: The iteration count K as a soft-to-hard knob. KL to the soft target rises as violation falls.

K	$\mathrm{KL}(\cdot\|p^\star_\lambda)$	Violation
1	0.081	0.216
2	0.197	0.155
3	0.396	0.118
5	0.824	0.082
10	1.622	0.059

finite-λ posterior. This parallels the role of the terminal projection: once the iterations bring samples near the feasible set, a final tightening step has little left to do.

Table 13: Relative distance of the realized control to $u^\star$ under injected adjoint noise. The best results are in **bold**. Fourier and wavelet coincide, since the retained subspaces agree in this setting.

Adjoint noise	TTC-Direct	TTC-FFT = TTC-WAV
0	0.605	0.605
1	0.798	**0.694**
2	1.207	**0.917**
3	1.681	**1.200**

Direct control carries the injected noise into the update, whereas the parameterized variants remove its component outside $\mathrm{Range}(\mathcal{R})$ and stay up to 29% closer to $u^\star$. Since all variants coincide at zero noise, the parameterization discards variance rather than signal. That Fourier and wavelet behave identically indicates that the filtering comes from the orthogonal projection itself rather than from the choice of basis, consistent with the spectral analysis of Appendix C.2.3.

C.3.2 More Constraints

The three constraints used in Section 5 cover complementary structure: Mean fixes a global statistic, ΔMean a trend, and Max a value at a specific timestamp. To assess whether the method extends beyond these, we evaluate three further constraint types.

Autocorrelation. The constraint fixes the lag-k autocorrelation of each generated trajectory, targeting periodic structure rather than a pointwise statistic. Its gradient is spread across the spectrum, making it the least aligned with the retained subspace among the constraints we consider.

Mask. The constraint fixes the series at a prescribed subset of timesteps, which is the partial-observation setting mentioned in the introduction.

OHLC. On Stock, the constraint enforces structural consistency between the open, high, low and close channels: at every timestep, the high dominates and the low is dominated by the remaining channels.

Table 14: Violation under three additional constraint types. Lower is better. The best results are in **bold**.

	Autocorrelation				Mask				OHLC
Method	Stock	ETTh	Energy	fMRI	Stock	ETTh	Energy	fMRI	Stock
GuidedDiffTime	1.623	2.140	6.927	10.284	132.947	50.535	327.345	327.953	0.034
Diffusion-TS	1.661	2.137	6.924	10.294	**0.000**	**0.000**	**0.000**	**0.000**	0.032
COP	1.639	2.135	6.935	10.329	0.270	0.478	1.417	4.295	0.035
COP-FT	1.653	2.133	6.933	10.316	0.269	0.478	1.417	4.295	0.034
CPS	0.102	0.062	0.112	0.089	**0.000**	**0.000**	**0.000**	**0.000**	0.018
TTC-FFT	**0.000**	**0.000**	**0.017**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.015**
TTC-WAV	**0.000**	**0.000**	**0.017**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	0.016

The autocorrelation constraint separates the methods most sharply. The guidance and terminal-projection baselines leave violations at essentially the uncontrolled level, and even CPS retains a

residual, whereas TTC reaches zero on three of four datasets and 0.017 on Energy. This is the regime Theorem 3.3 speaks to: the constraint gradient is spectrally spread, so a local surrogate has the least to work with, and trajectory-level control has the most to contribute.

On the mask constraint, Diffusion-TS and CPS also reach zero, but for Diffusion-TS this holds by construction, since it overwrites the observed entries directly. That mechanism applies only to constraints of the form "these coordinates equal these values," and the same method remains at the uncontrolled level on the autocorrelation constraint. On OHLC all methods perform reasonably and TTC is best by a smaller margin, consistent with the constraint being a per-timestep inequality that any locally corrective scheme can partially enforce.

C.3.3 Distributional Constraints

The formulation of Section 3 states the condition through a terminal loss evaluated on a single sample. This restriction, however, is one of exposition rather than of the formulation. The entropic objective is defined over path measures, the only per-sample object is the terminal loss itself, and the method requires nothing of it beyond differentiability. A condition defined on a batch is therefore equally admissible, with the adjoint initialized from the gradient of that batch functional instead.

