SLOWLY ANNEALED JUMP PROCESSES: THEORY AND APPLICATIONS TO MDM FINE-TUNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Annealing is a powerful strategy for sampling from complex distributions and reward-based fine-tuning of diffusion models, using a sequence of intermediate distributions to connect an easily sampled reference distribution to the desired target. However, if the target changes faster than the sampling or learning dynamics can adapt, substantial tracking error can persist. This motivates slow annealing, which gives the dynamics more time to adapt to the evolving target and recent work has established the effectiveness in both continuous-state sampling and guided generation. However, whether slow annealing yields comparable benefits for discrete jump processes and reward-based fine-tuning of masked diffusion models remains insufficiently understood. In this work, we develop a theory of slowly-annealed jump processes with forward KL tracking guarantees of slowly-annealed Glauber dynamics. We further derive conditional KL bounds for slowly-annealed masked diffusion fine-tuning under idealized learning dynamics. Experiments on tractable distributions and LLaDA-8B-Instruct demonstrate the benefits.

1 INTRODUCTION

Sampling from complex, high-dimensional distributions is a central challenge in probabilistic modeling. Multimodality can make it difficult for sampling algorithms to explore the target distribution effectively. Annealing provides a significant benefit by gradually transforming the simple prior distribution to complex target distribution by a sequence of intermediate distributions. This strategy has been shown to be effective in statiscal physics (Machta, 2010; Ying, 2023; Prat Pou et al., 2025), generative modeling (Sohl-Dickstein et al., 2015; Song & Ermon, 2019; Doucet et al., 2022; Wu & Xie, 2025), and reward-based fine-tuning of diffusion models (Chen et al., 2026)

A key factor in this tracking problem is the annealing speed. If the target evolves too quickly, the dynamics may have insufficient time to adapt, leaving substantial tracking error. Slow annealing addresses this mismatch by allowing more algorithmic time to traverse the same target path. Guo et al. (2025) quantify the dependence of sampling accuracy on annealing time for non-log-concave targets, providing theoretical support for slowing the target evolution. Building on this perspective, Nitanda et al. (2026) establish forward-KL tracking guarantees for slowly annealed Langevin dynamics and demonstrate benefits in continuous-state sampling and guided generation.

These scheduling questions also arise in discrete applications. In statistical physics, local spin updates can struggle to cross energy barriers, making equilibration sensitive to the annealing schedule. Recently, He et al. (2026) established reverse-KL guarantees for discrete simulated annealing, with applications to mean-field spin models. In reward-based fine-tuning of masked diffusion models, increasing the reward tilt similarly requires the model to adapt to shifting target distributions within a limited training budget (Chen et al., 2026; So et al., 2026). Although these settings employ different annealing schemes, both face the challenge of adapting to evolving target distributions. This motivates the following question: *Can we develop a common theoretical framework to quantify the benefits of slow annealing in discrete sampling and masked diffusion fine-tuning?* We address this question through KL differential inequalities that separate error dissipation from the effect of target motion. This approach yields forward-KL guarantees for annealed jump processes and conditional KL tracking bounds for masked diffusion fine-tuning under idealized learning dynamics. Our contributions are as follows:

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.

1. We establish forward-KL tracking guarantees for slowly annealed jump processes under modified log-Sobolev inequalities (Theorem 1). The bounds separate the decay of initialization error from the cost of target motion. We apply this analysis to annealed Glauber dynamics, obtaining forward-KL sampling guarantees for mean-field Ising and Potts models (Theorem 2 and Theorem 3).
2. We derive KL tracking bounds for slowly annealed Discrete Tilt Matching and Discrete Adjoint Matching under idealized learning dynamics, quantifying how annealing speed affects tracking accuracy using Lyapunov function with KL-divergence (Theorem 4 and Theorem 5).
3. We evaluate slowly annealed fine-tuning on tractable distributions and LLaDA-8B-Instruct. The experiments demonstrate that finer annealing schedules can improve distributional accuracy and task performance, while revealing a trade-off between the number of stages and the optimization budget per stage (Section 5).

2 PRELIMINARIES

We consider continuous-time Markov jump processes on a finite connected undirected graph (Ω, E). Let $(\pi_s)_{s\in[0,1]}$ be a C^1 path in $\mathcal{P}_{>0}(\Omega)$, the set of strictly positive probability distributions on Ω. For each s, let $p_s(x,y)$ be jump rates that are continuous in s, positive on edges, and zero otherwise. The corresponding generator is $L_s h(x) = \sum_{y:\{x,y\}\in E} p_s(x,y)\big(h(y) - h(x)\big)$. We assume that L_s is reversible with respect to π_s, so the edge capacities $c_s(x,y) := \pi_s(x)p_s(x,y)$ satisfy $c_s(x,y) = c_s(y,x)$. In particular, π_s is invariant under L_s.

Given a terminal time $T > 0$, let $(X_t)_{t\in[0,T]}$ be the Markov jump process with time-dependent generator $L_{t/T}$ and initial distribution ρ_0. Its law $\rho_t := \mathcal{L}(X_t)$ satisfies $\partial_t \rho_t = L^*_{t/T}\rho_t$, where L^*_s denotes the adjoint with respect to counting measure. Here, t is the running time and $s = t/T$ is the normalized path parameter. Increasing T slows the traversal of the prescribed target path while preserving the generator at each point along it.

For probability distributions ρ and π with $\pi > 0$, write $\mathrm{KL}(\rho\|\pi) := \sum_{x\in\Omega} \rho(x)\log(\rho(x)/\pi(x))$, with $0\log 0 = 0$. The Dirichlet form associated with L_s is

$$\mathcal{E}_s(h,k) := \frac{1}{2}\sum_{x\in\Omega}\sum_{y:\{x,y\}\in E} c_s(x,y)\big(h(y)-h(x)\big)\big(k(y)-k(x)\big).$$

For $\rho > 0$ and $f = \rho/\pi_s$, define the entropy dissipation by $\mathcal{I}_s(\rho\|\pi_s) := \mathcal{E}_s(f, \log f)$. It is nonnegative and extends to distributions with zero entries by lower semicontinuity.

We quantify entropy contraction through the modified log-Sobolev inequality (mLSI) (Bobkov & Tetali, 2006). Whereas the classical log-Sobolev inequality controls entropy through $\mathcal{E}_s(\sqrt{f}, \sqrt{f})$, the mLSI uses $\mathcal{E}_s(f, \log f)$, the entropy dissipation of the Markov dynamics.

Assumption 1 (**Modified log-Sobolev inequality over annealing path**). *The distribution π is told to satisfy the modified log-Sobolev inequality with a constant λ_π when there exists $\lambda_\pi > 0$ such that*

$$\mathcal{I}_\pi(\rho\|\pi) \;\geq\; 2\lambda_\pi\,\mathrm{KL}(\rho\|\pi) \qquad \textit{for all } \rho \in \mathcal{P}_{>0}(\Omega).$$

We say the annealing path $(\pi_s)_{s\in[0,S]}$ satisfies the pathwise modified log-Sobolev inequality (path-mSLI) when there exists a measurable function $s \mapsto \lambda_s > 0$ such π_s that the mLSI for all $s \in [0,1]$.

Indeed, fix s and let μ_u evolve under the frozen generator L_s. Writing $f_u = \mu_u/\pi_s$, mass conservation and reversibility give $\frac{\mathrm{d}}{\mathrm{d}u}\mathrm{KL}(\mu_u\|\pi_s) = \sum_x \mu_u(x) L_s \log f_u(x) = -\mathcal{I}_s(\mu_u\|\pi_s)$. Under Assumption 1, Gronwall's inequality therefore yields $\mathrm{KL}(\mu_u\|\pi_s) \leq e^{-2\lambda_s u}\mathrm{KL}(\mu_0\|\pi_s)$. For the annealed process, the reference distribution $\pi_{t/T}$ also changes with time, introducing an additional term in the entropy derivative. The transport action defined next provides a way to control this term.

Remark 1 (**Modified log-Sobolev inequality in continuous space**). *In continuous case, the Dirichlet form $\mathcal{E}(f,g) = \int \nabla f^\top \nabla g \mathrm{d}\pi$. Applying chain rule gives $\mathcal{E}(f, \log f) = 4\mathcal{E}(\sqrt{f}, \sqrt{f})$ which implies the equivalence of the LSI and mLSI in the continuous space. Therefore, the key differentiation of the mLSI occurs in the avaialblility in the discrete space where the Dirichlet forms do not admit a spatial gradient but must be defined on the graph (Ω, E).*

2.1 Annealed Jump Process for Statistical Physics

We quantify the cost of moving along the target path $(\pi_s)_{s\in[0,1]}$ through fluxes on the underlying graph. We use the discrete continuity-equation framework of discrete optimal transport (Maas, 2011; Erbar & Maas, 2012). Following the flux formulation of He et al. (2026), a flux $J_s : \Omega \times \Omega \to \mathbb{R}$ is antisymmetric, $J_s(x, y) = -J_s(y, x)$, and vanishes outside the edges. It is admissible for the target path if it satisfies the discrete continuity equation

$$\partial_s \pi_s(x) + \sum_{y:\{x,y\}\in E} J_s(x, y) = 0, \qquad x \in \Omega. \tag{1}$$

Here, $J_s(x, y)$ represents net transport from x to y. We write $\mathcal{J}(\pi)$ for the collection of measurable flux families satisfying (1) for almost every $s \in [0, 1]$. For an admissible flux, define its local transport energy by

$$G_s[J_s](x) := \frac{1}{2\pi_s(x)} \sum_{y:\{x,y\}\in E} \frac{J_s(x, y)^2}{c_s(x, y)}. \tag{2}$$

Its average under π_s is the quadratic flux energy, $\mathbb{E}_{\pi_s}[G_s[J_s]] = \sum_{\{x,y\}\in E} J_s(x, y)^2/c_s(x, y)$, where each undirected edge is counted once. Quadratic energies of this form arise in the dynamical formulation of discrete optimal transport (Maas, 2011; Erbar & Maas, 2012), and minimizing the integrated average energy over $J \in \mathcal{J}(\pi)$ gives the *discrete quadratic action* associated with the prescribed capacities $(c_s)_{s\in[0,1]}$, as used by He et al. (2026) in their analysis of simulated annealing.

Following Nitanda et al. (2026, Section 3 and Appendix 9.1), we use exponential moments of the local transport energy to obtain a change-of-measure bound for forward-KL analysis. This motivates the following definition of the discrete α-action.

Definition 1 (Discrete α-action). *For $\alpha > 0$ and $J \in \mathcal{J}(\pi)$, define $\mathfrak{E}_{\alpha,s}(J_s) := \frac{1}{\alpha} \log \mathbb{E}_{\pi_s}\big[\exp\big(\alpha G_s[J_s]\big)\big]$. The α-action associated with J and the intrinsic α-action of the path are, respectively,*

$$\mathcal{A}_\alpha(\pi, J) := \int_0^1 \mathfrak{E}_{\alpha,s}(J_s)\, ds, \qquad \mathcal{A}_\alpha(\pi) := \inf_{J\in\mathcal{J}(\pi)} \mathcal{A}_\alpha(\pi, J).$$

For $\alpha = 0$, set $\mathfrak{E}_{0,s}(J_s) := \mathbb{E}_{\pi_s}[G_s[J_s]]$ and define $\mathcal{A}_0(\pi, J)$ and $\mathcal{A}_0(\pi)$ by the same formulas.

At $\alpha = 0$, this definition recovers the quadratic action, while Jensen's inequality gives $\mathcal{A}_\alpha(\pi) \geq \mathcal{A}_0(\pi)$ for $\alpha > 0$. Moreover, the entropy variational inequality (Lemma 1) yields

$$\mathbb{E}_\rho[G_s[J_s]] \leq \frac{1}{\alpha}\mathrm{KL}(\rho\|\pi_s) + \mathfrak{E}_{\alpha,s}(J_s). \tag{3}$$

This bound controls transport energy under the evolving law (see Appendix B).

2.2 Masked Diffusion and Reward Tilting

For masked diffusion, $\tau \in [0, 1]$ denotes generation time, internal to a single sampling run and unrelated to the tilt strength a introduced below. Let $\mathcal{V}$ be a finite vocabulary, $\texttt{MASK} \notin \mathcal{V}$, and $\mathcal{X} = (\mathcal{V} \cup \{\texttt{MASK}\})^d$. Write $M(x) := \{i \in [d] : x^i = \texttt{MASK}\}$, and let $x^{i\leftarrow v}$ denote x with its ith entry replaced by v. Fix a nondecreasing C^1 reveal schedule $\kappa : [0, 1] \to [0, 1]$ with $\kappa(0) = 0$, $\kappa(1) = 1$, and $\kappa(\tau) < 1$ for $\tau < 1$; the masking kernel $K_\tau(x \mid z)$ independently retains each coordinate of z with probability $\kappa(\tau)$ and masks it otherwise. In particular $K_0(\cdot \mid z) = \delta_{(\texttt{MASK},\ldots,\texttt{MASK})}$ and $K_1(\cdot \mid z) = \delta_z$.

Given a reference ρ_1 on $\mathcal{V}^d$ and a reward $r : \mathcal{V}^d \to \mathbb{R}$, define the tilted distribution with respect to $a \in [0, A]$ (equivalent to the slow-annealing time s in Glauber dynamics) as

$$\rho_{1,a}(z) \propto \rho_1(z)e^{ar(z)}, \qquad \rho_{\tau,a}(x) := \sum_{z\in\mathcal{V}^d} K_\tau(x \mid z)\rho_{1,a}(z),$$

so that all tilt levels share the same masking kernel. In MDM scenario, $\rho_{1,a}$ corresponds to the π_s in the annealed jump process. L_s is a generator that evolves the distribution towards π_s, which in the MDM case it will be a learned generator Q_t that are fine-tuned to generate samples from $\rho_{1,a(t)}$.

Under the joint law $Z \sim \rho_{1,a}, Y_\tau \mid Z \sim K_\tau(\cdot \mid Z)$, define the unmasking posterior $\pi_a(v \mid \tau, x, i) := \mathbb{P}_a(Z^i = v \mid Y_\tau = x)$ for $\rho_{\tau,a}(x) > 0$ and $i \in M(x)$. The marginals $\{\rho_{\tau,a}\}_{\tau\in[0,1]}$ are realized by the jump rates

$$Q_{\tau,a}(x, x^{i\leftarrow v}) = \lambda(\tau)\, \pi_a(v \mid \tau, x, i), \qquad \lambda(\tau) := \frac{\dot\kappa(\tau)}{1-\kappa(\tau)}, \quad \tau < 1, \tag{4}$$

with all other off-diagonal rates zero. A parameterized predictor $\pi_\theta(v \mid \tau, x, i)$ defines model rates by replacing π_a in (4); fine-tuning at tilt level a aims to match the conditional posteriors of the reward-tilted target (Chen et al., 2026; So et al., 2026). We provide a detailed comparison of slowly annealed Glauber dynamics and slowly-annealed MDM fine-tuning at Appendix H.

3 Slowly-Annealed Jump Processes

We first establish a forward-KL bound for the continuous-time annealed dynamics, separating the decay of initialization error from the cost of tracking the target path in Section 3.1. We then give an exact implementation of annealed Glauber dynamics and bound the expected number of conditional sampling calls required to achieve a prescribed KL accuracy in Section 3.2.

3.1 Continuous-time convergence guarantees

The forward-KL analysis requires controlling transport energy under the evolving law ρ_t, whereas the action is defined under the target distribution $\pi_{t/T}$. The change-of-measure inequality (3) connects these two expectations. Combined with mLSI, it yields the following bound, which separates the decay of initialization error from the cost of tracking the target path.

Theorem 1 (Forward-KL tracking). *Assume that, for every $s \in [0,1]$, the frozen generator L_s satisfies λ_s-mLSI (Assumption 1) with $\lambda_s \geq 0$. Then, for every $T, \alpha > 0$ and $J \in \mathcal{J}(\pi)$ with $\mathcal{A}_\alpha(\pi, J) < \infty$,*

$$\mathrm{KL}(\rho_T\|\pi_1) \leq e^{-T\int_0^1 \lambda_s\, ds + \frac{1}{2\alpha T}}\, \mathrm{KL}(\rho_0\|\pi_0) + \frac{1}{2T}\int_0^1 e^{-T\int_s^1 \lambda_u\, du + \frac{1-s}{2\alpha T}}\, \mathfrak{E}_{\alpha,s}(J_s)\, ds. \tag{5}$$

In particular, taking the infimum over admissible fluxes gives

$$\mathrm{KL}(\rho_T\|\pi_1) \leq e^{-T\int_0^1 \lambda_s\, ds + \frac{1}{2\alpha T}}\, \mathrm{KL}(\rho_0\|\pi_0) + \frac{e^{1/(2\alpha T)}}{2T}\mathcal{A}_\alpha(\pi). \tag{6}$$

Theorem 1 bounds the terminal error $\mathrm{KL}(\rho_T\|\pi_1)$ by the sum of initialization and transport contributions. The initial mismatch $\mathrm{KL}(\rho_0\|\pi_0)$ is multiplied by $\exp(-T\int_0^1 \lambda_s\, ds + 1/(2\alpha T))$, where $T\int_0^1 \lambda_s\, ds$ measures the accumulated entropy contraction along the entire path. For fixed $\alpha > 0$, this factor vanishes exponentially as $T \to \infty$ whenever $\int_0^1 \lambda_s\, ds > 0$. Consequently, the contribution of $\mathrm{KL}(\rho_0\|\pi_0)$ disappears without requiring $\rho_0 = \pi_0$. The transport contribution in (5) is the second term. Here, $\mathfrak{E}_{\alpha,s}(J_s)$ measures the local transport cost of realizing the target velocity $\partial_s \pi_s$ through the admissible flux J_s. The factor $\exp(-T\int_s^1 \lambda_u\, du)$ captures the decay of this cost by the remaining entropy contraction $T\int_s^1 \lambda_u\, du$ between physical times Ts and T. The additional factor $\exp((1-s)/(2\alpha T))$ accounts for the accumulated change-of-measure correction from (3). Thus, the terminal effect of the local cost $\mathfrak{E}_{\alpha,s}(J_s)$ depends on both its magnitude and the subsequent dissipation $T\int_s^1 \lambda_u\, du$.

For a fixed path and $\alpha > 0$, the simplified bound (6) gives an $O(\mathcal{A}_\alpha(\pi)/T)$ transport contribution. Thus, slowing the traversal of the target path suppresses both sources of error, with the intrinsic action providing a bound optimized over all admissible fluxes. We defer the proof for the provided theorem at appendix B.

Remark 2 (Comparison with quadratic-action bounds (He et al., 2026)). *In our notation, He et al. (2026, Theorem 4.3) establish the* reverse-KL *bound* $\mathrm{KL}(\pi_1\|\rho_T) \leq \mathrm{KL}(\pi_0\|\rho_0) + \frac{\mathcal{A}_0(\pi)}{4T}$. *Theorem 1 instead controls* forward-KL $\mathrm{KL}(\rho_T\|\pi_1)$ *and multiplies the initial forward mismatch* $\mathrm{KL}(\rho_0\|\pi_0)$ *by* $\exp(-T\int_0^1 \lambda_s\, ds + 1/(2\alpha T))$, *capturing the decay of initialization error.*

These guarantees involve different assumptions and transport costs. Our bound uses λ_s-mLSI (Assumption 1) to quantify entropy contraction and the exponential action $\mathcal{A}_\alpha(\pi)$ to control transport energy under the evolving law. Their bound does not require mLSI but uses the quadratic action bound $\mathcal{A}_0(\pi) \le \mathcal{A}_\alpha(\pi)$.