To verify this, we replace the per-sample condition with a distributional one: rather than fixing the temporal mean of each sample, the target is the *distribution* of the temporal mean over a reference set. We compare two terminal costs: a pointwise cost, which penalizes the deviation of each sample from the target value, and an optimal-transport cost between the generated and reference distributions of the statistic.

Table 15: Distributional constraint results. Point denotes a pointwise terminal cost and Transport an optimal-transport terminal cost. For the standard-deviation ratio to the reference, closer to 1 is better; for energy distance and Context-FID, lower is better. The best results are in **bold**.

Dataset	Terminal cost	Std ratio to ref.	Energy dist. ↓	Context-FID ↓
Stock	Point	0.02	1.64	2.56
	Transport	**1.04**	**0.11**	**0.16**
Energy	Point	0.01	0.72	5.23
	Transport	**0.93**	**0.15**	**1.06**

The contrast is instructive. With a pointwise cost, the sampler drives every sample to the target value and the distribution collapses, with its standard deviation falling to 1–2% of the reference. The constraint is then satisfied in the per-sample sense, while the distributional condition it was meant to express is violated entirely. Replacing it with a transport cost recovers the reference spread, at ratios of 1.04 and 0.93, while also lowering both the energy distance and Context-FID.

A distributional condition is therefore within scope, provided the terminal cost is defined on the batch rather than on each sample. The sampler itself requires no modification, and the choice of terminal cost determines whether the intended condition is the one actually enforced.

C.3.4 Downstream Utility: Data Augmentation

The experiments above measure constraint satisfaction and distributional quality, but not whether satisfying a constraint makes generated data more useful. We test this in a data-scarce setting on Energy under the autocorrelation constraint of Appendix C.3.2, where TTC reaches near-zero violation, CPS retains a residual, and Diffusion-TS stays at the uncontrolled level (Table 14). We augment 40 real windows with generated windows, train a ridge forecaster from the past to the future of each window, and evaluate it on held-out real windows over 20 random splits.

Generations that satisfy the constraint improve the forecaster by roughly 60% more than those that violate it ($+22$–23% against $+14$–15%). Unconstrained samples, which are realistic but ignore the constraint, fall with the violating methods, so the gain comes from constraint satisfaction rather than from sample realism alone. Constrained samples can thus substitute for scarce real data, and TTC is the method that satisfies the autocorrelation constraint most completely.

Table 16: Forecaster trained on 40 real windows augmented with generated windows, evaluated on held-out real data (Energy, autocorrelation constraint, 20 splits, std ≈ 0.003). Lower MSE is better.

Training data	MSE ↓	Gain over real only
Real only	0.083	–
+ Unconstrained	0.071	+15%
+ Diffusion-TS	0.072	+14%
+ CPS	0.065	+22%
+ TTC	**0.064**	**+23%**

C.3.5 Longer Horizon

To confirm that the method continues to work at longer horizons, we repeat the experiments at sequence lengths L=96 and L=192 (2× and 4× the benchmark length). We train a Diffusion-TS backbone at each length on Stock, ETTh and Energy, keep it frozen, and repeat the comparison on the Mean, ΔMean and Max constraints. We exclude fMRI, whose series are too short to form windows at these lengths. We evaluate TTC-FFT and TTC-WAV against the guidance baselines (GuidedDiffTime, Diffusion-TS) and the projection baseline COP, reporting violation and Context-FID averaged over 5 samples per window. We omit CPS, which re-solves the projection at every denoising step and already runs an order of magnitude slower than COP at L=48, making it impractical at these lengths. All other settings match the L=48 experiments, and results are in Table 17.

TTC-FFT and TTC-WAV reach exactly zero violation on every constraint at both L=96 and L=192, as at L=48. The guidance baselines stay far from feasible on Mean and Max (violations of 5–7 on Stock, 1.4–6.5 on ETTh and 11–25 on Energy) at both horizons. COP also reaches near-zero violation, so the meaningful comparison among feasible methods is on distributional quality.

Among the feasible methods, TTC attains the best Context-FID on Mean and Max in all three datasets at L=96. Both parameterizations retain the DC component that a Mean gradient concentrates on, and the Fourier and wavelet variants split the Max cells, so the advantage comes from trajectory-level control rather than from one basis. ΔMean is the constraint where the parameterization helps least, since its gradient is broadband, and the baselines are correspondingly competitive there, as at L=48. Absolute Context-FID grows with L for every method, since longer sequences are harder to match. At L=192 the ordering holds on all but two cells, ETTh-Mean and Energy-Max, where COP edges TTC on Context-FID (20.6 vs 20.9 and 23.6 vs 30.3) while still leaving a residual violation (0.03 and 0.10) that TTC drives to exactly zero; on Energy-Max, COP also pays roughly 110× TTC's runtime for it (see below). The low Context-FID of the guidance baselines on Energy-Max (e.g., Diffusion-TS 14.8 and 20.1) is not a fair comparison, as those samples are infeasible (violation 23–25).

A final difference lies in runtime and, more importantly, in how it scales. TTC runs a fixed number of adjoint iterations per window, so its per-batch cost is roughly constant across constraints and grows about linearly with the horizon, roughly doubling each time the length doubles. COP is a single projection and is therefore several times faster than TTC on the affine Mean and ΔMean constraints and on the lower-dimensional datasets. Its cost, however, depends sharply on both the constraint and the dimension: the Max projection grows superlinearly with the horizon (roughly 5–6× per doubling on Stock and ETTh), and on the higher-dimensional Energy data (28 channels) COP becomes slower than TTC on Max, with the multiple growing from ∼29× at L=96 to ∼110× at L=192.

In short, across a 4× range of horizons TTC keeps exactly zero violation and a near-constant quality ordering while its runtime grows only linearly, whereas the projection baseline's cost grows superlinearly and, where it draws level with TTC on quality, still falls short of exact feasibility.

C.3.6 Extension to Soft Conditions: Forecasting

While the main paper focuses on hard structural constraints, the same framework applies when the condition is an input rather than the objective. For structural constraints, meeting the specification is the task itself: a sample that misses it has failed, the feasible set is well defined, and satisfaction is checked against a tolerance. In forecasting, the condition is instead a mask of observed history from which the remaining steps are generated; reproducing those values is not the deliverable, and there is no feasible set to project onto. The same terminal loss and tilted target are used in both cases,

Table 17: Long-horizon results at sequence lengths 96 and 192. The best results are in **bold**.

		Horizon 96						Horizon 192					
		Violation ↓			Context-FID ↓			Violation ↓			Context-FID ↓		
Data	Method	Mean	ΔMean	Max	Mean	ΔMean	Max	Mean	ΔMean	Max	Mean	ΔMean	Max
Stock	GuidedDiffTime	5.425	**0.000**	6.628	619.971	644.130	831.713	5.169	**0.000**	6.730	1473.250	2222.471	1704.870
	Diffusion-TS	6.897	**0.000**	6.660	568.584	1513.130	350.542	6.693	**0.000**	5.912	3259.265	2455.481	1174.078
	COP	0.021	**0.000**	0.019	0.565	895.021	525.043	0.021	**0.000**	0.018	1.538	1490.636	1845.865
	TTC-FFT	**0.000**	**0.000**	**0.000**	0.648	**518.992**	357.747	**0.000**	**0.000**	**0.000**	**0.726**	2146.694	831.764
	TTC-WAV	**0.000**	**0.000**	**0.000**	**0.412**	713.044	**350.268**	**0.000**	**0.000**	**0.000**	0.925	**1475.874**	**558.112**
ETTh	GuidedDiffTime	1.479	0.001	3.766	22.840	**19.337**	18.190	1.445	**0.000**	4.032	26.822	**22.843**	25.513
	Diffusion-TS	1.563	0.001	5.567	21.472	20.337	12.186	1.523	**0.000**	6.468	29.537	27.171	20.303
	COP	0.030	0.001	0.018	14.760	20.162	13.212	0.031	**0.000**	0.017	**20.644**	26.666	21.720
	TTC-FFT	**0.000**	**0.000**	**0.000**	14.960	20.828	**8.589**	**0.000**	**0.000**	**0.000**	20.960	29.287	**10.772**
	TTC-WAV	**0.000**	**0.000**	**0.000**	**14.518**	22.776	8.806	**0.000**	**0.000**	**0.000**	21.637	28.089	14.737
Energy	GuidedDiffTime	11.486	0.029	20.642	27.805	25.695	25.697	10.997	0.003	22.683	35.320	**30.910**	32.515
	Diffusion-TS	15.580	0.029	23.279	59.979	24.743	**14.847**	15.250	0.003	24.739	87.191	34.587	**20.089**
	COP	0.101	0.010	0.104	8.107	25.087	19.470	0.103	0.001	0.095	12.701	38.478	23.564
	TTC-FFT	**0.000**	**0.000**	**0.000**	7.587	26.897	16.552	**0.000**	**0.000**	**0.000**	12.528	36.184	30.344
	TTC-WAV	**0.000**	**0.000**	**0.000**	**7.450**	**24.292**	15.753	**0.000**	**0.000**	**0.000**	**12.477**	31.502	30.912