Remark 3 (Relation to continuous-space α-action (Nitanda et al., 2026)). *The continuous-space α-action uses the local energy $\alpha^{-1}\log \mathbb{E}_{\pi_s}[\exp(\alpha\|v_s\|^2)]$, where $\partial_s\pi_s + \nabla\cdot(\pi_s v_s) = 0$. Our analysis replaces $\|v_s\|^2$ with $G_s[J_s]$ and uses λ_s-mLSI in place of LSI, providing a jump-process counterpart of their Corollary 1 and the weighted estimate in Appendix 9.1. Moreover, whereas their analysis selects the transport velocity minimizing the $L^2(\pi_s)$ norm, our intrinsic discrete α-action $\mathcal{A}_\alpha(\pi)$ minimizes the α-cost itself over all admissible fluxes $J \in \mathcal{J}(\pi)$.*

Remark 4 (Relation to classical entropy-based annealing (Del Moral & Miclo, 1999)). *Del Moral & Miclo (1999) already analyzed forward KL and controled it by using $|\mathbb{E}_\rho[\partial_s \log \pi_s]| \le \|\partial_s \log \pi_s\|_\infty \sqrt{2\mathrm{KL}(\rho\|\pi_s)}$ (Del Moral & Miclo, 1999, Lemma 1). Our analysis instead uses $|\mathbb{E}_\rho[\partial_s \log \pi_s]| \le \sqrt{\mathcal{I}_s(\rho\|\pi_s)\mathbb{E}_\rho[G_s[J_s]]}$ (Eq. (10)), which pairs target motion with entropy dissipation and leads to the discrete α-action $\mathcal{A}_\alpha(\pi)$.*

3.2 Application: Slowly Annealed Glauber Dynamics

Sampling setup and annealed Glauber dynamics. Let $\Omega = \mathcal{V}^d$, where $\mathcal{V}$ is a finite spin alphabet, and let $(\pi_s)_{s\in[0,1]}$ be a C^1 path in $\mathcal{P}_{>0}(\Omega)$. Given $\varepsilon \in (0,1)$, our goal is to output a configuration with law ν satisfying $\mathrm{KL}(\nu\|\pi_1) \le \varepsilon$. We consider Gibbs paths $\pi_s(x) \propto \exp(-\beta_s H(x))$ for a Hamiltonian $H : \Omega \to \mathbb{R}$ and an inverse-temperature schedule β_s, such as the mean-field Ising and Potts models considered by He et al. (2026, Section 5). We initialize from $\rho_0 = \mathrm{Unif}(\mathcal{V}^d)$ by sampling the coordinates independently and uniformly. Write $x^{i\leftarrow a}$ for x with coordinate i replaced by a. We assume access to an oracle that, for any $s \in [0,1]$, $i \in [d]$, and configuration x, returns a fresh exact sample from $\pi_s(\cdot \mid x_{-i})$. For Gibbs paths, this oracle can be implemented by normalizing the weights $\exp(-\beta_s H(x^{i\leftarrow a}))$ over $a \in \mathcal{V}$, since the partition function cancels. Each update therefore requires at most $|\mathcal{V}|$ energy evaluations and one categorical sample. We measure complexity by the expected number of conditional sampling calls. The corresponding random-scan heat-bath generator is

$$L_s h(x) = \frac{1}{n}\sum_{i=1}^{d}\sum_{a\in\mathcal{V}} \pi_s(a \mid x_{-i})\big(h(x^{i\leftarrow a}) - h(x)\big). \tag{7}$$

This generator is reversible with respect to π_s. For $T > 0$, the annealed Glauber process has generator $L_{t/T}$, initial law ρ_0, and marginal law ρ_t. All mLSI constants λ_s and transport actions in this subsection correspond to the normalization in (7).

Algorithm. Since $L_s = P_s - I$, where P_s is the random-scan heat-bath kernel, we simulate the process using a rate-one Poisson clock. At each event time t, we perform a single-site update using the conditional distribution at $s = t/T$. The procedure is given in Algorithm 1. Each event before T requires one conditional sampling call, including updates that leave X unchanged. The final waiting-time draw that crosses T requires no oracle call.

Analysis of Algorithm 1. We next specialize the sampling guarantee to the mean-field Ising and Potts models considered by He et al. (2026, Section 5).

Theorem 2 (Mean-field Ising model). *Let $d \ge 2$ and $\beta > 0$, and consider $\mu_\beta(x) \propto \exp(\frac{\beta}{d}\sum_{i<j} x_i x_j)$ on $\{-1,+1\}^d$. Run Algorithm 1 along $\pi_s = \mu_{s\beta}$, initialized from $\rho_0 = \mu_0 = \mathrm{Unif}(\{-1,+1\}^d)$. Then its output law ρ_T satisfies*

$$\mathrm{KL}(\rho_T\|\mu_\beta) \le \frac{d^5\beta^2}{4T}.$$

Thus, ε-forward-KL accuracy requires $O(d^5\beta^2/\varepsilon)$ expected single-site updates.

Remark 5 (Comparison with He et al. (2026, Theorem 5.10.)). *For the same annealing path and uniform initialization, He et al. (2026, Theorem 5.10) achieve $\mathrm{KL}(\mu_\beta\|\nu) \le \varepsilon$ for their output law ν using $O(d^5\beta^2/\varepsilon)$ expected single-site updates. Theorem 2 provides the complementary forward-KL guarantee $\mathrm{KL}(\rho_T\|\mu_\beta) \le \varepsilon$ with the same dependence on d, β, and ε.*

Algorithm 1 Uniformized exact annealed Glauber dynamics

Require: Horizon $T > 0$; exact conditional sampling oracle for $(\pi_s)_{s\in[0,1]}$
Sample $X_i \sim \mathrm{Unif}(\mathcal{V})$ independently for $i = 1, \dots, d$
Set $t \leftarrow 0$
while $t < T$ **do**
  Sample $\Delta \sim \mathrm{Exp}(1)$ and set $t \leftarrow t + \Delta$
  if $t > T$ **then**
    break
  Set $s \leftarrow t/T$ and sample $I \sim \mathrm{Unif}(\{1, \dots, d\})$
  Query $A \sim \pi_s(\cdot \mid X_{-I})$
  Set $X \leftarrow X^{I\leftarrow A}$
return X

For the Potts model, we use the block version of Algorithm 1. At each Poisson event, choose a subset $I \subseteq [d]$ of size $q-1$ uniformly and jointly resample X_I from $\pi_{t/T}(\cdot \mid X_{-I})$.

Theorem 3 (**Mean-field Potts model**). *Let $d \geq q \geq 2$, $\beta \geq q/2$, and $\varepsilon \in (0,1)$. Consider $\mu_\beta(x) \propto \exp(\frac{\beta}{d}\sum_{i,j=1}^{d} \mathbf{1}_{\{x_i=x_j\}})$ on $[q]^d$. Set $\beta_0 = d\log q + \log(6/\varepsilon)$, $\beta_s = \beta_0 + s(\beta - \beta_0)$, and $\pi_s = \mu_{\beta_s}$. Initialize the block sampler from $\rho_0 = \mathrm{Unif}(\Omega_{\mathrm{mono}})$, where $\Omega_{\mathrm{mono}} := \{(a, \dots, a) : a \in [q]\}$. There exists a universal constant $C > 0$ such that its output law satisfies*

$$\mathrm{KL}(\rho_T \| \mu_\beta) \leq \frac{\varepsilon}{6q} + \frac{d^{Cq}(\beta - \beta_0)^2}{2T}.$$

Thus, choosing $T = \max\{1, d^{Cq}(\beta - \beta_0)^2/\varepsilon\}$ achieves ε-forward-KL accuracy using $O(d^{Cq}(\beta + d\log q + \log(1/\varepsilon))^2/\varepsilon)$ expected block updates.

Remark 6 (**Comparison with He et al. (2026, Theorem 5.22)**). *For the same annealing path and $(q-1)$-block updates, He et al. (2026, Theorem 5.22) establish a reverse-KL guarantee. Our theorem provides the complementary forward-KL guarantee while retaining polynomial expected block-update complexity for fixed q. Moreover, we initialize uniformly over monochromatic configurations, without the additional full-space uniform component used in their initialization.*

4 Slowly-Annealed Reward Fine-tuning

We now consider the application slowly-annealed jump process on reward fine-tuning of Masked Diffusion Model. We specifically consider the theoretical analysis of the provable benefit incurred by the slowly-annealed jump process on two fine-tuning tasks: Slowly-Annealed Discrete Tilt Matching (SA-DTM) and Slowly-Annealed Discrete Adjoint Matching (SA-DAM). The difference between slowly-annealed jump process in Section 3 and slowly-annealed reward fine-tuning is "learning the posterior (SA-DTM) and generator (SA-DAM) with slowly annealing with provable benefit" which the convergence in terms of KL divergence (inside the Lyapunov function) between learned and the target distribution is dependent on the annealing schedule while annealing the reward during fine-tuning.

In this section, we study slowly annealed reward fine-tuning of masked diffusion models using the reward-tilted family introduced in Section 2.2. We analyze slowly annealed variants of DTM in Section 4.1 and DAM in Section 4.2, deriving conditional KL tracking bounds that quantify the effect of annealing speed under idealized learning dynamics. All proofs are deferred to Appendix E.

4.1 Slowly-Annealed Discrete Tilt Matching

Discrete Tilt Matching (DTM) (Chen et al., 2026) is a likelihood-free fine-tuning method for masked diffusion language models (dLLMs) that reformulates reward fine-tuning as state-level matching of local unmasking posteriors under progressive reward tilting via annealing. DTM advances along the tilt path in increments of size h: from rollouts of the current model at tilt a, it minimizes a masked-position cross-entropy in which each rollout is reweighted by $e^{hr(x_1)}$, so the population minimizer is the tilted posterior $\pi_{a+h}(\cdot \mid x_\tau, i)$ with no likelihood evaluation. Since π_{a+h} is inaccessible, the

target slot holds a random distribution $T_c(\cdot \mid x_\tau, i, x_1)$ interpolating between the one-hot indicator of the revealed token ($c = 0$) and the current posterior $\pi_a(\cdot \mid x_\tau, i)$. Discrete Tilt Matching naturally admits annealing schedule with $h = 6$ as an optimal schedule. Yet, this annealing step is based on an empirical observation which our analysis demonstrates the empirical benefit of slow-annealing.

Let $\mathrm{osc}(r) = \sup_{x,x' \in \mathrm{supp}\,\rho_1} |r(x) - r(x')|$ denote the oscillation of the reward function. Let the Lyapunov function of DTM $H(s)$ over the annealing schedule $a : [0, T] \to [0, A]$ be defined as

$$H(t) = \mathbb{E}_{\rho_{\tau,a(t)}}[\mathrm{KL}\left(\pi_a(\cdot \mid x_\tau, i) \parallel \pi_\theta(\cdot \mid x_\tau, i)\right)].$$

We let Assumption 1 hold along the annealing path $\{\rho_{1,a}\}_{a\in[0,A]}$. The convergence rate of the respective Lyapnuv function with respect to the annealing schedule is in fact governed by the following proposition:

Theorem 4. *Let the slowdown schedule be as $a : [0, T] \to [0, A]$. Then for $\alpha(t) = \frac{\lambda_0 \delta_{a(t)}}{2}$ and $\alpha_\chi = C\alpha(t)$ for some constant C,*

$$H(T) \le e^{-\frac{\alpha_\chi T}{2} + \mathrm{osc}(r)A} H(0) + \frac{\mathrm{osc}^2(r)}{2\alpha_\chi} \int_0^T e^{-\frac{\alpha_\chi (T-u)}{2} + \mathrm{osc}(r)(A - a(u))} \dot{a}(u)^2 \, du. \tag{8}$$

Intuitively, the analysis provides the analysis on the collapse of KL divergence between the annealing path π_a and the learned path π_θ in reverse-KL manner as in the Guo et al. (2025). We see that Equation (8) evaluates the terminal Lyapunov function at T, and decomposible into the initial contribution $H(0)$ term and the action term that is quadratic in $\dot{a}$. Therefore, when α_χ and $\mathrm{osc}(r)$ are fixed and positive, lowering $\dot{a}$ (i.e. applying slowdown to the annealing speed) allows improved convergence rate of the KL divergence between the annealing path and the target.

4.2 Slowly-Annealed Discrete Adjoint Matching

Similarly, the annealing and its slowdown counterpart brings an improvement on the Discrete Adjoint Matching (DAM) (So et al., 2026). DAM is a fine-tuning method for MDM (and to general discrete diffusion models) that solves entropy-regularized reward optimization by matching the model's rate to a statistically-derived discrete adjoint estimator of the optimal rate. In contrast to DTM, DAM is a one-shot matching that does not leverage an annealing in the training, and minimizes a single-step matching loss

$$\mathcal{L}_{\mathrm{DAM}}(\theta) := p_\tau^{\bar{u}}(Y \mid X_\tau)^{-1} \cdot D_{\mathrm{gKL}}\left(u_\tau^\theta(Y, X_\tau),\ (u_\tau^{\mathrm{base}} \hat{a}_\tau)(Y, X_\tau)\right).$$

Yet, we shall apply annealing on the tilt strength A such that the tilted target gradually transforms by the DAM loss which provides a provable benefit in terms of the sample size for accurate estimation of the adjoint $\hat{a}_\tau$ (see Appendix E.2 Proposition 1 and 2). We argue that the slow-annealing improves the convergence guarantee of the Lyapunov function which is KL divergence between the learned generator Q^θ. In contrast to comparison between unannealed vs. annealed, the annealing speed has benefit in terms of the We define the Lyapunov function across the family of tilted marginals $\{\rho^{*,a}\}_{a\in[0,A]}$ with respect to context $c = (\tau, x)$ as

$$H(t) := \mathbb{E}_{\mathbb{P}^{*,a(t)}}\left[\int_0^1 \lambda_\tau^{\mathrm{base}}(X_\tau) k_t(\tau, X_\tau) d\tau\right] = \int_0^1 \sum_x w_t(c) k_t(\tau, x) d\tau,$$

where $w_t(c) := \rho_\tau^{*,a(t)}(x)\lambda_\tau^{\mathrm{base}}(x)$, $k_t(c) := \mathrm{KL}(Q_t^\theta(\cdot \mid c) \| Q^{*,a(t)}(\cdot \mid c))$, and for the number of jumps $\mathfrak{N}$, we have $\mathbb{E}_{\mathbb{P}^{\theta,a}}\left[\int_0^1 \lambda_\tau^{\mathrm{base}}(X_\tau) \mathrm{d}\tau\right] = \mathfrak{N}$.

Theorem 5. *Assume the above assumptions for DAM analysis and the annealing path satisfy Assumption 1 where the infimum exists such that $\alpha_t = \inf_c(\alpha_t(c))$. The convergence rate of the Lyapunov function with mirror flow rate a_0 across $t \in [0, T]$ is given by*

$$H(T) \le e^{-\frac{a_0 \alpha_\chi T}{2} + \mathrm{osc}(r)A} H(0) + \frac{\mathfrak{N}\,\mathrm{osc}^2(r)}{4a_0\alpha_\chi} \int_0^T e^{-\frac{a_0 \alpha_\chi (T-u)}{2} + \mathrm{osc}(r)(A - a(u))} \dot{a}(u)^2 \, \mathrm{d}u. \tag{9}$$

Similarly with the SA-DTM, Equation (9) of SA-DAM also demonstrates the theoretical guarantee that slower annealing improves the convergence rate of the reward fine-tuning. Similar to the Proposition 4, we have the terminal Lyapunov function of the SA-DAM improves when $\dot{a}$ is smaller in both terms when other hyper-parameters are fixed.

Remark 7 (Finite-budget fine-tuning with slow-annealing). *For both SA-DTM and SA-DAM, the analysis is under idealized training dynamics and sufficient training, which the rate of convergence is improved with smaller $\dot{a}$. SA-DAM requires more complex training procedure compared to SA-DTM since 1) it requires sampling process 2) its one-shot loss function. This shows that the finite-budget training may often show a result that contradicts the theory but sufficient budget reproduces to support the theoretical contribution. Details are left at section 5 and appendix G.*

5 Experiments

We validate the efficiency of slowdown on annealing schedule on various reward fine-tuning task. We specifically consider the two MDM fine-tuning methods Discrete Tilt Matching (DTM) and Discrete Adjoint Matching (DAM) (Chen et al., 2026; So et al., 2026). We validate whether the theoretical guarantees established above is empirically validated in various scenarios by KL divergence and entropy at small task and by mean reward and accuracy at complex task.

5.1 Small Toy Distribution

Setup For the small tractable distribution, we consider slowdown on a simple tractable task of length-4 sequence of 5 vocabularies. In such case, there are total 5^4 possible sequences and distribution over it. Furthermore, the base marginal is defined as a seed unary with a pairwise Potts-like energy function. The reward function is defined as $r = \frac{\#\{i:x_i=x_{i+1}\}}{L-1} \in [0,1]$ such that $r = 1$ when all tokens are the same. We test over 3 different tilt strength $\{8, 16, 24\}$ and measure the wall clock time for the training where both training and evaluation is done on a single NVIDIA H200 GPU.

The empirical result at figure 1 provides a gradual decrease in the terminal KL divergence between the model distribution and the target distribution. At small scale, at single-step training with $N = 1$, the KL divergence remains at 1.84×10^{-1} after training over 32768 steps across stage. On the other hand, a fine-grained slowdown schedule provides significantly lower KL divergence. The KL divergence is reduced by $10\times$ to 7.61×10^{-2} by letting $N = 16$ while the increase in the wall clock time is remains at 5.67%. Furthermore, slowdown by $N = 128$ further decreases the KL divergence fo 1.60×10^{-2}. For larger oscillation, similar trend occurs such that the slowdown from $N = 1$ to $N = 128$ decreases the KL divergence at the end of training from 1.50×10^{1} to 7.94×10^{0}, approximately halving the KL divergence within the 17.1% in the wall clock time increase. The benefit of SA-DTM is further validated in the entropy which at large scale $A \cdot \mathrm{osc} = 16, 24$, the increase in entropy is evident. At $A \cdot \mathrm{osc} = 16$, the entropy increases from 1.17 to 1.34 by 14.5% and at $A \cdot \mathrm{osc} = 24$, the entropy increases from 0.75 to 1.03 by 37.3%.

Similarly for SA-DAM, at $A \cdot \mathrm{osc} = 8$, the KL divergence decreases significantly from 1.27×10^{-2} to 9.18×10^{-5} (only 0.72% of identity slowdown case) with approximately doubled wall-clock time. There are two different phenomenon observed for DAM: 1) the entropy is already saturated at $N = 1$ and 2) the KL divergence increases and entropy decreases for $N = 128$ compared to $N = 64$ indicating a degradation of the quality. Observation 1 indicates the learning of DAM is faster compared to the DTM with respect to the gradient steps while taking significantly longer per gradient step. Observation 2 at first glance seems counter-intuitive which demonstrates that the more slowdown results in a degradation of the quality. However, this is due to the lack of training budget per annealing grid. Further increasing the annealing grid alleviates this U-shaped phenomenon and provides the theoretically-supported improved convergence. We provide the detailed results with full table at appendix G.