but L here supplies information to generate from rather than a target to be attained, so no tolerance and no projection appear in this setting. As a preliminary demonstration, we condition on historical values of length h and predict the remaining $48 - h$ steps, evaluating with mean squared error (MSE), continuous ranked probability score (CRPS), and average quantile loss (QL) at $\tau \in \{0.1, 0.9\}$. We use the single-iteration variants so that the sampling cost matches that of the baselines, and set the control weight to $\lambda = 10^4$.

As shown in Table 18, TTC-FFT$_s$ attains the best average rank on all three metrics (1.75, 1.58 and 1.42), with TTC-WAV$_s$ second on QL and tied for second on MSE, suggesting that the framework extends beyond hard constraint satisfaction to soft conditional generation. The advantage is not uniform across datasets, and its structure is informative. On Stock and fMRI, the single-iteration variants are best or tied for best on every metric and every horizon, reducing MSE relative to GuidedDiffTime by roughly 86% on Stock and improving on Diffusion-TS by a further 5–7%. On ETTh and Energy, by contrast, Diffusion-TS remains stronger on MSE and CRPS, by 59–75% on ETTh MSE and 14–20% on Energy MSE. The ordering reverses on QL, where TTC-FFT$_s$ is best on Stock, ETTh and Energy at every horizon and beats Diffusion-TS in 9 of the 12 settings, improving quantile loss by up to 24% on Energy. Since QL at $\tau \in \{0.1, 0.9\}$ probes the tails of the predictive distribution while MSE and CRPS are dominated by its center, this pattern suggests that trajectory-level control improves calibration in the tails even where it does not improve central accuracy, as one would expect from a method that reshapes the whole sample path rather than the conditional mean.

Table 18: Forecasting results in mean squared error (MSE), continuous ranked probability score (CRPS) and quantile loss (QL), with their average rank. The best results are in **bold** and the second best are underlined. DownArrow (↓) indicates lower is better.

		Stock			ETTh			Energy			fMRI			AvgRank ↓
Metric	Method	12	24	36	12	24	36	12	24	36	12	24	36	(Rank)
MSE ↓	GuidedDiffTime	1.113	1.115	1.093	0.165	0.164	0.165	0.393	0.394	0.392	0.120	0.119	0.118	4.00 (4)
	Diffusion-TS	0.160	0.174	0.166	**0.048**	**0.051**	**0.061**	**0.132**	**0.131**	**0.132**	0.111	0.115	0.117	2.00 (2)
	TTC-FFT$_s$	**0.151**	**0.162**	**0.157**	0.084	0.081	0.098	0.150	0.153	0.158	**0.108**	**0.109**	0.114	**1.75 (1)**
	TTC-WAV$_s$	**0.151**	0.164	0.158	0.086	0.084	0.101	0.150	0.152	0.154	0.110	**0.109**	**0.111**	2.00 (2)
CRPS ↓	GuidedDiffTime	0.933	0.934	0.923	0.292	0.291	0.292	0.477	0.477	0.477	0.264	0.263	0.263	4.00 (4)
	Diffusion-TS	0.310	0.319	0.312	**0.153**	**0.155**	**0.167**	**0.252**	**0.252**	**0.254**	0.254	0.258	0.261	2.00 (2)
	TTC-FFT$_s$	**0.301**	**0.314**	**0.306**	0.209	0.206	0.223	0.273	0.277	0.282	**0.251**	**0.252**	0.258	**1.58 (1)**
	TTC-WAV$_s$	0.302	0.316	0.307	0.210	0.209	0.226	0.282	0.284	0.287	0.253	0.253	**0.255**	2.42 (3)
QL ↓	GuidedDiffTime	0.400	0.403	0.396	0.070	0.070	0.070	0.139	0.138	0.138	0.060	**0.060**	0.060	3.50 (4)
	Diffusion-TS	0.154	0.158	0.154	0.057	0.056	0.059	0.107	0.104	0.099	**0.059**	**0.060**	0.060	2.50 (3)
	TTC-FFT$_s$	**0.147**	**0.154**	**0.150**	**0.049**	**0.049**	**0.058**	**0.081**	**0.082**	**0.083**	0.060	**0.060**	0.063	**1.42 (1)**
	TTC-WAV$_s$	0.148	0.155	**0.150**	**0.049**	0.050	0.060	0.086	0.089	0.091	**0.059**	**0.060**	**0.059**	1.67 (2)

C.3.7 Additional Quantitative Results with Standard Deviations

Table 19 reports the results in terms of mean and standard deviation over multiple runs. Our methods generally exhibit lower standard deviations than baselines, particularly on Stock and Energy where guidance-based methods show high variability.

Table 19: Quality comparison on statistical constraints using Context-FID and Correlation (with standard deviations). DownArrow (↓) indicates lower is better.