5.2 Large Scale Fine-tuning

To test the scalability of our result on more complex task, we test our theoretical result on tasks of Maze Planning and Llada-8B-Instruct fine-tuning (Sudoku and Countdown). We fix the training budget as 5120 gradient steps and distribute them equally across the annealing grid. The best result for the maze planning in terms of the reward and validity is achieved at the most fine-grained annealing grid with $N = 40$. While the pretrained base model has produced 83.59% of valid paths, annealing with stronger annealing slowdown $N = 4$ to $N = 40$ provides increasing valid path rate of 84.11% to 86.65%. The mean reward increases from 0.324 to 0.349. Similarly on Sudoku task, we observe a

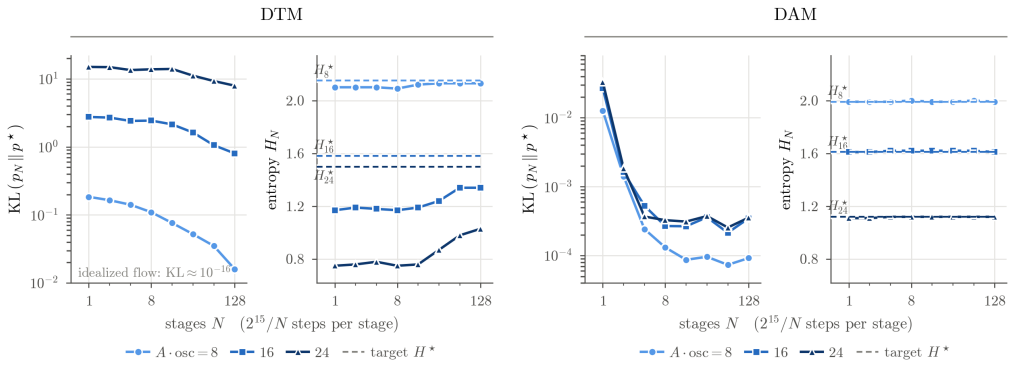

Figure 1: Comparison between DTM and DAM models on tractable target distribution.

Table 1: DTM performance across varying step horizons h alongside mean reward and accuracy. We set the final reward tilt $A = 5$ for maze Planning and $A = 100$ for Sudoku and CountDown.

	Maze Planning		**Sudoku**		**CountDown**	
N	Mean Reward	Validity (%)	Mean Reward	Accuracy (%)	Mean Reward	Accuracy (%)
Base	0.324	83.59	0.540	27.00	0.272	20.31
4	0.329	84.11	1.708	85.41	0.285	20.70
5	0.332	84.50	1.952	97.61	0.353	28.12
10	0.345	85.48	1.971	98.56	0.374	30.47
20	0.346	85.93	1.977	98.85	0.501	44.53
40	0.349	86.65	1.980	99.00	0.469	41.02

monotonic increase in the mean reward from 1.708 to 1.980 as the annealing slowdown is applied by $10\times$. Concurrently, the mean accuracy increases from 85.41% to 99.00% as the slowdown is strengthened. More specifically, the performance increases significantly between $N = 4$ and $N = 5$ while reaches approximate plateau at around 99.00%, which is very close to the optimal value. Similar observation is made at the Countdown task except for the strongest slowdown $N = 40$ case. While both mean reward and accuracy monotonically increases across the slowdown strength N, the mean ward and accuracy both degraded at $N = 40$ compared to $N = 20$ from 0.501 and 44.53% to 0.469 and 41.02%. Similar to the discrete adjoint matching experiment on a small scale, this is likely due to the training budget per stage similar to table 3. The higher complexity of the countdown task requires larger number of gradient steps at each stage, which lacks sufficient training per stage results in degradation of the performance again. Chen et al. (2026) refers to 800 steps as gradient descent budget at each step while our $N = 40$ with 256 gradient steps are undertrained, which similar to the result obtained by increasing gradient step per annealing grid on DAM has resolved the U-shaped optimality problem and resulted in a monotonic convergence.

6 Conclusion and Discussion

We present slowly-annealed jump process, the discrete counterpart of slowly-annealed Langevin Dynamics by introducing the modified log-Sobolev inequality. We show the success of slow annealing in the continuous state stochastic process directly transfers to the discrete state space specifically on the Poisson jump process. We further demonstrate the application of slowly-annealed jump process in the discrete diffusion model which is characterized by the Levy-type stochastic integral. We first derive the forward KL bound for the slowly-annealed fine-tuning of the masked diffusion model. We validate the theoretical studies on both small scale and large scale fine-tuning tasks which at small scale, it demonstrates an improved KL divergence gap to the target density with slow annealing. At large scale task of Maze planning and LLaDA-8B-Instruct fine-tuning, the empirical success of the slowly-annealed masked diffusion fine-tuning validate the theoretical analysis. This work argues the importance of slow-annealing in MDM fine-tuning to provide a sufficient time for the learning dynamics to track the target evolution for both theoretical and empirical success.

AI USE STATEMENT

In this work, AI are used for the following purposes:

1. Generation of proof sketch for Section 3, 4 and corresponding appendices.
2. Code generation for the experiments on DTM, DAM.
3. Anonymizing the code for submission.
4. Construction of the tables (formatting Table 1, 2, 3, 4).
5. Visualization of Figure 3.
6. Writing section F.2. The format was adopted from (Chen et al., 2026, Appendix D.2).

ETHICS STATEMENT

This work is theoretical in nature: we establish convergence guarantees for reward-guided fine-tuning algorithms (SA-DTM and SA-DAM) applied to masked discrete diffusion models. We do not collect, use, or release any new datasets, nor do our experiments (if any) involve human subjects, personal data, or sensitive attributes. We do not foresee direct negative societal impacts stemming from the theoretical contributions of this paper.

REPRODUCIBILITY STATEMENT

We provide the anonymized github repository for reproducibility and leave the experimental details at Appendix F. https://anonymous.4open.science/r/sa_mdm_finetuning-FE2C

REFERENCES

Sergey G Bobkov and Prasad Tetali. Modified logarithmic sobolev inequalities in discrete settings. *Journal of Theoretical Probability*, 19(2):289–336, 2006.

Yuyuan Chen, Shiyi Wang, Peter Potaptchik, Jaeyeon Kim, and Michael Samuel Albergo. Discrete tilt matching. In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=H2Wueh9MMy.

Pierre Del Moral and Laurent Miclo. On the convergence and applications of generalized simulated annealing. *SIAM Journal on Control and Optimization*, 37(4):1222–1250, 1999.

Arnaud Doucet, Will Sussman Grathwohl, Alexander G. D. G. Matthews, and Heiko Strathmann. Score-based diffusion meets annealed importance sampling. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho (eds.), *Advances in Neural Information Processing Systems*, 2022. URL https://openreview.net/forum?id=9cU2iW3bz0.

Matthias Erbar and Jan Maas. Ricci curvature of finite markov chains via convexity of the entropy. *Archive for Rational Mechanics and Analysis*, 206(3):997–1038, 2012.

Wei Guo, Molei Tao, and Yongxin Chen. Provable benefit of annealed langevin monte carlo for non-log-concave sampling. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=P6IVIoGRRg.

Yuchen He, Tianhui Jiang, Sihan Wang, and Chihao Zhang. Discrete optimal transport: Rapid convergence of simulated annealing algorithms, 2026. URL https://arxiv.org/abs/2605.05877.

Jan Maas. Gradient flows of the entropy for finite markov chains. *Journal of Functional Analysis*, 261(8):2250–2292, 2011.

J. Machta. Population annealing with weighted averages: A monte carlo method for rough free-energy landscapes. *Phys. Rev. E*, 82:026704, Aug 2010. doi: 10.1103/PhysRevE.82.026704. URL https://link.aps.org/doi/10.1103/PhysRevE.82.026704.

Atsushi Nitanda, Dake Bu, Yueming Lyu, and Tanya Veeravalli. Slowly annealed langevin dynamics: Theory and applications to training-free guided generation, 2026. URL https://arxiv.org/abs/2605.07950.

Arnau Prat Pou, Enrique Romero, Jordi Martí, and Ferran Mazzanti. Mean field initialization of the annealed importance sampling algorithm for an efficient evaluation of the partition function using restricted boltzmann machines. *Entropy*, 27(2), 2025. ISSN 1099-4300. doi: 10.3390/e27020171. URL https://doi.org/10.3390/e27020171.

Oswin So, Brian Karrer, Chuchu Fan, Ricky T. Q. Chen, and Guan-Horng Liu. Discrete adjoint matching. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=VXB4xxAgOf.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In Francis Bach and David Blei (eds.), *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 2256–2265, Lille, France, 07–09 Jul 2015. PMLR. URL https://proceedings.mlr.press/v37/sohl-dickstein15.html.

Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché Buc, E. Fox, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/3001ef257407d5a371a96dcd947c7d93-Paper.pdf.

Chenyu Wang, Masatoshi Uehara, Yichun He, Amy Wang, Avantika Lal, Tommi Jaakkola, Sergey Levine, Aviv Regev, Hanchen, and Tommaso Biancalani. Fine-tuning discrete diffusion models via reward optimization with applications to DNA and protein design. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=G328D1xt4W.

Dongze Wu and Yao Xie. Annealing flow generative models towards sampling high-dimensional and multi-modal distributions. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=iT8qIzNSAl.

Lexing Ying. Annealed importance sampling for ising models with mixed boundary conditions. *Journal of Computational Mathematics*, 41(3):542–550, 2023. doi: 10.4208/jcm.2211-m2022-0172. URL https://doi.org/10.4208/jcm.2211-m2022-0172.

A Useful Facts

Lemma 1 (**Entropy variational inequality**). *Let π and ρ be probability measures on a finite set Ω with $\rho \ll \pi$. Then for every function $g : \Omega \to \mathbb{R}$ and every $\alpha > 0$,*

$$\mathbb{E}_\rho[g] \le \frac{1}{\alpha}\mathrm{KL}(\rho\|\pi) + \frac{1}{\alpha}\log \mathbb{E}_\pi \exp(\alpha g).$$

Proof. By the variational formula for relative entropy, for any function $h : \Omega \to \mathbb{R}$,

$$\log \mathbb{E}_\pi e^h = \sup_{\nu \ll \pi} \left\{ \mathbb{E}_\nu h - \mathrm{KL}(\nu\|\pi) \right\}.$$

Taking $\nu = \rho$ gives

$$\mathbb{E}_\rho h - \mathrm{KL}(\rho\|\pi) \le \log \mathbb{E}_\pi e^h.$$

Applying this with $h = \alpha g$ yields

$$\alpha \mathbb{E}_\rho[g] - \mathrm{KL}(\rho\|\pi) \le \log \mathbb{E}_\pi e^{\alpha g}.$$

Dividing by α proves the claim. □

B Proof of Theorem 1

Proof of Theorem 1. Set

$$s = \frac{t}{T}, \qquad f_t(x) := \frac{\rho_t(x)}{\pi_s(x)}, \qquad H_t := \mathrm{KL}(\rho_t \| \pi_s).$$

We differentiate H_t along the inhomogeneous dynamics. Since

$$H_t = \sum_x \rho_t(x) \log f_t(x),$$

we have

$$\frac{d}{dt} H_t = \sum_x \partial_t \rho_t(x) \log f_t(x) - \frac{1}{T} \sum_x f_t(x) \partial_s \pi_s(x).$$

Here we used $\sum_x \partial_t \rho_t(x) = 0$.

The first term is the entropy dissipation of the frozen chain. Since p_s is reversible with respect to π_s,

$$\sum_x \partial_t \rho_t(x) \log f_t(x) = -\mathcal{I}_s(\rho_t \| \pi_s).$$

Therefore

$$\frac{d}{dt} H_t = -\mathcal{I}_s(\rho_t \| \pi_s) - \frac{1}{T} \sum_x f_t(x) \partial_s \pi_s(x).$$

Using the continuity equation

$$\partial_s \pi_s + \operatorname{div} J_s = 0,$$

we rewrite the second term as

$$-\sum_x f_t(x) \partial_s \pi_s(x) = \sum_x f_t(x) \operatorname{div} J_s(x).$$

By discrete integration by parts,

$$\sum_x f_t(x) \operatorname{div} J_s(x) = \sum_{\{x,y\} \in E} J_s(x,y) \big(f_t(x) - f_t(y)\big),$$

where an arbitrary orientation of each edge is fixed. Hence

$$\frac{d}{dt} H_t = -\mathcal{I}_s(\rho_t \| \pi_s) + \frac{1}{T} \sum_{\{x,y\} \in E} J_s(x,y) \big(f_t(x) - f_t(y)\big).$$

We now bound the forcing term. Let

$$\theta(a,b) := \frac{a-b}{\log a - \log b}$$

be the logarithmic mean. Then

$$f_t(x) - f_t(y) = \theta(f_t(x), f_t(y)) \big(\log f_t(x) - \log f_t(y)\big).$$

By Cauchy–Schwarz,

$$\begin{aligned}
&\left| \sum_{\{x,y\} \in E} J_s(x,y) \big(f_t(x) - f_t(y)\big) \right| \\
&\le \left[\sum_{\{x,y\} \in E} c_s(x,y) \big(f_t(x) - f_t(y)\big) \big(\log f_t(x) - \log f_t(y)\big) \right]^{1/2} \\
&\quad \times \left[\sum_{\{x,y\} \in E} \theta(f_t(x), f_t(y)) \frac{J_s(x,y)^2}{c_s(x,y)} \right]^{1/2} \\
&= \sqrt{\mathcal{I}_s(\rho_t \| \pi_s)} \left[\sum_{\{x,y\} \in E} \theta(f_t(x), f_t(y)) \frac{J_s(x,y)^2}{c_s(x,y)} \right]^{1/2}.
\end{aligned}$$

Using the elementary inequality

$$\theta(a,b) \le \frac{a+b}{2},$$

we obtain

$$\begin{aligned}\sum_{\{x,y\}\in E} \theta(f_t(x), f_t(y)) \frac{J_s(x,y)^2}{c_s(x,y)} &\le \frac{1}{2} \sum_{\{x,y\}\in E} \big(f_t(x) + f_t(y)\big) \frac{J_s(x,y)^2}{c_s(x,y)} \\ &= \sum_x \rho_t(x) \frac{1}{2\pi_s(x)} \sum_{y:\{x,y\}\in E} \frac{J_s(x,y)^2}{c_s(x,y)} \\ &= \mathbb{E}_{\rho_t} G_s^J.\end{aligned}$$

Therefore

$$\left| \sum_{\{x,y\}\in E} J_s(x,y)\big(f_t(x) - f_t(y)\big) \right| \le \sqrt{\mathcal{I}_s(\rho_t\|\pi_s)} \sqrt{\mathbb{E}_{\rho_t} G_s^J}. \tag{10}$$

It follows that

$$\frac{d}{dt} H_t \le -\mathcal{I}_s(\rho_t\|\pi_s) + \frac{1}{T}\sqrt{\mathcal{I}_s(\rho_t\|\pi_s)} \sqrt{\mathbb{E}_{\rho_t} G_s^J}.$$

By Young's inequality,

$$\frac{1}{T}\sqrt{\mathcal{I}_s(\rho_t\|\pi_s)} \sqrt{\mathbb{E}_{\rho_t} G_s^J} \le \frac{1}{2}\mathcal{I}_s(\rho_t\|\pi_s) + \frac{1}{2T^2}\mathbb{E}_{\rho_t} G_s^J.$$

Thus

$$\frac{d}{dt} H_t \le -\frac{1}{2}\mathcal{I}_s(\rho_t\|\pi_s) + \frac{1}{2T^2}\mathbb{E}_{\rho_t} G_s^J.$$

By the modified log-Sobolev assumption,

$$\mathcal{I}_s(\rho_t\|\pi_s) \ge 2\lambda_s H_t.$$

Therefore

$$\frac{d}{dt} H_t \le -\lambda_s H_t + \frac{1}{2T^2}\mathbb{E}_{\rho_t} G_s^J.$$

We now apply the entropy variational inequality (Lemma 1). For every $\alpha > 0$,

$$\mathbb{E}_{\rho_t} G_s^J \le \frac{1}{\alpha}\mathrm{KL}(\rho_t\|\pi_s) + \frac{1}{\alpha}\log \mathbb{E}_{\pi_s} \exp(\alpha G_s^J).$$

By the definition of $\mathcal{E}_\alpha^{\mathrm{disc}}$, this gives

$$\mathbb{E}_{\rho_t} G_s^J \le \frac{1}{\alpha} H_t + \mathcal{E}_\alpha^{\mathrm{disc}}(\pi_s, J_s).$$

Consequently,

$$\frac{d}{dt} H_t \le \left(-\lambda_s + \frac{1}{2\alpha T^2}\right) H_t + \frac{1}{2T^2}\mathcal{E}_\alpha^{\mathrm{disc}}(\pi_s, J_s).$$

Define

$$\bar{H}_s := H_{Ts}.$$

Since $d\bar{H}_s/ds = T\, dH_t/dt$, we get

$$\frac{d}{ds}\bar{H}_s \le \left(-T\lambda_s + \frac{1}{2\alpha T}\right)\bar{H}_s + \frac{1}{2T}\mathcal{E}_\alpha^{\mathrm{disc}}(\pi_s, J_s).$$

Applying Grönwall's inequality on $[0,1]$ yields

$$\begin{aligned}\bar{H}_1 \le \exp\left(-T\int_0^1 \lambda_s\, ds + \frac{1}{2\alpha T}\right)\bar{H}_0 \\ + \frac{1}{2T}\int_0^1 \exp\left(-T\int_s^1 \lambda_u\, du + \frac{1-s}{2\alpha T}\right)\mathcal{E}_\alpha^{\mathrm{disc}}(\pi_s, J_s)\, ds.\end{aligned}$$

Since

$$\bar{H}_1 = \mathrm{KL}(\rho_T \| \pi_1), \qquad \bar{H}_0 = \mathrm{KL}(\rho_0 \| \pi_0),$$

this proves the first bound.

For the simplified bound, use $\lambda_s \geq 0$ to obtain

$$\exp\left(-T \int_s^1 \lambda_u \, du + \frac{1-s}{2\alpha T}\right) \leq e^{\frac{1}{2\alpha T}}.$$

Therefore

$$\mathrm{KL}(\rho_T \| \pi_1) \leq e^{-T\int_0^1 \lambda_s \, ds + \frac{1}{2\alpha T}} \mathrm{KL}(\rho_0 \| \pi_0) + \frac{e^{\frac{1}{2\alpha T}}}{2T} \mathcal{A}_\alpha^{\mathrm{disc}}(\pi, J).$$

Finally, taking the infimum over admissible fluxes J gives the intrinsic α-action bound. □

C Proof of Theorem 2

C.1 Main argument

The main step is to construct an admissible flux whose local transport energy is bounded uniformly over configurations. Such a bound controls both the discrete α-action and the transport term in the entropy differential inequality. We obtain it by reducing the Ising model to a one-dimensional graph, constructing a flux there, and lifting it back to the original state space.