		Context-FID ↓			Correlation ↓		
Dataset	Method	Mean	ΔMean	Max	Mean	ΔMean	Max
Stock	GuidedDiffTime	465.188±226.009	367.728±100.287	295.443±31.108	0.262±0.011	0.371±0.003	0.290±0.004
	Diffusion-TS	773.924±352.743	470.734±270.456	389.098±58.691	0.077±0.009	0.279±0.009	0.292±0.008
	TTC-FFT$_s$ (No proj)	9.711±1.265	219.968±35.879	43.942±12.268	0.190±0.003	0.271±0.008	0.279± 0.009
	TTC-WAV$_s$ (No proj)	10.981±1.441	363.200±33.216	48.738±7.668	0.187±0.001	0.265±0.009	0.394±0.013
	TTC-FFT (No proj)	1.854± 0.436	252.430±12.068	69.799±19.512	0.221±0.002	0.269±0.007	0.120± 0.009
	TTC-WAV (No proj)	1.720±0.702	194.939±92.987	238.138±26.671	0.230±0.006	0.286±0.013	0.062±0.009
	GuidedDiffTime + proj	608.311±365.886	379.406±124.265	538.001±156.659	0.282±0.008	0.278±0.011	0.402±0.006
	Diffusion-TS + proj	13.717±3.486	656.096±256.985	276.481±89.792	0.133±0.004	0.298±0.005	0.431±0.009
	COP	0.149±0.028	416.229±139.359	483.345±68.356	0.017±0.007	0.293±0.007	0.508±0.026
	COP-FT	0.183±0.050	316.462±49.257	261.026±90.783	0.015±0.006	0.304±0.010	0.501±0.007
	CPS	0.379±0.048	428.990±330.297	24.428±4.424	0.230±0.006	0.266±0.011	0.648±0.006
	TTC-FFT$_s$	0.192±0.042	283.406±45.508	29.460±11.396	0.020±0.005	0.276±0.010	0.326±0.036
	TTC-WAV$_s$	0.216±0.047	363.999±124.396	28.367±6.677	0.039±0.030	0.278±0.005	0.428±0.013
	TTC-Direct	0.201±0.018	321.151±113.317	11.367±1.179	0.035±0.005	0.282±0.014	0.479±0.020
	TTC-FFT	0.177±0.022	209.189±24.776	16.235±1.692	0.034±0.011	0.248±0.005	0.472±0.007
	TTC-WAV	0.185±0.011	238.360±74.036	16.639±0.718	0.029±0.009	0.279±0.013	0.590±0.006
ETTh	GuidedDiffTime	12.041±0.447	12.592±2.280	10.811±1.420	0.317±0.006	0.330±0.007	0.319±0.007
	Diffusion-TS	13.550±1.929	14.372±1.644	6.609±0.860	0.340±0.009	0.315±0.008	0.383±0.004
	TTC-FFT$_s$ (No proj)	10.679±0.668	12.316±0.912	10.612±0.658	0.299±0.003	0.324±0.006	0.484± 0.007
	TTC-WAV$_s$ (No proj)	10.103±0.998	12.887±1.026	9.029±0.775	0.288±0.007	0.324±0.006	0.442±0.007
	TTC-FFT (No proj)	7.458±0.121	11.089±0.839	8.610±1.040	0.128±0.013	0.312±0.003	0.273±0.015
	TTC-WAV (No proj)	7.664±0.311	11.756±0.116	8.769±0.461	0.110±0.015	0.318±0.008	0.239±0.004
	GuidedDiffTime + proj	12.497±2.336	12.216±3.498	11.378±2.665	0.319±0.009	0.330±0.011	0.364±0.010
	Diffusion-TS + proj	9.538±1.684	12.498±2.666	12.571±2.198	0.199±0.003	0.327±0.007	0.392±0.011
	COP	9.267±0.140	12.454±0.778	8.307±0.378	0.199±0.016	0.341±0.005	0.363±0.005
	COP-FT	9.597±0.598	14.878±0.598	8.346±0.981	0.212±0.026	0.333±0.005	0.362±0.002
	CPS	5.271±0.244	13.143±0.883	7.094±0.501	0.189±0.012	0.313±0.002	0.242±0.005
	TTC-FFT$_s$	9.184±0.281	11.791±1.742	6.196±0.219	0.167±0.008	0.326±0.004	0.398±0.003
	TTC-WAV$_s$	9.330±0.559	12.620±1.161	7.084±0.030	0.171±0.021	0.316±0.002	0.375±0.003
	TTC-Direct	7.743±0.372	13.454±1.028	7.078±0.196	0.189±0.005	0.326±0.002	0.437±0.010
	TTC-FFT	7.835±0.221	11.761±1.439	5.092±0.828	0.167±0.012	0.310±0.010	0.429±0.011
	TTC-WAV	7.501±0.305	11.593±0.658	6.832±0.157	0.168±0.004	0.313±0.009	0.391±0.007
Energy	GuidedDiffTime	17.974±0.865	17.604±1.369	18.513±1.699	8.024±0.048	7.923±0.042	7.713±0.047
	Diffusion-TS	46.035±14.534	20.226±4.653	15.403±2.264	8.274±0.216	7.994±0.278	8.552±0.200