Symmetry reduction and flux lifting. For the mean-field Ising model, coordinate permutations and the global spin flip $x \mapsto -x$ preserve both the target distribution and the heat-bath capacities. Configurations related by these operations have the same absolute magnetization $|\sum_i x_i|$. Grouping them together produces a path graph with $\lfloor n/2 \rfloor + 1$ vertices. We will construct a flux on this smaller graph.

We first describe a general lifting construction, which will also be used for the Potts model. On a finite graph (Ω, E), let Γ be a fixed finite group of bijections of Ω preserving the target distributions and capacities: $\pi_s(gx) = \pi_s(x)$ and $c_s(gx, gy) = c_s(x, y)$ for every $g \in \Gamma$, $s \in [0,1]$, and $x, y \in \Omega$. Capacities are extended by zero outside E. The *orbit* of x is the set $\{gx : g \in \Gamma\}$. These orbits partition Ω, and π_s is uniform within each orbit.

Let $\overline{\Omega} = \Omega/\Gamma$ denote the set of orbits. The quotient graph has an edge between distinct orbits A, B whenever an edge of E connects them. Write $\overline{E}$ for its edge set, and define

$$\overline{\pi}_s(A) := \sum_{x \in A} \pi_s(x), \qquad \overline{c}_s(A, B) := \sum_{x \in A} \sum_{y \in B} c_s(x, y), \quad A \neq B.$$

Admissible fluxes and local energies $\overline{G}_s$ on this graph are defined as in (1) and (2), using $(\overline{\pi}_s, \overline{c}_s)$.

The following lemma distributes the flux between two orbits over the original edges in proportion to their capacities.

Lemma 2 (Flux lifting under symmetry). *Under the preceding symmetry assumptions, let $\overline{J}$ be an admissible flux for $(\overline{\pi}_s)_{s\in[0,1]}$. For $x \in A$ and $y \in B$, define*

$$J_s(x, y) := \begin{cases} \dfrac{c_s(x, y)}{\overline{c}_s(A, B)} \overline{J}_s(A, B), & \{x, y\} \in E, \ A \neq B, \\ 0, & \textit{otherwise}. \end{cases} \tag{11}$$

Then $J \in \mathcal{J}(\pi)$, and its local transport energy satisfies

$$G_s[J_s](x) = \overline{G}_s[\overline{J}_s](A), \qquad x \in A, \quad s \in [0,1]. \tag{12}$$

The proof is given in Section C.2. In particular, a uniform bound on the projected local energy gives the same bound on the original state space.

Pointwise transport energy. For Ising, the quotient graph is a path, so the continuity equation determines the flux across each edge: it equals the negative derivative of the total probability on the left of that edge. The unimodality and capacity estimates of He et al. (2026, Lemma 5.5 and Eq. (5)) allow us to bound this flux in terms of the probabilities at the edge endpoints. Combining these estimates with Lemma 2 yields the following bound.

Lemma 3 (Local transport energy for mean-field Ising). *Let $n \geq 2$ and let $s \mapsto \beta_s \in [0, \infty)$ be C^1. Set $\pi_s = \mu_{\beta_s}$ on $\Omega = \{-1, +1\}^n$, and let c_s be the capacities of the normalized single-site heat-bath generator (7). There exists a continuous admissible flux $J \in \mathcal{J}(\pi)$ such that*

$$\sup_{x \in \Omega} G_s[J_s](x) \leq \frac{n^5}{2} |\beta_s'|^2, \qquad s \in [0, 1]. \tag{13}$$

Consequently, for every $\alpha > 0$,

$$\mathcal{A}_\alpha(\pi) \leq \mathcal{A}_\alpha(\pi, J) \leq \frac{n^5}{2} \int_0^1 |\beta_s'|^2 \, ds. \tag{14}$$

The proof is given in Section C.3. The same flux satisfies (14) for every $\alpha > 0$.

Completion of the argument. Fix $\beta > 0$ and take $\beta_s = s\beta$. Since $\pi_0 = \mu_0$ is uniform on $\{-1, +1\}^n$, the prescribed initialization satisfies $\rho_0 = \pi_0$ and hence $\mathrm{KL}(\rho_0 \| \pi_0) = 0$.

Each frozen heat-bath generator is irreducible and reversible on the finite state space. Moreover, its edge rates are continuous and strictly positive along the compact parameter interval $[0, 1]$. Consequently, the generators satisfy Assumption 1 with a common constant $\lambda_* > 0$. We only need the existence of this constant, not a quantitative lower bound.

Let J be the admissible flux supplied by Lemma 3. Since $\beta_s' = \beta$, the same flux satisfies, for every $\alpha > 0$,

$$\begin{aligned} \mathcal{A}_\alpha(\pi) \leq \mathcal{A}_\alpha(\pi, J) &= \int_0^1 \mathfrak{E}_{\alpha, s}(J_s) \, ds \\ &\leq \frac{n^5}{2} \int_0^1 |\beta_s'|^2 \, ds = \frac{n^5 \beta^2}{2}. \end{aligned}$$

In particular, the action is finite for every $\alpha > 0$.

Applying Theorem 1, specifically (6), with $\lambda_s = \lambda_*$ gives

$$\begin{aligned} \mathrm{KL}(\rho_T \| \mu_\beta) &\leq e^{-T\lambda_* + \frac{1}{2\alpha T}} \underbrace{\mathrm{KL}(\rho_0 \| \pi_0)}_{=0} + \frac{e^{1/(2\alpha T)}}{2T} \mathcal{A}_\alpha(\pi) \\ &\leq \frac{e^{1/(2\alpha T)}}{2T} \mathcal{A}_\alpha(\pi, J) \\ &\leq e^{1/(2\alpha T)} \frac{n^5 \beta^2}{4T}. \end{aligned}$$

For fixed $T > 0$, the law ρ_T does not depend on α, and this inequality holds for every $\alpha > 0$. Letting $\alpha \to \infty$ therefore yields

$$\mathrm{KL}(\rho_T \| \mu_\beta) \leq \frac{n^5 \beta^2}{4T}.$$

This limit uses the uniform bound on $\mathcal{A}_\alpha(\pi, J)$ established above.

By the exactness of Algorithm 1, its output X_{out} has law ρ_T. The proposal times form a rate-one Poisson process, so the number of events before T satisfies $N_T \sim \mathrm{Pois}(T)$. Each event requires exactly one conditional sampling call, including updates that leave the configuration unchanged. The final waiting-time draw that crosses T requires no such call. Thus the expected number of conditional sampling calls is $\mathbb{E}[N_T] = T$.

For $\varepsilon \in (0, 1)$, choose $T_\varepsilon := n^5 \beta^2 / (4\varepsilon)$. Then

$$\mathrm{KL}\big(\mathcal{L}(X_{\mathrm{out}}) \| \mu_\beta\big) \leq \frac{n^5 \beta^2}{4T_\varepsilon} = \varepsilon,$$

$$\mathbb{E}[\text{number of conditional sampling calls}] = \mathbb{E}[N_{T_\varepsilon}] = T_\varepsilon = \frac{n^5 \beta^2}{4\varepsilon}.$$

Therefore, $O(n^5\beta^2/\varepsilon)$ expected single-site conditional sampling calls suffice, proving Theorem 2.

C.2 Proof of Lemma 2

Proof. We work under the general symmetry assumptions of the lemma. Since π_s is invariant under Γ, all states in an orbit A have equal probability. Thus $\pi_s(x) = \overline{\pi}_s(A)/|A|$ for $x \in A$.

For distinct orbits A, B, the sum $\sum_{y\in B} c_s(x,y)$ is independent of $x \in A$. Indeed, given $x, x' \in A$, choose $g \in \Gamma$ with $gx = x'$. The map g sends B bijectively onto itself, so capacity invariance gives $\sum_{y\in B} c_s(x',y) = \sum_{y\in B} c_s(gx, gy) = \sum_{y\in B} c_s(x,y)$. Summing this common value over $x \in A$ yields

$$\sum_{y\in B} c_s(x,y) = \frac{\overline{c}_s(A,B)}{|A|}, \qquad x \in A. \tag{15}$$

The lifted flux is antisymmetric because the capacities are symmetric and $\overline{J}_s$ is antisymmetric. It vanishes outside E by construction. At every s at which the projected continuity equation holds, for $x \in A$ we have

$$\begin{aligned}\sum_{y:\{x,y\}\in E} J_s(x,y) &= \sum_{B:\{A,B\}\in\overline{E}} \frac{\overline{J}_s(A,B)}{\overline{c}_s(A,B)} \sum_{y\in B} c_s(x,y) \\ &= \frac{1}{|A|} \sum_{B:\{A,B\}\in\overline{E}} \overline{J}_s(A,B) \\ &= -\frac{\partial_s \overline{\pi}_s(A)}{|A|} = -\partial_s \pi_s(x).\end{aligned}$$

The lift is measurable because the original flux and capacities are measurable and the denominators are positive on quotient edges. Therefore $J \in \mathcal{J}(\pi)$.

For the energy identity, substitute (11) into the definition of G_s and use (15):

$$\begin{aligned}G_s[J_s](x) &= \frac{1}{2\pi_s(x)} \sum_{B:\{A,B\}\in\overline{E}} \frac{\overline{J}_s(A,B)^2}{\overline{c}_s(A,B)^2} \sum_{y\in B} c_s(x,y) \\ &= \frac{1}{2|A|\pi_s(x)} \sum_{B:\{A,B\}\in\overline{E}} \frac{\overline{J}_s(A,B)^2}{\overline{c}_s(A,B)} \\ &= \frac{1}{2\overline{\pi}_s(A)} \sum_{B:\{A,B\}\in\overline{E}} \frac{\overline{J}_s(A,B)^2}{\overline{c}_s(A,B)} = \overline{G}_s[\overline{J}_s](A).\end{aligned}$$

This proves (12). □

C.3 Proof of Lemma 3

Throughout, let $n \geq 2$ and $\beta_\cdot \in C^1([0,1];[0,\infty))$. Let Γ be generated by coordinate permutations and the global spin flip on $\Omega = \{-1,+1\}^n$. Both $\pi_s = \mu_{\beta_s}$ and its heat-bath capacities c_s are Γ-invariant. The orbits are

$$A_k := \left\{ x : \left| \sum_{i=1}^n x_i \right| = m_k \right\}, \qquad m_k := (n \bmod 2) + 2k, \qquad 0 \leq k < N := \lfloor n/2 \rfloor + 1.$$

A single-spin flip changes the signed magnetization by ± 2. Thus distinct neighboring orbits are consecutive, and every consecutive pair is connected. The quotient graph is the path $\{0, \ldots, N-1\}$, with projected probabilities and capacities

$$\overline{\pi}_s(k) := \sum_{x\in A_k} \pi_s(x), \qquad \overline{c}_s(k,\ell) := \sum_{x\in A_k} \sum_{y\in A_\ell} c_s(x,y), \quad k \neq \ell.$$

We use the following properties established by He et al. (2026).

Lemma 4 (**Properties of the projected Ising chain** (He et al., 2026, Lemma 5.5 and Eq. (5))). *For every $s \in [0,1]$, the following hold:*

1. ***Unimodality.*** *There exists $r_s \in \{0,\dots,N-1\}$ such that $\overline{\pi}_s(k)$ is nondecreasing for $0 \le k \le r_s$ and nonincreasing for $r_s \le k \le N-1$.*
2. ***Capacity lower bound.*** *For every $0 \le k \le N-2$,*

$$\overline{c}_s(k,k+1) \ge \frac{1}{2n}\min\{\overline{\pi}_s(k), \overline{\pi}_s(k+1)\}. \tag{16}$$

Fact 6 (**Projected score bound**). *Set $B_s := n|\beta_s'|/2$. Then*

$$|\partial_s \overline{\pi}_s(k)| \le B_s \overline{\pi}_s(k), \qquad 0 \le k < N.$$

Proof. The orbit sizes and projected probabilities are

$$a_k := |A_k| = \big(2 - \mathbf{1}_{\{m_k=0\}}\big)\binom{n}{(n+m_k)/2}, \qquad \overline{\pi}_s(k) = \frac{a_k e^{\beta_s m_k^2/(2n)}}{Z_s},$$

where $Z_s := \sum_{\ell=0}^{N-1} a_\ell e^{\beta_s m_\ell^2/(2n)}$. Indeed, $\sum_{i<j} x_i x_j = ((\sum_i x_i)^2 - n)/2$, and the factor $e^{-\beta_s/2}$ cancels in the normalization. Differentiating gives

$$\begin{aligned}\partial_s \log \overline{\pi}_s(k) &= \frac{\beta_s' m_k^2}{2n} - \frac{\sum_{\ell=0}^{N-1} a_\ell e^{\beta_s m_\ell^2/(2n)} \beta_s' m_\ell^2/(2n)}{Z_s} \\ &= \frac{\beta_s'}{2n}\left(m_k^2 - \sum_{\ell=0}^{N-1} \overline{\pi}_s(\ell) m_\ell^2\right).\end{aligned}$$

Since $0 \le m_\ell^2 \le n^2$,

$$\begin{aligned}|\partial_s \overline{\pi}_s(k)| &= \overline{\pi}_s(k)\,|\partial_s \log \overline{\pi}_s(k)| \\ &\le \overline{\pi}_s(k)\frac{|\beta_s'|}{2n}\,n^2 = B_s \overline{\pi}_s(k).\end{aligned}$$

□

Fact 7 (**Projected flux**). *For $0 \le k \le N-2$, define*

$$\overline{J}_s(k,k+1) := -\sum_{\ell=0}^{k} \partial_s \overline{\pi}_s(\ell) = \sum_{\ell=k+1}^{N-1} \partial_s \overline{\pi}_s(\ell), \tag{17}$$

extend it antisymmetrically, and set it to zero outside the edges. Then $\overline{J}$ is continuous and admissible, and

$$|\overline{J}_s(k,k+1)| \le B_s N \min\{\overline{\pi}_s(k), \overline{\pi}_s(k+1)\}. \tag{18}$$

Proof. The two expressions in (17) agree because $\sum_{\ell=0}^{N-1} \partial_s \overline{\pi}_s(\ell) = 0$. Continuity follows from $\overline{\pi} \in C^1$.

Write $j_k := \overline{J}_s(k,k+1)$ and set $j_{-1} = j_{N-1} = 0$. For every $0 \le k < N$, interpreting empty sums as zero,

$$\begin{aligned}\partial_s \overline{\pi}_s(k) + \sum_{\substack{0 \le \ell < N \\ |\ell-k|=1}} \overline{J}_s(k,\ell) &= \partial_s \overline{\pi}_s(k) + j_k - j_{k-1} \\ &= \partial_s \overline{\pi}_s(k) - \sum_{\ell=0}^{k} \partial_s \overline{\pi}_s(\ell) + \sum_{\ell=0}^{k-1} \partial_s \overline{\pi}_s(\ell) = 0.\end{aligned}$$

Thus $\overline{J}$ is admissible.

Fix s and a mode $r = r_s$ from Lemma 4. Using both expressions in (17) and Fact 6,

$$|\overline{J}_s(k,k+1)| \le \min\left\{\sum_{\ell=0}^{k}|\partial_s\overline{\pi}_s(\ell)|,\ \sum_{\ell=k+1}^{N-1}|\partial_s\overline{\pi}_s(\ell)|\right\}$$
$$\le B_s\min\left\{\sum_{\ell=0}^{k}\overline{\pi}_s(\ell),\ \sum_{\ell=k+1}^{N-1}\overline{\pi}_s(\ell)\right\}.$$

Unimodality bounds the left tail when $k < r$ and the right tail when $k \ge r$:

$$\min\left\{\sum_{\ell=0}^{k}\overline{\pi}_s(\ell),\ \sum_{\ell=k+1}^{N-1}\overline{\pi}_s(\ell)\right\} \le \begin{cases}(k+1)\overline{\pi}_s(k), & k<r,\\ (N-k-1)\overline{\pi}_s(k+1), & k\ge r,\end{cases}$$
$$\le N\min\{\overline{\pi}_s(k),\overline{\pi}_s(k+1)\}.$$

The last step uses $\overline{\pi}_s(k) \le \overline{\pi}_s(k+1)$ when $k < r$ and the reverse inequality when $k \ge r$. Combining these bounds proves (18). □

Fact 8 (Projected local transport energy). *The flux of Fact 7 satisfies*

$$\max_{0\le k<N}\overline{G}_s[\overline{J}_s](k) \le 2nB_s^2N^2 \le \frac{n^5}{2}|\beta_s'|^2.$$

Proof. For every projected edge $\{k,\ell\}$, (18) and (16) give

$$\frac{\overline{J}_s(k,\ell)^2}{\overline{c}_s(k,\ell)} \le \frac{B_s^2N^2\min\{\overline{\pi}_s(k),\overline{\pi}_s(\ell)\}^2}{(2n)^{-1}\min\{\overline{\pi}_s(k),\overline{\pi}_s(\ell)\}}$$
$$= 2nB_s^2N^2\min\{\overline{\pi}_s(k),\overline{\pi}_s(\ell)\}.$$

Each vertex has at most two neighbors, so

$$\overline{G}_s[\overline{J}_s](k) = \frac{1}{2\overline{\pi}_s(k)}\sum_{\substack{0\le\ell<N\\|\ell-k|=1}}\frac{\overline{J}_s(k,\ell)^2}{\overline{c}_s(k,\ell)}$$
$$\le nB_s^2N^2\sum_{\substack{0\le\ell<N\\|\ell-k|=1}}\frac{\min\{\overline{\pi}_s(k),\overline{\pi}_s(\ell)\}}{\overline{\pi}_s(k)}$$
$$\le 2nB_s^2N^2 = \frac{n^3N^2}{2}|\beta_s'|^2 \le \frac{n^5}{2}|\beta_s'|^2,$$

where $N \le n$ for $n \ge 2$. □

Proof of Lemma 3. Apply Lemma 2 to the flux from Fact 7. The lifted flux J is admissible, and Fact 8 yields

$$\sup_{x\in\Omega}G_s[J_s](x) = \max_{0\le k<N}\overline{G}_s[\overline{J}_s](k)$$
$$\le 2nB_s^2N^2 \le \frac{n^5}{2}|\beta_s'|^2.$$

This proves (13). Moreover, $\overline{J}_s$ and the capacities are continuous in s, and the quotient capacities are strictly positive on quotient edges. The lifting formula therefore gives a continuous flux J.

For every $\alpha > 0$,

$$\mathfrak{E}_{\alpha,s}(J_s) = \frac{1}{\alpha}\log\sum_{x\in\Omega}\pi_s(x)e^{\alpha G_s[J_s](x)}$$
$$\le \frac{1}{\alpha}\log\left(e^{\alpha n^5|\beta_s'|^2/2}\sum_{x\in\Omega}\pi_s(x)\right) = \frac{n^5}{2}|\beta_s'|^2.$$

Integrating gives

$$\mathcal{A}_\alpha(\pi) \le \mathcal{A}_\alpha(\pi,J) = \int_0^1\mathfrak{E}_{\alpha,s}(J_s)\,ds \le \frac{n^5}{2}\int_0^1|\beta_s'|^2\,ds,$$

which proves (14). □

D Proof of Theorem 3

The proof combines an initialization bound with a pointwise transport-energy estimate for the $(q-1)$-block heat-bath dynamics. The latter yields an α-action bound uniform in $\alpha > 0$, allowing us to apply Theorem 1 and then let $\alpha \to \infty$. We give the main argument first and prove the transport-energy estimate in Section D.2.

D.1 Main argument

Path, dynamics, and initialization. Fix $n \geq q \geq 2$, $\beta \geq q/2$, and $\varepsilon \in (0,1)$. For $x \in \Omega = [q]^n$, write $M_a(x) := |\{i : x_i = a\}|$ and

$$S(x) := \frac{1}{n}\sum_{a=1}^{q} M_a(x)^2, \qquad \mu_b(x) := \frac{e^{bS(x)}}{Z_b}, \qquad Z_b := \sum_{x\in\Omega} e^{bS(x)}.$$

Consider the prescribed path

$$\beta_0 := n\log q + \log(6/\varepsilon), \qquad \beta_s := (1-s)\beta_0 + s\beta, \qquad \pi_s := \mu_{\beta_s}.$$

Since $\beta_0 \geq q\log 2 > q/2$, we have $\beta_s \geq \min\{\beta_0, \beta\} \geq q/2$ for every $s \in [0,1]$.

Set $k := q-1$. The normalized block heat-bath generator is

$$L_s h(x) = \binom{n}{k}^{-1} \sum_{\substack{I\subseteq[n]\\|I|=k}} \sum_{a\in[q]^I} \pi_s(a \mid x_{-I})\big(h(x^{I\leftarrow a}) - h(x)\big).$$

Its capacities are $c_s(x,y) := \pi_s(x)p_s(x,y)$. All transport energies and actions below refer to this normalization.

Let $\mathcal{M} := \{(a,\dots,a) : a \in [q]\}$ and initialize from $\rho_0 = \mathrm{Unif}(\mathcal{M})$. This requires one uniformly sampled color. Write ρ_t for the law of the process with generator $L_{t/T}$.

Initialization error. At the chosen inverse temperature β_0, the target π_0 places nearly all its mass on the monochromatic configurations $\mathcal{M}$. The following bound quantifies the forward-KL error incurred by initializing uniformly on $\mathcal{M}$.

Fact 9 (Forward-KL initialization bound). *The prescribed initialization satisfies*

$$\mathrm{KL}(\rho_0\|\pi_0) \leq q^{n-1}e^{-\beta_0} = \frac{\varepsilon}{6q}.$$

Proof. For $x \in \mathcal{M}$, we have $S(x) = n$. For $x \notin \mathcal{M}$, choose an occupied color a and set $m := M_a(x) \in \{1,\dots,n-1\}$. Then

$$nS(x) = m^2 + \sum_{b\neq a} M_b(x)^2 \leq m^2 + (n-m)^2 \leq (n-1)^2 + 1,$$

so $S(x) \leq n - 2 + 2/n \leq n-1$. All configurations in $\mathcal{M}$ have the same Gibbs probability, and hence $\rho_0 = \pi_0(\cdot \mid \mathcal{M})$. Therefore,

$$\begin{aligned}
\mathrm{KL}(\rho_0\|\pi_0) &= \log\frac{1}{\pi_0(\mathcal{M})} = \log\frac{Z_{\beta_0}}{qe^{\beta_0 n}} \\
&= \log\left(1 + \frac{\sum_{x\notin\mathcal{M}} e^{\beta_0 S(x)}}{qe^{\beta_0 n}}\right) \\
&\leq \frac{(q^n - q)e^{\beta_0(n-1)}}{qe^{\beta_0 n}} \leq q^{n-1}e^{-\beta_0} = \frac{\varepsilon}{6q}.
\end{aligned}$$

□

Pointwise transport energy. We project onto color-count vectors modulo color permutations and use the edgewise flux estimate from the proof of He et al. (2026, Theorem 5.19). We then reroute each edge flow through a conditional mode of the corresponding block. This gives capacity bounds sufficient to control the local transport energy. The symmetry-lifting lemma transfers this bound to the original configuration space.

Lemma 5 (Local transport energy for mean-field Potts). *Let $n \geq q \geq 2$ and let $\beta. \in C^1([0,1]; [q/2, \infty))$. Set $\pi_s = \mu_{\beta_s}$, and let c_s be the capacities of the normalized $(q-1)$-block heat-bath generator above. There exists a universal constant $C > 0$ and a measurable admissible flux $J \in \mathcal{J}(\pi)$ such that, for almost every $s \in [0,1]$,*

$$\sup_{x \in \Omega} G_s[J_s](x) \leq n^{Cq} |\beta_s'|^2. \tag{19}$$

Consequently, the same flux satisfies, for every $\alpha > 0$,

$$\mathcal{A}_\alpha(\pi) \leq \mathcal{A}_\alpha(\pi, J) \leq n^{Cq} \int_0^1 |\beta_s'|^2 \, ds. \tag{20}$$

The proof is given in Section D.2. For the prescribed linear path, $\beta_s' = \beta - \beta_0$, so

$$\mathcal{A}_\alpha(\pi) \leq \mathcal{A}_\alpha(\pi, J) \leq n^{Cq} (\beta - \beta_0)^2, \qquad \alpha > 0. \tag{21}$$

Completion of the argument. Every frozen block heat-bath chain is reversible and irreducible: any single-coordinate change can be realized by a block update with positive probability. Since the state space is finite and the positive edge rates vary continuously along the compact parameter interval, the generators satisfy Assumption 1 with a common constant $\lambda_* > 0$. No quantitative lower bound on λ_* is needed.

Set $H_0 := \mathrm{KL}(\rho_0 \| \pi_0)$. Applying Theorem 1 and (21), for every $\alpha > 0$ we obtain

$$\begin{aligned} \mathrm{KL}(\rho_T \| \mu_\beta) &\leq e^{-T\lambda_* + \frac{1}{2\alpha T}} H_0 + \frac{e^{1/(2\alpha T)}}{2T} \mathcal{A}_\alpha(\pi) \\ &\leq e^{1/(2\alpha T)} \left(e^{-T\lambda_*} H_0 + \frac{n^{Cq}(\beta - \beta_0)^2}{2T} \right). \end{aligned}$$

For fixed $T > 0$, the law ρ_T is independent of α. Letting $\alpha \to \infty$ and using Fact 9 therefore gives

$$\begin{aligned} \mathrm{KL}(\rho_T \| \mu_\beta) &\leq e^{-T\lambda_*} H_0 + \frac{n^{Cq}(\beta - \beta_0)^2}{2T} \\ &\leq \frac{\varepsilon}{6q} + \frac{n^{Cq}(\beta - \beta_0)^2}{2T}. \end{aligned}$$

Choose

$$T_\varepsilon := \max\left\{ 1, \, \frac{n^{Cq}(\beta - \beta_0)^2}{\varepsilon} \right\}.$$

Then

$$\mathrm{KL}(\rho_{T_\varepsilon} \| \mu_\beta) \leq \frac{\varepsilon}{6q} + \frac{n^{Cq}(\beta - \beta_0)^2}{2T_\varepsilon} \leq \frac{\varepsilon}{6q} + \frac{\varepsilon}{2} < \varepsilon.$$

The block-update version of Algorithm 1 simulates this process exactly: at each event of a rate-one Poisson clock, it chooses a uniformly random block of size $q-1$ and resamples that block from its current conditional distribution. Thus its output has law ρ_T. The number of conditional sampling calls is $N_T \sim \mathrm{Pois}(T)$, including updates that leave the configuration unchanged. The final waiting-time draw that crosses T requires no conditional sampling call. Consequently,

$$\mathbb{E}[\text{number of block conditional sampling calls}] = \mathbb{E}[N_{T_\varepsilon}] = T_\varepsilon = O\left(\frac{n^{Cq} \big(\beta + n \log q + \log(1/\varepsilon)\big)^2}{\varepsilon} \right),$$

where we used $|\beta - \beta_0| \leq \beta + n \log q + \log(6/\varepsilon)$. This proves Theorem 3.

D.2 Proof of Lemma 5

Throughout, let $n \geq q \geq 2$, $k := q-1$, and $\beta_\cdot \in C^1([0,1]; [q/2, \infty))$. Let $\Gamma = S_n \times S_q$ act on $[q]^n$ by coordinate permutations and color permutations. Both $\pi_s = \mu_{\beta_s}$ and the capacities c_s of the normalized block heat-bath generator are Γ-invariant.

Each orbit is identified with a sorted color-count vector, so

$$\overline{\Omega} := \Omega/\Gamma \simeq \left\{ m \in \mathbb{Z}^q_{\geq 0} : m_1 \geq \cdots \geq m_q,\ \sum_{a=1}^{q} m_a = n \right\}.$$

Writing $N := |\overline{\Omega}|$, we have

$$N \leq \binom{n+q-1}{q-1} \leq (n+1)^{q-1} \leq n^{2(q-1)}. \tag{22}$$

For distinct orbits A, B, define

$$\overline{\pi}_s(A) := \sum_{x \in A} \pi_s(x), \qquad \overline{c}_s(A,B) := \sum_{x \in A} \sum_{y \in B} c_s(x,y).$$

Let $\overline{E}$ be the quotient edge set. For an antisymmetric flux F on this graph, write

$$\overline{G}_s[F](A) := \frac{1}{2\overline{\pi}_s(A)} \sum_{B : \{A,B\} \in \overline{E}} \frac{F(A,B)^2}{\overline{c}_s(A,B)}.$$

A projected flux bound. We first recall the edgewise estimate used in the proof of He et al. (2026, Theorem 5.19).

Lemma 6 (**Projected Potts flux** (He et al., 2026, proof of Theorem 5.19))**.** *There exists a measurable admissible flux $\overline{F}$ for $(\overline{\pi}_s)_{s\in[0,1]}$ such that*

$$|\overline{F}_s(A,B)| \leq K_s \min\{\overline{\pi}_s(A), \overline{\pi}_s(B)\}, \qquad K_s := 2Nq!e^{q-1}n|\beta_s'|. \tag{23}$$

Proof. The pointwise construction and estimate are given in the cited proof. To obtain a measurable family, order the finitely many simple paths between each pair of vertices. For each s, select the first path satisfying the mass bounds in that construction. These bounds are inequalities between continuous functions of s, so the selected unit flows are measurable.

For the transport plan, order the vertices and successively match the first remaining source and sink, transferring the minimum of their remaining masses. The source masses are $(-\partial_s \overline{\pi}_s(A))_+$ and the sink masses are $(\partial_s \overline{\pi}_s(A))_+$. This finite procedure uses only measurable comparisons and arithmetic operations. The resulting flux is therefore measurable and retains the stated bound. □

Conditional modes and capacity bounds. Fix one orientation of each edge in $\overline{E}$ and denote the resulting set by $\mathcal{O}$. Set

$$L := |\mathcal{O}| \leq \frac{N(N-1)}{2}, \qquad D := \binom{n}{k} q^k.$$

For each $e = (A,B) \in \mathcal{O}$, fix configurations $x^e \in A$, $y^e \in B$, and a block $I_e \subseteq [n]$ of size k such that $x^e_{-I_e} = y^e_{-I_e}$. Such witnesses exist by the definition of the quotient graph.

Choose

$$z^e \in \operatorname*{arg\,max}_{z:\, z_{-I_e} = x^e_{-I_e}} S(z),$$

using a fixed ordering to break ties, and let C_e be its orbit. Since $\beta_s > 0$, the configuration z^e is a conditional mode for every s. In particular, z^e and C_e are independent of s.

Fact 10 (**Capacities through a conditional mode**)**.** *For every $e = (A,B) \in \mathcal{O}$ and $U \in \{A,B\} \setminus \{C_e\}$,*

$$\frac{\overline{\pi}_s(U)}{D} \leq \overline{c}_s(U, C_e) \leq \overline{\pi}_s(C_e).$$

Proof. Let $u = x^e$ when $U = A$ and $u = y^e$ when $U = B$. There are q^k possible completions of u_{-I_e}. By the definition of z^e,

$$\begin{aligned}\pi_s(z^e_{I_e} \mid u_{-I_e}) &= \frac{e^{\beta_s S(z^e)}}{\sum_{a\in[q]^{I_e}} e^{\beta_s S(u^{I_e\leftarrow a})}} \\ &\geq \frac{e^{\beta_s S(z^e)}}{q^k e^{\beta_s S(z^e)}} = q^{-k}.\end{aligned}$$

The block I_e is selected with probability $\binom{n}{k}^{-1}$, and hence $p_s(u, z^e) \geq D^{-1}$.

By the symmetry row-sum identity (15),

$$\begin{aligned}\overline{c}_s(U, C_e) &= |U| \sum_{w\in C_e} c_s(u, w) \\ &\geq |U|\pi_s(u)p_s(u, z^e) \\ &\geq \frac{|U|\pi_s(u)}{D} = \frac{\overline{\pi}_s(U)}{D}.\end{aligned}$$

For the upper bound, use reversibility and the fact that the total exit rate is at most one:

$$\begin{aligned}\overline{c}_s(U, C_e) &= \overline{c}_s(C_e, U) \\ &= \sum_{z\in C_e} \pi_s(z) \sum_{w\in U} p_s(z, w) \\ &\leq \sum_{z\in C_e} \pi_s(z) = \overline{\pi}_s(C_e).\end{aligned}$$

□

Rerouting the projected flux. For each $e = (A, B) \in \mathcal{O}$, let R^e be the antisymmetric unit flow along $A \to C_e \to B$, with self-loops omitted. Thus the route has one or two edges. Define

$$\overline{J}^e_s := \overline{F}_s(A, B)R^e, \qquad \overline{J}_s := \sum_{e\in\mathcal{O}} \overline{J}^e_s.$$

Fact 11 (Admissibility of the rerouted flux). *The family $\overline{J}$ is measurable and admissible for $(\overline{\pi}_s)_{s\in[0,1]}$.*

Proof. Each R^e is fixed in s, so measurability follows from that of $\overline{F}$. The flux is antisymmetric and supported on $\overline{E}$. For every $U \in \overline{\Omega}$,

$$\sum_V R^e(U, V) = \mathbf{1}_{\{U=A\}} - \mathbf{1}_{\{U=B\}},$$

because the incoming and outgoing flows at C_e cancel. This identity also holds when a self-loop has been omitted. Consequently,

$$\begin{aligned}\sum_V \overline{J}_s(U, V) &= \sum_{e=(A,B)\in\mathcal{O}} \overline{F}_s(A, B)\big(\mathbf{1}_{\{U=A\}} - \mathbf{1}_{\{U=B\}}\big) \\ &= \sum_V \overline{F}_s(U, V) = -\partial_s\overline{\pi}_s(U).\end{aligned}$$

Thus the projected continuity equation holds. □

Fact 12 (Local energy of the rerouted flux). *For every $s \in [0, 1]$,*

$$\sup_{U\in\overline{\Omega}} \overline{G}_s[\overline{J}_s](U) \leq L^2K_s^2D^2.$$

Proof. Fix $e = (A, B)$ and write $f_e := \overline{F}_s(A, B)$. Every retained route edge has the form $\{U, C_e\}$ with $U \in \{A, B\} \setminus \{C_e\}$. By (23) and Fact 10,

$$|f_e| \le K_s \overline{\pi}_s(U), \qquad \overline{c}_s(U, C_e) \ge \frac{\overline{\pi}_s(U)}{D}, \qquad \overline{\pi}_s(C_e) \ge \frac{\overline{\pi}_s(U)}{D}.$$

Since $D \ge 1$, either endpoint $V \in \{U, C_e\}$ satisfies $\overline{\pi}_s(V) \ge \overline{\pi}_s(U)/D$. The contribution of this edge to the local energy at V is therefore bounded by

$$\frac{f_e^2}{2\overline{\pi}_s(V)\overline{c}_s(U, C_e)} \le \frac{K_s^2 \overline{\pi}_s(U)^2}{2(\overline{\pi}_s(U)/D)(\overline{\pi}_s(U)/D)} = \frac{K_s^2 D^2}{2}.$$

Each vertex is incident to at most two edges of this route, including when $C_e = A$ or $C_e = B$. Hence

$$\sup_V \overline{G}_s[\overline{J}_s^e](V) \le K_s^2 D^2.$$

For the superposition of all L routes, Cauchy–Schwarz gives

$$\begin{aligned}
\overline{G}_s[\overline{J}_s](U) &= \frac{1}{2\overline{\pi}_s(U)} \sum_{V:\{U,V\}\in\overline{E}} \frac{\left(\sum_{e\in\mathcal{O}} \overline{J}_s^e(U, V)\right)^2}{\overline{c}_s(U, V)} \\
&\le \frac{L}{2\overline{\pi}_s(U)} \sum_{V:\{U,V\}\in\overline{E}} \sum_{e\in\mathcal{O}} \frac{\overline{J}_s^e(U, V)^2}{\overline{c}_s(U, V)} \\
&= L \sum_{e\in\mathcal{O}} \overline{G}_s[\overline{J}_s^e](U) \\
&\le L^2 K_s^2 D^2.
\end{aligned}$$

□

Proof of Lemma 5. By Fact 11, the projected flux $\overline{J}$ is measurable and admissible. Moreover, Fact 12 gives

$$\begin{aligned}
\sup_U \overline{G}_s[\overline{J}_s](U) &\le L^2 K_s^2 D^2 \\
&\le \frac{N^4}{4} \left(2Nq! e^{q-1} n |\beta_s'|\right)^2 D^2 \\
&= N^6 (q!)^2 e^{2(q-1)} n^2 D^2 |\beta_s'|^2.
\end{aligned}$$

For $n \ge q \ge 2$,

$$D = \binom{n}{q-1} q^{q-1} \le n^{2(q-1)}, \qquad q! \le n^q, \qquad e^{2(q-1)} \le n^{4(q-1)}.$$

Together with (22), these bounds imply

$$\begin{aligned}
N^6 (q!)^2 e^{2(q-1)} n^2 D^2 &\le n^{12(q-1)+2q+4(q-1)+2+4(q-1)} \\
&= n^{22q-18} \le n^{22q}.
\end{aligned}$$

Thus a universal constant C suffices.

Lift $\overline{J}$ using Lemma 2. The resulting flux J is measurable and admissible, and its local energy satisfies

$$\sup_{x\in\Omega} G_s[J_s](x) = \max_{U\in\overline{\Omega}} \overline{G}_s[\overline{J}_s](U) \le n^{Cq} |\beta_s'|^2.$$

This proves (19).

Finally, for every $\alpha > 0$,

$$\begin{aligned}
\mathfrak{E}_{\alpha,s}(J_s) &= \frac{1}{\alpha} \log \sum_{x\in\Omega} \pi_s(x) e^{\alpha G_s[J_s](x)} \\
&\le \frac{1}{\alpha} \log \left(e^{\alpha n^{Cq} |\beta_s'|^2} \sum_{x\in\Omega} \pi_s(x) \right) = n^{Cq} |\beta_s'|^2.
\end{aligned}$$

Integrating yields

$$\mathcal{A}_\alpha(\pi) \le \mathcal{A}_\alpha(\pi, J) = \int_0^1 \mathfrak{E}_{\alpha,s}(J_s)\, ds \le n^{Cq} \int_0^1 |\beta_s'|^2\, ds,$$

which proves (20). □

E Proofs for Section 4

Here, we provide the formal statement and mathematical proof of the propositions discussed in the Section 4.