	TTC-FFT$_s$ (No proj)	6.737±0.225	20.882±1.006	15.135±0.998	7.669±0.186	7.978±0.254	6.912±0.132
	TTC-WAV$_s$ (No proj)	6.895±0.228	16.368±1.233	13.142±1.664	7.729±0.198	8.109±0.166	7.205±0.199
	TTC-FFT (No proj)	6.806±0.319	18.048±0.872	15.609±1.128	5.846±0.147	7.939±0.037	7.424±0.221
	TTC-WAV (No proj)	6.055±0.199	20.852±2.040	15.002±1.364	5.717±0.119	7.997±0.190	6.852±0.108
	GuidedDiffTime + proj	19.164±1.866	21.674±7.659	18.676±3.664	8.076±0.066	8.226±0.126	8.221±0.245
	Diffusion-TS + proj	5.312±1.785	19.273±4.659	11.427±3.125	2.911±0.218	7.813±0.269	8.472±0.247
	COP	2.927±0.274	21.515±5.052	12.942±0.567	1.438±0.068	7.967±0.106	8.146±0.124
	COP-FT	3.050±0.033	17.943±1.525	16.092±0.960	1.409±0.084	7.744±0.090	8.151±0.120
	CPS	2.353±0.117	20.290±2.623	4.863±0.871	1.436±0.044	7.743±0.275	5.093±0.075
	TTC-FFT$_s$	2.194±0.046	17.079±2.552	9.750±0.608	1.025±0.121	8.006±0.078	7.084±0.038
	TTC-WAV$_s$	2.312±0.084	18.180±3.126	10.418±0.522	1.099±0.098	7.870±0.067	7.484±0.096
	TTC-Direct	1.905±0.055	18.584±5.027	8.470±1.589	1.272±0.015	8.027±0.088	6.655±0.066
	TTC-FFT	1.783±0.127	17.130±1.761	6.681±0.625	1.307±0.060	7.910±0.031	5.894±0.036
	TTC-WAV	1.756±0.067	18.771±2.531	9.975±0.527	1.223±0.024	7.641±0.184	6.318±0.132
fMRI	GuidedDiffTime	19.181±0.231	19.235±0.627	14.212±0.346	3.455±0.025	3.419±0.022	3.404±0.022
	Diffusion-TS	18.716±0.126	19.361±0.160	18.140±0.210	3.361±0.048	3.422±0.023	3.181±0.019
	TTC-FFT$_s$ (No proj)	19.496±0.182	18.854±0.126	2.386±0.015	3.432±0.044	3.413±0.019	4.739±0.022
	TTC-WAV$_s$ (No proj)	19.068±0.184	19.667±0.124	2.814±0.023	3.361±0.042	3.417±0.015	4.574±0.028
	TTC-FFT (No proj)	18.836±0.184	19.590±0.428	2.472±0.051	3.415±0.041	3.401±0.011	4.561±0.023
	TTC-WAV (No proj)	19.172±0.174	19.251±0.576	2.643±0.072	3.405±0.050	3.414±0.035	4.592±0.034
	GuidedDiffTime + proj	20.429±0.668	20.217±0.698	15.567±0.532	3.395±0.054	3.417±0.026	3.898±0.028
	Diffusion-TS + proj	20.573±0.532	19.864±0.429	13.862±0.433	3.350±0.042	3.359±0.038	4.056±0.025
	COP	19.425±0.301	18.930±0.487	14.606±0.471	3.379±0.039	3.395±0.033	3.935±0.022
	COP-FT	19.425±0.316	19.116±0.227	14.528±0.302	3.362±0.005	3.435±0.008	3.977±0.042
	CPS	21.547±0.400	16.650±0.642	11.582±0.131	3.927±0.013	3.173±0.027	3.834±0.026
	TTC-FFT$_s$	19.172±0.110	19.440±0.238	3.641±0.172	3.356±0.027	3.325±0.020	4.970±0.049
	TTC-WAV$_s$	18.918±0.426	18.788±0.507	4.716±0.104	3.376±0.014	3.317±0.022	4.819±0.079
	TTC-Direct	19.291±0.192	19.334±0.274	6.087±0.152	3.400±0.028	3.270±0.029	6.028±0.089
	TTC-FFT	18.714±0.344	18.692±0.599	3.336±0.219	3.310±0.012	3.340±0.028	4.749±0.003
	TTC-WAV	19.252±0.343	18.446±0.213	3.764±0.148	3.341±0.026	3.312±0.038	4.621±0.020

C.4 VISUALIZATION

We visualize constraint generation results on the four benchmark datasets (Stock, ETTh, Energy, fMRI) for three constraint types (Mean, ΔMean, Max) in Figure 5. In each panel, the blue curve is the ground-truth trajectory, the red curve is the generated trajectory, and the orange element encodes the constraint target.



Figure 5: Visualization of time series generation situations under various constraints.