We consider the MDM fine-tuning:

Setup Consider the case where the pretrained MDM $\rho_1(x)$ is given, and it is adequately fine-tuned via respective tilt distribution method

$$\rho_{1,a}(x) \propto \rho_1(x) e^{ar(x)}, \tag{24}$$

where $r(x)$ is a reward function. This is a common setting that is available in various reward fine-tuning text (refer to Wang et al. (2025); So et al. (2026); Chen et al. (2026)). The MDM generation method follows the standard MDM inference procedure where the rate matrix is characterized by

$$R_\tau(x, x[x^i \leftarrow v]) = g(\tau) \cdot \mathbb{P}[x_1^i = v \mid x_\tau = x].$$

When the control as in equation (24) is applied, the respective rate matrix is adjusted by

$$R_{\tau,a}(x, x[x^i \leftarrow v]) = g(\tau) \cdot \mathbb{P}_a[x_1^i = v \mid x_\tau = x_{\tau,a}]$$

where the discrete stochastic interpolant $x_{\tau,a}$ is defined with $\alpha(\tau) = \mathbb{P}[T^i < \tau]$ as

$$x_{\tau,a}^i := \begin{cases} x_1^i, & \tau \ge T^i, \\ \mathbf{m}, & \tau < T^i, \end{cases} \quad T^i \sim \dot{\alpha}(\tau)\, d\tau, \quad x_1 \sim \rho_{1,a}.$$

The posterior probability of the unmasked tokens are given by

$$\pi_a(v \mid x_\tau, i) := \mathbb{P}_a[x_1^i = v \mid x_\tau] = \mathbb{E}_a[\mathbf{1}_{\{x_1^i = v\}} \mid x_\tau]. \tag{25}$$

Before any proof on MDM fine-tuning, we recall the mLSI in general distribution and the path mLSI:

Assumption 2 (Modified log-Sobolev inequality over annealing path). *The distribution π is told to satisfy the modified log-Sobolev inequality with a constant λ_π when there exists $\lambda_\pi > 0$ such that*

$$\mathcal{I}_\pi(\rho \| \pi) \;\ge\; 2\lambda_\pi \,\mathrm{KL}(\rho \| \pi) \qquad \textit{for all } \rho \in \mathcal{P}_{>0}(\Omega).$$

We assume the path mLSI across the annealing path $(\rho_{1,a(t)})_{t\in[0,T]}$ at all setting in the proof.

E.1 Slowly-Annealed Discrete Tilt Matching

In Discrete Tilt Matching (DTM), the goal is to finetune a generative model such that the model now generates the samples from the tilted distribution in Equation (24). For this, the ideal target objective is defined as

$$\min_\theta \mathbb{E}_{a+h}\Big[\sum_{i: x_\tau^i = \mathbf{m}} \mathrm{CE}(\pi_{a+h}(\cdot \mid x_\tau, i) \parallel \pi_\theta(\cdot \mid x_\tau, i)) \Big],$$

while this is intractable to generate samples from $\pi_{a+h}(\cdot \mid x_\tau, i)$ where $\mathbb{E}_{a+h}$ represents the expectation with respect to the probability distribution induced by $x_{a+h,A}^i$. To address this, DTM introduces various reformulation techniques such as Radon-Nikodym Derivative, Esscher Transform to represent the posterior probability $\pi_{a+h}(\cdot \mid x_\tau, i)$ in a tractable form. The unmasking probability at time a is given by

$$\mathbb{P}_a(x_1 \mid x_\tau) = \frac{\rho_1(x_1) e^{ar(x_1)} K_\tau(x_\tau \mid x_1)}{\sum_{x_1'} \rho_1(x_1') e^{ar(x_1')} K_\tau(x_\tau \mid x_1')} = \frac{\rho_1(x_1 \mid x_\tau) e^{ar(x_1)}}{\mathbb{E}_0\left[e^{ar(x_1)} \mid x_\tau\right]}, \tag{26}$$

Furthermore, as it is known that the one-hot indicator function $\mathbf{1}\{x_1^i = v\}$ inside the equation is of high variance, the one-hot indicator function is replaced by a prediction network.

$$\begin{aligned}\mathbb{E}_a\left[e^{hr(x_1)}\pi_{a+h}(v \mid x, i) \mid x_\tau\right] &= \mathbb{E}_a\left[e^{hr(x_1)}\mathbf{1}\{x_1^i = v\} \mid x_\tau\right] \\ &= \mathbb{E}_a\left[e^{hr(x_1)}T_c(v \mid x_\tau, i, x_1) \mid x_\tau\right] \\ &\left(e.g. = \mathbb{E}_a\left[c\pi_a(v \mid x_\tau, i) + \left(e^{hr(x_1)} - c\right)\mathbf{1}\{v = x_1^i\} \;\middle|\; x_\tau\right]\right)\end{aligned}$$

This naturally results in the loss function given by

$$\mathcal{L}^{\text{c-DTM}}_{a\to a+h}(\theta) = \int_0^1 g(\tau)\mathbb{E}_a\left[\sum_{i\in[L],\, x_\tau^i = \mathbf{m}} e^{hr(x_1)} \cdot \mathrm{CE}\big(T_c(\cdot \mid x_\tau, i, x_1) \parallel \pi_\theta(\cdot \mid x_\tau, i)\big)\right] d\tau$$

which the unique minimizer of the loss function gives the ground-truth tilted unmasking posterior which is $\pi_{\theta_*}(\cdot \mid x_\tau, i) = \pi_{a+h}(\cdot \mid x_\tau, i)$.

Assumptions Assume the following for the analysis of the DTM. Let the finite token space be $\mathcal{V}$ and the state space be the union of token space and the masked state of length $L < \infty$ be an integer consisting of the full space of $\mathcal{X} = (\mathcal{V} \times \{m\})^L$.

- **(A1) Bounded Oscillation.** $r : \mathcal{X} \to \mathbb{R}$ with $\mathrm{osc}(r) = \sup_{x,x' \in \mathrm{supp}\,\rho_1} |r(x) - r(x')| < \infty$.
- **(A2) Schedule.** $a \in C^1([0,T];[0,A])$, $\dot{a} \geq 0$, $a(0) = 0$, $a(T) = A$ where $\int_0^T \dot{a}(t)\,dt = A$.
- **(A3) Gradient Lower Bound.** Define a set $\mathcal{R}_t := \{\theta : \mathrm{ess\,inf}_c \min_v \pi_\theta(v \mid c) \geq \delta_{a(t)}/2\}$. There exists λ_0 that satisfy for all $t \in [0,T], \theta \in \mathcal{R}_t$,
$$\lambda_0 \int \nu_s(\mathrm{d}c)\|g_c(\theta)\|_2^2 \leq \left\| \int \nu_s(\mathrm{d}c) J_c(\theta)^T g_c(\theta) \right\|_2^2$$
- **(A4) Nondegeneracy.** $\pi_{\theta(0)}(v \mid x_\tau, i) > 0$ for all $v \in \mathcal{V}$ and all (x_τ, i).
- **(A5) Interpolant Structure.** The unmasking times $T^i \sim \dot{a}(t)\,dt$ are drawn independently of x_1, so the mask set $M(x_\tau) := \{i : x_\tau^i = m\}$ has a law independent of a.
- **(A6) Coverage.** $\log(\pi_{a(t)}/\pi_{\theta(t)}) < \infty$ on $\mathrm{supp}\,\pi_{a(t)}$, uniformly on $[0,T]$.

We let $a = a(t)$ otherwise specified for notational brevity.

Lemma 7 (KL dissipation)**.** *For the context $c = (\tau, x, i)$ with $i \in M(x)$, drawn from the proposal ν_t, let $p_c^a := \pi_a(\cdot \mid c)$ denote the tilted target, $q_{\theta,c} := \pi_\theta(\cdot \mid c)$ denote the model. The loss and Lyapunov functional in the setting is given by*

$$\mathcal{L}_t(\theta) = \int \nu_t(\mathrm{d}c)\,\mathrm{CE}(p_c^a \| q_{\theta,c}), \qquad H(t) = \int \nu_t(\mathrm{d}c)\,\mathrm{KL}(p_c^a \| q_{\theta,c}).$$

*Assume **(A3)** and let $\theta(t) \in \mathcal{R}_t$. Then along the learning dynamics $\dot{\theta} = -\nabla_\theta \mathcal{L}_t$,*

$$\partial_t H(t) \leq -\alpha(t)H(t), \qquad \alpha(t) := \frac{\lambda_0 \delta_{a(t)}}{2}.$$

Proof. By fixing time t, a, ν_t and mass floor δ_a are fixed. The cross entropy is defined as $\mathrm{CE}(p_c^a \| q_{\theta,c}) = \mathrm{KL}(p_c^a \| q_{\theta,c}) + \mathrm{Ent}(p_c^a)$ and where the entropy term is independent of θ. $\nabla_\theta \mathcal{L}_t$ is equivalent to the gradient of the loss and the Lyapunov function which gives the gradient along the path $\dot{\theta} = -\nabla_\theta \mathcal{L}_t$ as

$$\begin{aligned}\partial_t H_t = \langle \nabla_\theta H, \dot{\theta} \rangle = -\|\nabla_\theta \mathcal{L}_t\|_2^2 &= -\left\| \int \nu_t(\mathrm{d}c) J_c^T g_c(\theta) \right\|_2^2 \\ &\leq -\lambda_0 \int \nu_t(\mathrm{d}c)\|g_c(\theta)\|_2^2\end{aligned}$$

where $g_c := q_{\theta,c} - p_c^a$ denote the logistic gradient of $\mathrm{CE}(p_c^a \| q_{\theta,c})$. Then, for $\ell_c = p_c^a / q_{\theta,c}$ being the density ratio, the fact $\pi_c(v)^2 \left(\ell_c(v) - 1\right)^2 = (p_c^a(v) - q_{\theta,c}(v))^2$ gives the inequality

$$\|g_c\|_2^2 = \sum_v q_{\theta,c}(v) q_{\theta,c}(v) \left(\ell_c(v) - 1\right)^2 \geq \left(\min_v q_{\theta,c}(v) \left(\ell_c(v) - 1\right)^2\right) \geq \frac{\delta_{a(t)}}{2} \mathcal{E}_{\pi_c}(\ell_c, \ell_c)$$

where the last step used $\sum_v q_{\theta,c}(v)(\ell_c(v) - 1)^2 = \mathrm{Var}_{q_{\theta,c}}(\ell_c) = \mathcal{E}_{q_{\theta,c}}(\ell_c, \ell_c)$. Then, by modified LSI Assumption 2, we shall obtain

$$\mathcal{E}_{q_{\theta,c}}(\ell_c, \ell_c) = \chi^2(p_c \| q_{\theta,c}) \geq \mathrm{Ent}_{q_{\theta,c}}(\ell_c) = \mathrm{KL}(p_c \| q_{\theta,c}).$$

Together, we conclude with the inequality

$$\partial_t H \leq -\lambda_0 \frac{\delta_a}{2} \int \nu_t(dc) \mathrm{KL}(p_c^a \| q_{\theta,c}) = -\alpha(t) H(t).$$

□

Lemma 8. *Let P and Q be probability measures on a space $\mathcal{X}$ such that $P \ll Q$. Define the log-likelihood ratio $u(x) := \log \frac{dP}{dQ}(x)$. Suppose there exists a constant R such that $u(x) \leq R$ for P-almost all $x \in \mathcal{X}$. Then,*

$$\mathrm{Var}_P(u) \leq \mathbb{E}_P[u^2] \leq \phi(R) D_{\mathrm{KL}}(P \,\|\, Q)$$

where

$$\phi(R) := \frac{R^2}{e^{-R} - 1 + R}.$$

Proof. Consider the function $h(u) := \frac{u^2}{e^{-u} - 1 + u}$ for $u \neq 0$, extended continuously to $h(0) = 2$. Note that the denominator $g(u) := e^{-u} - 1 + u$ is strictly positive for all $u \neq 0$, and $g(0) = 0$. Furthermore, h is monotonically non-decreasing on $\mathbb{R}$.

Given our one-sided bound $u(x) \leq R$, the monotonicity of h yields $h(u) \leq h(R) = \phi(R)$ pointwise. Rearranging this inequality gives:

$$u^2 \leq \phi(R)(e^{-u} - 1 + u).$$

Now, taking the expectation of both sides with respect to P, we obtain two inequalities:

$$\mathbb{E}_P[e^{-u}] = \int_{\mathcal{X}} \frac{\mathrm{d}Q}{\mathrm{d}P} \mathrm{d}P \leq \int_{\mathcal{X}} \mathrm{d}Q \leq 1$$

and

$$\mathbb{E}_P[u] = D_{\mathrm{KL}}(P \,\|\, Q).$$

Applying these bounds to our expectation yields:

$$\mathbb{E}_P[u^2] \leq \phi(R)\big(\mathbb{E}_P[e^{-u}] - 1 + \mathbb{E}_P[u]\big) \leq \phi(R)\big(1 - 1 + D_{\mathrm{KL}}(P \,\|\, Q)\big).$$

Simplifying this expression gives the desired upper bound $\mathbb{E}_P[u^2] \leq \phi(R) D_{\mathrm{KL}}(P \,\|\, Q)$. Finally, since variance is bounded above by the second moment ($\mathrm{Var}_P(u) = \mathbb{E}_P[u^2] - \mathbb{E}_P[u]^2 \leq \mathbb{E}_P[u^2]$), the proof is complete. □

Slowly-Annealed Discrete Tilt Matching To apply a Slowly-Annealed Variant of the DTM, we consider that the annealing schedule determined by $a : [0, T] \to [0, A]$. Across the family of annealed marginal $\{\rho_{1,a}\}_{a \in [0,A]}$, we consider the annealing slowdown, specifically linear case given by $a(t) = tS/T$. In such case, we obtain the $\dot{a}(t) = S/T$ as the shrinking forcing term against the not rescaled dissipation term.

From Equation (26), we have $\partial_a \log \mathbb{P}_a(x_1 \mid x_\tau) = r(x_1) - \mathbb{E}_a[r \mid x_\tau]$, which for arbitrary f that is independent on a,

$$\partial_a \mathbb{E}_a[f \mid x_\tau] = \mathbb{E}_a\left[f \cdot (r - \mathbb{E}_a[r \mid x_\tau]) \mid x_\tau\right] = \mathrm{Cov}_a(f, r \mid x_\tau)$$

which $f = \mathbf{1}_{\{x_1^i = v\}}$, we have

$$\partial_a \log \pi_a(v \mid x_\tau, i) = \mathbb{E}_a\left[r(x_1) \mid x_\tau, x_1^i = v\right] - \mathbb{E}_a\left[r(x_1) \mid x_\tau\right] =: \psi_a(v \mid x_\tau, i) \tag{27}$$

from Equation (25). Subsequently, with respect to the oscillation of the reward defined as $\mathrm{osc}(r) := \sup_{x,x' \in \mathrm{supp}(\rho_1)} \big(r(x) - r(x')\big)$, we have

$$\sup_{v,x_\tau,i} |\psi_a(v \mid x_\tau, i)| \le \mathrm{osc}(r), \qquad \sum_v \pi_a(v \mid x_\tau, i)\psi_a(v \mid x_\tau, i) = 0.$$

For the Fisher-Rao metric $\mathrm{FR}(\phi, \phi') = 2 \arccos\left(\sum_v \sqrt{\phi(v)\phi'(v)}\right)$, by Popoviciu inequality and using the Fisher-Rao speed $|\dot\rho|_t := \lim_{\delta \to 0} \frac{\mathrm{FR}(\rho_{t+\delta}, \rho_t)}{|\delta|}$ (alternative to the W_2 speed in Guo et al. (2025)), we obtain the following lemma

Lemma 9. *The variance across the annealing path* $\mathrm{Var}_{\pi_a}(\psi_a)$ *with* ψ_a *at Equation* (27) *is bounded by*

$$|\dot\pi_a(\cdot \mid x_\tau, i)|_{\mathrm{FR}}^2 \le \mathrm{osc}^2(r).$$

Proof.

$$\dot\pi_a(v) = \pi_a(v \mid x_\tau, i)\, \partial_a \log \pi_a(v \mid x_\tau, i) = \pi_a(v \mid x_\tau, i)\, \psi_a(v \mid x_\tau, i),$$

$$\begin{aligned}
|\dot\pi_a(\cdot \mid x_\tau, i)|_{\mathrm{FR}}^2 &= \sum_{v \in \mathrm{supp}} \frac{\dot\pi_a(v)^2}{\pi_a(v \mid x_\tau, i)} \\
&= \sum_v \pi_a(v \mid x_\tau, i)\, \psi_a(v \mid x_\tau, i)^2 \\
&= \mathbb{E}_{v \sim \pi_a}[\psi_a^2] |\dot\pi_a(\cdot \mid x_\tau, i)|_{\mathrm{FR}}^2 \\
&= \mathrm{Var}_{v \sim \pi_a}[\psi_a(v \mid x_\tau, i)] \le \mathrm{osc}^2(r).
\end{aligned}$$

□

We lastly have the score identities which naturally follow from above

$$\begin{aligned}
\partial_a \log \rho_{\tau,a}(x_\tau) &= \mathbb{E}_a[r \mid x_\tau] - \mathbb{E}_a[r], \qquad |\cdot| \le \mathrm{osc}(r), \\
\partial_s \log \rho_{\tau,a}(x_\tau) &= \dot a(\mathbb{E}_a[r \mid x_\tau] - \mathbb{E}_a[r]), \qquad |\cdot| \le \dot a\, \mathrm{osc}(r).
\end{aligned}$$

The derivative with respect to t follows from each fixed t, $t \mapsto \rho_{\tau,a(t)}$ is a curve on the probability simplex; ∂_s acts only through a, with no ∂_t anywhere. If $a(\cdot)$ is differentiable at t, :

$$\partial_t \log \rho_{\tau,a(t)}(x_\tau) = \dot a(t) \cdot \partial_a \log \rho_{\tau,a}(x_\tau)\Big|_{a=a(t)} = \dot a(s)\left(\mathbb{E}_{a(t)}[r \mid x_\tau] - \mathbb{E}_{a(t)}[r]\right).$$

Now, here is the central proposition for the Slowly-Annealed Discrete Tilt Matching motivated by the theoretical analysis as in Nitanda et al. (2026) as the central proof.

Proof of Theorem 4. Letting $k_t(x_\tau, i) = \mathrm{KL}\left(\pi_a(\cdot \mid x_\tau, i) \,\|\, \pi_\theta(\cdot \mid x_\tau, i)\right)$ for brevity of notation, we can compute the time derivative of the target function

$$\frac{d}{dt} H(t) = \mathbb{E}_{\rho_{\tau,a}}[\partial_t k_t] + \mathbb{E}_{\rho_{\tau,a}}[\partial_t \log \rho_{\tau,a} \cdot k_t].$$

The first term is the gradient flow of KL. We want to find out the alternative expression of

$$\partial_t k_t = \underbrace{\dot a(t)\, \partial_a k_t}_{\text{Flow Contraction}} + \underbrace{\langle \dot\theta(t), \nabla_\theta k_t \rangle}_{\text{Dissipation}}.$$

Dissipation Term By Jensen's inequality, we have $\mathrm{KL} \le \chi^2$. Now, for the function

$$\zeta(u) := \frac{e^u - u - 1}{u^2} = \sum_{m \ge 0} \frac{u^m}{(m+2)!}, \qquad \zeta(0); = \frac{1}{2},$$

Assumption **(A6)** gives an pointwise bound $e^u - 1 \le u + \zeta(R)u^2$. Taking expectation and using $\chi^2 = \mathbb{E}_P\left[e^u\right] - 1$ and $\mathbb{E}[u] = \mathrm{KL}$, we obtain

$$\chi^2 \le \mathrm{KL} + \zeta(R)\mathbb{E}_P[u^2] \le \mathrm{KL} + \zeta(R)\phi(R)\mathrm{KL}.$$

Since $\psi(R)\phi(R) = \frac{e^R - 1 - R}{R^2} \cdot \frac{R^2}{e^{-R} - 1 + R}$, we obtain the lower and upper bound in this case by

$$\mathrm{KL} \le \chi^2 \le \underbrace{1 + \frac{e^R - 1 - R}{e^{-R} - 1 + R}}_{C(R)} \mathrm{KL}\,.$$

We thus obtain the upper bound of the dissipation term using Lemma 7:

$$\langle \dot{\theta}, \nabla_\theta k \rangle = -\sum_v \frac{\pi_a(v)}{\pi_\theta(v)} \dot{\pi}_\theta(v) \le -\alpha(t)\mathrm{KL}(\pi_a \| \pi_\theta) \le -\alpha_\chi(t)\chi^2(\pi_a \| \pi_\theta), \tag{28}$$

where $\alpha_\chi = \alpha(t)/C(R)$.

Flow Contraction Term We can represent the time derivative of k_s with respect to a with $\ell = \frac{\pi_a}{\pi_\theta}$ as

$$\begin{aligned}
\partial_a \underbrace{\mathrm{KL}(\pi_a \parallel \pi_\theta)}_{=k_s} &= \sum_v \big(\partial_a \pi_a(v)\big) \log \ell(v) + \underbrace{\sum_v \pi_a(v)\psi_a(v)}_{=0} \\
&= \mathbb{E}_{\pi_a}[\psi_a \log \ell] - \mathbb{E}_{\pi_a}[\psi_a]\mathbb{E}_{\pi_a}[\log \ell] \\
&= \mathrm{Cov}_{\pi_a}(\psi_a, \log \ell)
\end{aligned}$$

Now, under regularity condition ($\pi_\theta(v) > 0$ for $\forall v \in \mathrm{supp}(\pi_a)$) which gives a simple lemma $\log(\ell) \le \chi^2(\pi_a \parallel \pi_\theta)$, the target itself moves by annealing, which application of Young's inequality ($xy \le \frac{\epsilon}{2}x^2 + \frac{1}{2\epsilon}y^2$, $\epsilon = \alpha_\chi$) yields

$$\begin{aligned}
\dot{a}\ \mathrm{Cov}_{\pi_a}(\psi_a, \log(\ell)) \ &\le\ \dot{a}\sqrt{\mathrm{Var}_{\pi_a}(\psi_a)}\sqrt{\log(\ell)} \\
&\le\ \dot{a}\sqrt{\mathrm{Var}_{\pi_a}(\psi_a)}\sqrt{\chi^2(\pi_a \parallel \pi_\theta)} \\
&\le\ \frac{\dot{a}^2}{2\alpha_\chi}\mathrm{Var}_{\pi_a}(\psi_a) + \frac{\alpha_\chi}{2}\chi^2.
\end{aligned} \tag{29}$$

Then, adding Equation (28) and (29), we obtain the contraction due to the first term as

$$\partial k_t \ \le\ -\frac{\alpha_\chi}{2}\, k_t \ +\ \frac{\dot{a}^2}{2\alpha_\chi} \underbrace{\mathrm{Var}_{\pi_a}(\psi_a)}_{\le \mathrm{osc}^2(r)}.$$

Marginal Evolution Subsequently, the covariance matrix property above combined with the trivial fact $\mathrm{KL}(\cdot\|\cdot) \ge 0$ gives

$$\mathbb{E}_{\rho_{\tau,a}}[\partial_t \log \rho_{\tau,a} \cdot k_t] = \mathrm{Cov}_{\rho_{\tau,a}}(\partial_t \log \rho_{\tau,a}, k_t) \le \dot{a}\,\mathrm{osc}(r) \cdot \mathbb{E}_{\rho_{\tau,a}}[k_t] = \dot{a}\,\mathrm{osc}(r)H.$$

Combining the two with taking the expectation, we obtain the final convergence analysis given by

$$\frac{d}{dt}H(t) \le -\left(\frac{\alpha_\chi}{2} - \dot{a}\,\mathrm{osc}(r)\right) H(t) + \frac{\dot{a}^2}{2\alpha_\chi}\,\mathrm{osc}^2(r).$$

Therefore, when α_χ and $\mathrm{osc}(r)$ are fixed and positive, lowering $\dot{a}$ allows improved convergence rate of the KL divergence between the annealing path and the target. Applying Grönwall gives the desired result. □

E.2 Slowly-Annealed Discrete Adjoint Matching

While DTM is natively annealing-based algorithm to fine-tune a base model, Discrete Adjoint Matching (DAM) is based on the one-shot optimization matching the parametrized rate directly against the full-tilt optimal rate $u^*(\cdot)$. Similar to the DTM, its goal is to fine-tune such that it generates sample from the tilted distribution

$$\rho^\star(\boldsymbol{X}) \equiv \rho^{u^\star}(\boldsymbol{X}) \propto \rho^{\text{base}}(\boldsymbol{X}) e^{Ar(X_1)}$$

DAM object is an estimator of the exponential value difference formulated as a stochastic optimal control problem. The adjoint objective contains the representation of the exponential value difference under the optimal distribution given by

$$u^\star_\tau(y,x) = u^{\text{base}}_\tau(y,x)\, \mathbb{E}_{\mathbb{P}^\star(\boldsymbol{X}|X_\tau=x)} \left[\tilde{a}_\tau(y;\boldsymbol{X}) = \tilde{a}_1(y;\boldsymbol{X}) + \int_\tau^1 \sum_z u^{\text{base}}_\sigma(z,y)\, \tilde{a}_\sigma(z;\boldsymbol{X})\, \mathrm{d}\sigma \right],$$

with $\tilde{a}_1(y;\boldsymbol{X}) = e^{r(y)-r(X_1)}$ as the initial adjoint[1] Subsequently, idealized DAM objective matches the Bregman divergence

$$u^\star_\tau(y,x) = \arg\min_u \mathbb{E}_{\rho^\star(\boldsymbol{X}|x_\tau=x)} \left[\mathbb{D}_\phi \left(u^\theta_\tau(y,x), u^{\text{base}}_\tau(y,x) \tilde{a}_\tau(y;\boldsymbol{X}) \right) \right] \tag{30}$$

for a convex function ϕ.

Now, suppose we are fine-tuning the neural network to train u^θ such that it progressively anneals for sampling with from u^{base}_t to $u^\star_\tau$. Without annealing, the DAM objective is to minimize the approximate divergence

$$\mathcal{L}_{\text{DAM}}(\theta) := \frac{1}{p^{\bar{u}}_\tau(Y \mid x_\tau)} \cdot \mathbb{D}_{\text{gKL}} \left(u^\theta_\tau(Y, x_\tau), \left(u^{\text{base}}_\tau \hat{a}_\tau \right)(Y, x_\tau) \right)$$

where the discrete adjoint $\hat{a}_\tau$ is approximated by importance sampling

$$\hat{a}_\tau\big(y; Z, \{X_1^{(k)}\}\big) := \frac{\rho^{\text{base}}_{1|\tau}(Z \mid y)}{\rho^u_{1|\tau}(Z \mid y)} e^{Ar(Z)} \cdot \left(\frac{1}{K} \sum_{k=1}^K \frac{\rho^{\text{base}}_{1|\tau}\big(X_1^{(k)} \mid x\big)}{\rho^u_{1|\tau}\big(X_1^{(k)} \mid x\big)} e^{Ar(X_1^{(k)})} \right)^{-1}. \tag{31}$$

Now, we consider the annealing of DAM by $a : [0,T] \to [0,A]$ such that the target progressively moves following the family of marginals $\{\rho^{\text{base}}(\boldsymbol{X}) e^{ar(X_1)}\}_{a \in [0,A]}$.

While annealing is not directly considered in the DAM, the convergence of DAM will be irrelevant to the annealing schedule in vanilla DAM. However, considering that the discrete adjoint is approximated by importance sampling in Equation (31), it is possible to consider discrete adjoint approximation as an Annealed Importance Sampling (AIS) over the time horizon. Annealed-DAM is AIS on the geometric path in which the fine-tuned model plays the role of the transition kernel.

Assumptions Assume the following for the analysis of the DAM. Let the finite token space be $\mathcal{V}$ and the state space be the union of token space and the masked state of length $L < \infty$ be an integer consisting of the full space of $\mathcal{X} = (\mathcal{V} \times \{m\})^L$.

- **(A1) Bounded Oscillation.** $r : \mathcal{X} \to \mathbb{R}$ with $\text{osc}(r) = \sup_{x,x' \in \text{supp}\,\rho_1} |r(x) - r(x')| < \infty$.
- **(A2) Shared Support.** For all $\tau \in [0,1]$, $x \in \mathcal{X}$ and $u \in u_j{}_{j=0}^N \cup u^{\text{base}}$,

$$\text{supp}\big(\rho^u_{1|\tau}(\cdot \mid x)\big) = \text{supp}\big(\rho^{\text{base}}_{1|\tau}(\cdot \mid x)\big).$$

- **(A3) Oracle Regularity.** Given exact CTMC simulation (no discretization/τ-leaping error), we have $X_1^{(k)} \sim \rho^{u_j}_{1|\tau}(\cdot \mid x)$ being an exact i.i.d. and $\frac{\rho^{\text{base}}_{1|\tau}(v|x)}{\rho^{u_j}_{1|\tau}(v|x)}$ computable in a closed for for all $v \in \text{supp}(\rho)$.

[1] To distinguish the adjoint and the scale of the reward, I denote the adjoint as a with tilde $\tilde{a}_{(\cdot)}$ and the scale without tilde.

(A4) Annealing Schedule. Fix horizon $A > 0$ and stage count $N \in \mathbb{N}$. Set

$$h := \frac{A}{N}, \qquad a_j := jh \ \ (j = 0, \dots, N),$$

and require

$$h \cdot \operatorname{osc}(r) \le 1 \qquad \Longleftrightarrow \qquad N \ge A \cdot \operatorname{osc}(r).$$

(A5) Perfect Tracking. For each stage $j = 1, \dots, N$, the model used to generate stage-j rollouts exactly matches the previous stage's target law:

$$\rho^{u_j}_{1|\tau}(\cdot \mid x_\tau) = \rho_{1,a_{j-1}}(\cdot \mid x_\tau) \qquad \forall \tau, x_\tau.$$

Or alternatively, for some $\delta \ge 0$, we have

$$\left| \log \frac{d\rho_{1,a_{j-1}}(\cdot \mid x_\tau)}{d\rho^{u_j}_{1|t}(\cdot \mid x_\tau)} \right|_\infty \le \delta.$$

Discrete Adjoint Matching is Exponentially Complex We show the proof for the exponential complexity of vanilla Discrete Adjoint Matching to demonstrate the benefit of annealing in Discrete Adjoint Matching which characterizes the failure mode of the DAM under quantifiable sample size.

Proposition 1. *Let* $\operatorname{osc}(r)$ *be the oscillation where* r *is reward function of the tilt.*
(i) Sharp upper bound. *For every* $A > 0$*,*

$$1 + \chi^2_A(x_\tau) \le \exp\left(\frac{A^2 \operatorname{osc}^2(r)}{4}\right).$$

(ii) Hardness instance. *There exists a reward* r *(the one-hot reward* $r = \operatorname{osc}(r) \cdot \mathbf{1}_{\mathcal{G}}, p := \rho_1^{base}(\mathcal{G} \mid x_\tau) = e^{-A \operatorname{osc}(r)}$*) such that*

$$1 + \chi^2_A(x_\tau) \ge \frac{e^{A \operatorname{osc}(r)}}{4} \qquad \textit{for all } A > 0.$$

(iii) Sample complexity lower bound. *For that same instance, let* $\hat{Z}_A := \frac{1}{K} \sum_{k=1}^K e^{Ar(X_1^{(k)})}$ *with* $X_1^{(k)} \overset{iid}{\sim} \rho_1^{base}(\cdot \mid x_\tau)$*, and assume* $A \operatorname{osc}(r) \ge \log 2$*. If*

$$K \le \frac{1}{2} e^{A \operatorname{osc}(r)},$$

then with probability at least $1/2$*,* $|\log \hat{Z}_A - \log Z_A(x_\tau)| \ge \log(3/2)$*. Consequently, achieving relative-log accuracy* $< \log(3/2)$ *with probability* $> 1/2$ *requires*

$$K = \Omega(e^{A \operatorname{osc}(r)}).$$

Proof. Consider the Kish formula

$$\begin{aligned}
\frac{\mathbb{E}[W^2]}{\mathbb{E}[W]^2} &= \frac{\mathbb{E}_0\left[e^{2Ar} \mid x_\tau\right]}{Z_A(x_\tau)^2} \\
&= \mathbb{E}_0\left[\left(\frac{\rho_{1,A}(\cdot \mid x_\tau)}{\rho_1(\cdot \mid x_\tau)}\right)^2 \,\middle|\, x_\tau\right] \\
&= 1 + \chi^2\left(\rho_{1,A}(\cdot \mid x_\tau) \,\|\, \rho_1(\cdot \mid x_\tau)\right) =: 1 + \chi^2_A(x_\tau).
\end{aligned}$$

Considering the conditional generating function $g(a) := \log Z_a(x_\tau)$, we have

$$g'(a) = \bar{r}^a_t(x_\tau) = \mathbb{E}_a[r \mid x_\tau], \qquad g''(a) = \operatorname{Var}_a(r \mid x_\tau).$$

Then, since $g(0) = 0$, the natural follow up using the second-central-difference identity is that

$$\log\left(1 + \chi^2_A(x_\tau)\right) = g(2A) - 2g(A) + g(0) = \int_0^A \int_0^A g''(u+v)\, du\, dv = \int_0^A \int_0^A \operatorname{Var}_{u+v}(r \mid x_\tau)\, du\, dv.$$

Since Popoviciu's inequality gives $\mathrm{Var}_a(r \mid x_\tau) \le \mathrm{osc}^2(r)/4$, we complete (i) by

$$1 + \chi^2_A(x_\tau) \le \exp\left(\frac{A^2 \,\mathrm{osc}^2(r)}{4}\right).$$

To show the lower bound, we first have

$$\log\left(1 + \chi^2_A(x_\tau)\right) = \log \mathbb{E}_{\rho_{1,A}}\left[\frac{d\rho_{1,A}}{d\rho_1}\right] \ge \mathbb{E}_{\rho_{1,A}}\left[\log \frac{d\rho_{1,A}}{d\rho_1}\right] = \mathrm{KL}\left(\rho_{1,A}(\cdot \mid x_\tau) \,\|\, \rho_1(\cdot \mid x_\tau)\right),$$

Take the standard one-hot reward $r = \mathrm{osc}(r) \cdot \mathbf{1}_{\mathcal{G}}$ with conditional probability $p := \rho_1(\mathcal{G} \mid x_\tau)$. Direct computation gives:

$$1 + \chi^2_A = \frac{1 - p + p\, e^{2A\,\mathrm{osc}(r)}}{\left(1 - p + p\, e^{A\,\mathrm{osc}(r)}\right)^2},$$

and with $p = e^{-A\,\mathrm{osc}(r)}$

$$1 + \chi^2_A = \frac{1 - e^{-A\,\mathrm{osc}(r)} + e^{A\,\mathrm{osc}(r)}}{\left(2 - e^{-A\,\mathrm{osc}(r)}\right)^2} \ge \exp\left(A\,\mathrm{osc}(r)\right)$$

which completes the proof for (ii).

Let $N \sim \mathrm{Bin}(K, p)$ count hits in the event $\mathcal{G}$. The estimate $\hat{Z}_A$ then becomes

$$\begin{aligned}
\hat{Z}_A &= \sum_{k=1}^{K} e^{Ar(X_1^{(k)})} = N \cdot e^{A\mathrm{osc}(r)} + (K - N) \cdot 1 \\
&= \frac{N e^{A\,\mathrm{osc}(r)} + K - N}{K} \\
&= 1 + \frac{N}{K}\left(e^{A\,\mathrm{osc}(r)} - 1\right).
\end{aligned}$$

By Bernoulli's inequality, we have

$$\mathbb{P}[N = 0] = (1 - p)^K \ge 1 - Kp.$$

With $p = e^{-A\,\mathrm{osc}(r)}$ and $K \le e^{A\,\mathrm{osc}(r)}/2 = 1/(2p)$, we have $\mathbb{P}[N = 0] \ge 1/2$ which gives

$$Z_A = (1 - p) + p e^{A\,\mathrm{osc}(r)} = 2 - p$$

so

$$|\log \hat{Z}_A - \log Z_A| = \log(2 - p).$$

Lastly, since $A\,\mathrm{osc}(r) \ge \log 2$, we have $p \le 1/2$ and $2 - p \ge 3/2$, therefore

$$\mathbb{P}\left[|\log \hat{Z}_A - \log Z_A| \ge \log(3/2)\right] \ge \mathbb{P}[N = 0] \ge \frac{1}{2}.$$

Hence, any sampling with sample size less than $K \le \frac{1}{2} e^{A\,\mathrm{osc}(r)}$ will incur a relative-log error of greater than $\log(3/2)$ as a failure mode. Equivalently, achieving error below $\log(3/2)$ with probability over $1/2$ requires

$$K > \frac{1}{2} e^{A\,\mathrm{osc}(r)}, \qquad \text{i.e. } K = \Omega\left(e^{A\,\mathrm{osc}(r)}\right).$$

□

Corollary 1 (adjoint hardness)**.** *In the vanilla regime $u = u^{base}$, the one-shot adjoint estimator satisfies $\hat{a}_\tau(y; Z, \{X_1^{(k)}\}) = e^{Ar(Z)}/\hat{Z}_A(x)$. On the event of sample complexity lower bound (probability $\ge 1/2$ when $K \le \frac{1}{2} e^{A\mathrm{osc}(r)}$), every adjoint evaluation sharing x is inflated by the same estimation error $Z_A(x) \ge 3/2$ relative to its true value. (Refer to Equation (31) for the need of estimation Z).*

E.3 Annealed Discrete Adjoint Matching

In contrast, we show that the annealed Discrete Adjoint Matching allows polynomial-order complexity of the sampling. Annealed Discrete Adjoint Matching will follow the loss as

$$\mathcal{L}_a(\theta) = \mathbb{E}\left[\frac{1}{p_\tau^{\bar{u}}(Y \mid x_\tau)}\mathbb{D}_{\mathrm{gKL}}\left(u_\tau^\theta(Y, x_\tau), (u_\tau^{\mathrm{base}}\hat{a}_\tau^a)(Y, x_\tau)\right)\right],$$

which the model is now the AIS regeneration kernel. Under the Annealed DAM case, we have improved efficiency in the sampling over the one-shot DAM. The proof directly computes the sample complexity lower bound of the χ^2 distribution for the accurate estimation of the normalizing constant $\hat{Z}_A$ with certain log accuracy which by Popoviciu's inequality. The exponent of the scale can be interpreted as the total variation of the scaled tilt, which naturally will be complex when the shape is ill-posed and specific value is extremely large to enforce a stronger deviation from the base distribution. On the other hand, we are able to relax this exponential complexity with respect to the $A\,\mathrm{osc}(r)$ as a polynomial complexity by annealing. The following proposition show that the exponentially complex DAM estimator sampling is addressed by introducing annealing with a specific choice of number and required samples per annealing stage.

Proposition 2 (Annealed estimator)**.** *For a fixed target accuracy* $0 \le \varepsilon \le \min\left\{\frac{A\,\mathrm{osc}(r)}{2}, 6\right\}$*, set*

$$h = \frac{2\varepsilon}{A\,\mathrm{osc}^2(r)}, \qquad N = \left\lceil \frac{A^2\,\mathrm{osc}^2(r)}{2\varepsilon} \right\rceil, \qquad K = \left\lceil \frac{384}{\varepsilon} \right\rceil.$$

Then (under assumptions above) the annealed estimator $\widehat{m}(y, x)$ *of the log-control* $m^{\star,A}(y, x)$ *satisfies*

$$\mathbb{E}[(\widehat{m}(y, x) - m^{\star,A}(y, x))^2] < \varepsilon^2, \tag{32}$$

at total rollout cost across the annealed Discrete Adjoint Matching

$$N \cdot K = O\left(\frac{A^2\,\mathrm{osc}^2(r)}{\varepsilon^2}\right)$$

Proof. Let the expected reward (by tilt) is given as $\bar{r}_\tau^a(x_\tau) := \mathbb{E}_a[r \mid x_\tau]$. To compute the normalized-weight second moment as again a conditional χ^2, we have the cumulant shift property of exponential tilting, $\log \mathbb{E}_{a_{j-1}}[e^{\lambda r} \mid x_\tau] = g(a_{j-1} + \lambda) - g(a_{j-1})$. We obtain the second-difference identity

$$\begin{aligned}\log\left(1 + \chi_j^2(x_\tau)\right) &= g(a_{j-1} + 2h) - 2g(a_{j-1} + h) + g(a_{j-1}) \\ &= \int_0^h \int_0^h \mathrm{Var}_{a_{j-1}+u+v}(r \mid x_\tau)\, du\, dv \\ &\le \frac{h^2\,\mathrm{osc}^2}{4}.\end{aligned} \tag{33}$$

Consequently $\chi_j^2 \le e^{h^2\,\mathrm{osc}^2/4} - 1 \le \frac{h^2\,\mathrm{osc}^2}{2}$ whenever $h\,\mathrm{osc} \le 1$. This allows the establishment of improved convergence of

$$\sum_{j=1}^N \log\left(1 + \chi_j^2\right) \le N \cdot \frac{h^2\,\mathrm{osc}^2}{4} = \frac{Ah\,\mathrm{osc}^2}{4} \xrightarrow[h\to 0]{} 0,$$

compared to the $\mathcal{O}\left(\frac{A^2\,\mathrm{osc}^2(r)}{4}\right)$ in the one-shot case.

Because masking times are independent of token values, $h_\tau^a(x) := \mathbb{E}_0[e^{ar(x_1)} \mid x_\tau = x]$ depends on x only through its revealed content, so $\partial_t h_t^a = 0$, and

$$h^a(x[x^i \leftarrow v]) = \mathbb{E}_0[e^{ar(x_1)} \mid x_\tau = x, x_1^i = v].$$

Consider the optimal tilt-rate is given by $u_\tau^{\star,a}(y, x) = u_\tau^{\mathrm{base}}(y, x)h^a(y)/h^a(x)$. The log-control is defined as

$$m^{\star,A}(y, x) := \log \frac{u_\tau^{\star,A}(y, x)}{u_\tau^{\mathrm{base}}(y, x)} = \log h^A(y) - \log h^A(x),$$

which gives the velocity identity below: since $\partial_a \log h^a = \partial_a g = \bar{r}^a_\tau$ and $\log h^0 = \log Z_0 = 0$,

$$\log h^A(x) = \int_0^A \bar{r}^a_\tau(x)\, da \implies m^{\star,A}(y,x) = \int_0^A [\bar{r}^a_\tau(y) - \bar{r}^a_\tau(x)]\, da$$

Assume perfect-tracking reference coupling $\rho^{u_j}_{1|\tau}(\cdot \mid x_\tau) = \rho_{1,a_{j-1}}(\cdot \mid x_\tau)$. We consider the self-normalized importance sampling (SNIS)-estimated velocity-form estimand batch per stage

$$\hat{\bar{r}}^{a_j}(x) = \sum_{k=1}^K \tilde{w}_k r\big(X_1^{(k)}\big), \qquad \tilde{w}_k = \frac{w_k}{\sum_l w_l}, \qquad w_k \propto \frac{\rho^{\text{base}}_{1|\tau}\big(X_1^{(k)} \mid x\big)}{\rho^{u_j}_{1|\tau}\big(X_1^{(k)} \mid x\big)} e^{a_j r\left(X_1^{(k)}\right)}.$$

Normalize w so that $\mathbb{E}_{\mathbb{Q}}[w] = 1$ where $\mathbb{Q} = \rho_{1,a_{j-1}}(\cdot \mid x)$, and write $\mu := \bar{r}^{a_j}(x)$ so $\mathbb{E}_{\mathbb{Q}}[wr] = \mu$ and $|r - \mu| \le \operatorname{osc}(r)$ pointwise. Define

$$S_K := \frac{1}{K}\sum_k w_k, \qquad G_K := \frac{1}{K}\sum_k w_k\big(r(X_1^{(k)}) - \mu\big), \qquad \text{so that} \qquad (\hat{\bar{r}} - \mu)\, S_K = G_K.$$

We can then rearrange the last identity to give

$$\hat{\bar{r}} - \mu = G_K + (\hat{\bar{r}} - \mu)(1 - S_K), \tag{34}$$

which can be used to derive three momentum-related results below.

1. We first obtain $\mathbb{E}[S_K] = 1$ by the normalization and $\mathbb{E}[(1 - S_K)^2] = \chi^2_j / K$ by:

$$\mathbb{E}\left[(1 - S_K)^2\right] = \operatorname{Var}(S_K) = \frac{1}{K}\operatorname{Var}_{\mathbb{Q}}(w) = \frac{1}{K}\left(\mathbb{E}_{\mathbb{Q}}[w^2] - 1\right) = \frac{\chi^2_j}{K}, \tag{35}$$

2. We have $\mathbb{E}[G_K] = 0$

$$\mathbb{E}_{\mathbb{Q}}[w(r - \mu)] = \mathbb{E}_{\mathbb{Q}}[wr] - \mu\mathbb{E}_{\mathbb{Q}}[w] = \mu - \mu \cdot 1 = 0.$$

3. Lastly, we have $\mathbb{E}[G_K^2] = \operatorname{Var}(G_K) = \frac{1}{K}\operatorname{Var}_{\mathbb{Q}}(w(r-\mu)) = \frac{1}{K}\mathbb{E}_{\mathbb{Q}}\left[w^2(r-\mu)^2\right]$. From the inequality on the pointwise bound, μ and r takes values in an same interval of length $\operatorname{osc}(r)$

$$\mathbb{E}[G_K^2] = \frac{1}{K}\mathbb{E}_{\mathbb{Q}}\left[w^2(r-\mu)^2\right] \le \frac{\operatorname{osc}^2(r)}{K}\mathbb{E}_{\mathbb{Q}}[w^2] = \frac{\operatorname{osc}^2(r)(1+\chi^2_j)}{K}. \tag{36}$$

Let $a = G_K$ and $b = (\hat{\bar{r}} - \mu)(1 - S_K)$, and applying the inequality $(a+b)^2 \le 2a^2 + 2b^2$, from Equation (34), we obtain

$$\begin{aligned}(\hat{\bar{r}} - \mu)^2 &= \big(G_K + (\hat{\bar{r}} - \mu)(1 - S_K)\big)^2 \\ &\le 2G_K^2 + 2(\hat{\bar{r}} - \mu)^2(1 - S_K)^2.\end{aligned}$$

Taking expectation on both sides, and using the fact that $|\hat{\bar{r}} - \mu| \le \operatorname{osc}(r)$ almost surely,

$$\mathbb{E}\left[(\hat{\bar{r}} - \mu)^2(1 - S_K)^2\right] \le \operatorname{osc}(r)^2\mathbb{E}\left[(1 - S_K)^2\right] = \frac{\operatorname{osc}(r)^2\chi^2_j}{K}. \tag{37}$$

Combining the results,

$$\mathbb{E}\left[\left(\hat{\bar{r}}^{a_j}(x) - \underbrace{\bar{r}^{a_j}(x)}_{=\mu}\right)^2\right] \le \underbrace{\frac{2\operatorname{osc}(r)^2\left(1+\chi^2_j\right)}{K}}_{\text{eq. (36)}} + \underbrace{\frac{2\operatorname{osc}^2(r)\chi^2_j}{K}}_{\text{eq. (37)}} \le \frac{4\operatorname{osc}^2(r)\left(1+\chi^2_j\right)}{K}.$$

Let $1 + \chi^2_j \le \exp\left(\frac{h^2\operatorname{osc}^2(r)}{4}\right)$ be given by Equation (33). We have $h\operatorname{osc}(r) \le 1$ since $h = A/N$ and $N \le A\operatorname{osc}(r)$, therefore

$$\operatorname{osc}^2(r)\left(1+\chi^2_j\right) \le \operatorname{osc}^2(r)\exp\left(\frac{h^2\operatorname{osc}^2(r)}{4}\right) \le \operatorname{osc}^2(r)\exp\left(\frac{1}{4}\right).$$

This results in the upper bound of the MSE

$$\mathcal{R} := \mathbb{E}\left[\left(\hat{\bar{r}}^{a_j}(x) - \bar{r}^{a_j}(x)\right)^2\right] \lesssim \frac{\mathrm{osc}^2(r)}{K}. \tag{38}$$

Directly taking expectation of Equation (34), we have

$$\begin{aligned}\mathbb{E}[\hat{\bar{r}} - \mu] &= \underbrace{\mathbb{E}[G_K]}_{=0} + \mathbb{E}\left[(\hat{\bar{r}} - \mu)(1 - S_K)\right] \\ &\leq \sqrt{\mathbb{E}\left[(\hat{\bar{r}} - \mu)^2\right]} \cdot \sqrt{\mathbb{E}\left[(1 - S_K)^2\right]} = \sqrt{\mathcal{R}} \cdot \frac{\chi_j}{\sqrt{K}}\end{aligned} \tag{39}$$

using Cauchy-Schwartz inequality. On $[0, 1]$, the inequality from convexity $e^u - 1 \leq (e-1)u \leq 2u$ follows $\chi_j^2 \leq e^{h^2\,\mathrm{osc}^2(r)/4} - 1 \leq \frac{h^2\,\mathrm{osc}^2(r)}{2}$ which implies $\chi_j \leq \frac{h\,\mathrm{osc}(r)}{\sqrt{2}}$ and we end up with the upper bound of the estimation error using Equation (39)

$$\left|\mathbb{E}[\hat{\bar{r}} - \mu]\right| \lesssim \underbrace{\frac{h\,\mathrm{osc}(r)}{\sqrt{K}}}_{\text{eq. (35)}} \cdot \underbrace{\frac{\mathrm{osc}(r)}{\sqrt{K}}}_{\text{eq. (38)}} = \mathcal{O}\left(\frac{h\,\mathrm{osc}^2(r)}{K}\right).$$

Define $\mathbb{P}_2$ measure (the tilt of $\mathbb{Q}$ via $w^2 \propto e^{2hr}$) as

$$\mathrm{d}\mathbb{P}_2 := \frac{w^2}{\mathbb{E}_{\mathbb{Q}}[w^2]}\mathrm{d}\mathbb{Q}.$$

From the pointwise bound of $|r - \mu| \leq \mathrm{osc}(r)$ and Equation (36),

$$\begin{aligned}\mathbb{E}_{\mathbb{Q}}\left[w^2(r-\mu)^2\right] &= (1 + \chi_j^2)\,\mathbb{E}_{\mathbb{P}_2}\left[(r-\mu)^2\right] \\ &= (1 + \chi_j^2)\left[\mathrm{Var}_{a_j+h}(r) + \left(\bar{r}^{a_j+h} - \bar{r}^{a_j}\right)^2\right].\end{aligned}$$

Then, Popoviciu gives $\mathrm{Var}_{a_j+h}(r) \leq \mathrm{osc}^2(r)/4$ and the combined with velocity Lipschitz bound also gives

$$|\partial_a \bar{r}^a| = \mathrm{Var}_a(r) \leq \mathrm{osc}^2(r)/4.$$

Integrating this derivative gives

$$|\bar{r}^{a_j+h} - \bar{r}^{a_j}| \leq \frac{h\,\mathrm{osc}^2(r)}{4}$$

therefore

$$\mathbb{E}_{\mathbb{P}_2}\left[(r-\mu)^2\right] \leq \frac{\mathrm{osc}^2(r)}{4} + \frac{h^2\,\mathrm{osc}^4(r)}{16} \leq \frac{5\,\mathrm{osc}^2(r)}{16}.$$

MSE Analysis via Log Control Recall the log control

$$\begin{aligned}m^{\star,A}(y,x) :&= \log\frac{u_\tau^{\star,A}(y,x)}{u_\tau^{\mathrm{base}}(y,x)} = \log h^A(y) - \log h^A(x) \\ &= \int_0^A \left[\bar{r}_\tau^a(y) - \bar{r}_\tau^a(x)\right] da\end{aligned}$$

where the last line comes from $\log h^A(x) = \int_0^A \bar{r}_\tau^a(x)\,da$. We analyze the MSE via three different channels:

1. **Riemann Sum**: Velocity Lipschitz $|\partial_a \bar{r}^a| = \mathrm{Var}_a(r) \leq \mathrm{osc}^2(r)/4$. Left-endpoint error on one subinterval for an L-Lipschitz integrand is $\leq Lh^2/2$; two nodes, N subintervals:
$$|m_N - m^{\star,A}| \leq 2 \cdot N \cdot \frac{h^2}{2} \cdot \frac{\mathrm{osc}^2(r)}{4} = \frac{Ah\,\mathrm{osc}^2(r)}{4}.$$
2. **Bias**: Per stage per node, $|\mathbb{E}[\hat{\bar{r}}^{a_j}] - \bar{r}^{a_j}| \leq 2h\,\mathrm{osc}^2(r)/K$. These are deterministic offsets, so they add in absolute value — quadrature weight h, two nodes, N stages, $Nh = A$:
$$|\mathbb{E}[\widehat{m}] - m_N| \leq h \cdot N \cdot 2 \cdot \frac{2h\,\mathrm{osc}^2(r)}{K} = \frac{4Ah\,\mathrm{osc}^2(r)}{K}.$$

3. **Variance**: Fresh batches are independent across stages allowing addition of variances at each stage. Per stage per node MSE $\leq 6\,\mathrm{osc}^2(r)/K$, and $\mathrm{Var}(X-Y) \leq 2\,\mathrm{Var}\,X + 2\,\mathrm{Var}\,Y$:

$$\mathrm{Var}(\widehat{m}) = h^2 \sum_j \mathrm{Var}\left(\hat{\hat{r}}^{a_j}(y) - \hat{\hat{r}}^{a_j}(x)\right) \leq h^2 N \cdot \frac{24\,\mathrm{osc}^2(r)}{K} = \frac{24Ah\,\mathrm{osc}^2(r)}{K}.$$

We can bound the total error in the log-control

$$\widehat{m} - m^{\star,A} = \underbrace{(\widehat{m} - \mathbb{E}[\widehat{m}])}_{\text{MC Variance}} + \underbrace{(\mathbb{E}[\widehat{m}] - m_N)}_{\text{SNIS Bias}} + \underbrace{(m_N - m^{\star,A})}_{\text{Discretization Error}},$$

and result in

$$\mathbb{E}\left[\left(\widehat{m}(y,x) - m^{\star,A}(y,x)\right)^2\right] \leq \left(\frac{A^2\,\mathrm{osc}^2(r)}{4N} + \frac{4A^2\,\mathrm{osc}^2(r)}{KN}\right)^2 + \frac{24A^2\,\mathrm{osc}^2(r)}{KN}. \tag{40}$$

Now, fix a target $0 < \varepsilon \leq \min\{A\,\mathrm{osc}(r)/2,\ 6\}$ and take

$$h = \frac{2\varepsilon}{A\,\mathrm{osc}^2(r)}, \qquad N = \left\lceil \frac{A^2\,\mathrm{osc}^2(r)}{2\varepsilon} \right\rceil, \qquad K = \left\lceil \frac{384}{\varepsilon} \right\rceil.$$

We then have for each term of the Equation (40) as

$$\frac{A^2\,\mathrm{osc}^2(r)}{4N} = \frac{\varepsilon}{2}, \qquad \frac{4A^2\,\mathrm{osc}^2(r)}{KN} = \frac{8\varepsilon}{K} \leq \frac{\varepsilon}{8}, \qquad \frac{24A^2\,\mathrm{osc}^2(r)}{KN} = \frac{48\varepsilon}{K} \leq \frac{\varepsilon^2}{8},$$

resulting in the upper bound

$$\mathbb{E}\left[\left(\widehat{m}(y,x) - m^{\star,A}(y,x)\right)^2\right] < \varepsilon^2.$$

Hence from Definition E.3, we have the complexity of $N \cdot K$ as $\mathcal{O}\left(\frac{A^2\,\mathrm{osc}^2(r)}{\varepsilon^2}\right)$, hence polynomial bound. □

The adjoint matching loss Equation (30) can be re-written as the function of a log-control error $\delta = \log u_\tau^\theta - \log u_\tau$:

$$\mathbb{D}_{\mathrm{gKL}} = u(1 - e^\delta + \delta e^\delta)$$

at a single jump case, which $\delta \to 0$ gives $\mathbb{D}_{\mathrm{gKL}} \approx \frac{u}{2}\delta^2$. Therefore, the MSE bound in the log control bound in Equation (32) suffices to become an estimate of the error in the gKL contribution bound. Especially, $\mathrm{KL}\left(\mathbb{P}^{\star,a(t)} \parallel \mathbb{P}^{\theta(t)}\right)$ is a locally quadratic over log-control error establishing a linear translation between the two bounds giving $\mathbb{D}_{\mathrm{gKL}} \lesssim \frac{u}{2}\varepsilon^2$.

E.4 Slowly-Annealed Discrete Adjoint Matching

Now the goal is to show how slowdown affects the convergence rate of the Annealed Discrete Adjoint Matching.

In contrast to the DTM, DAM is built on the jump-target distribution $Q_\tau(\cdot \mid x)$ rather than the per-site unmasking posterior $\pi_a(\cdot \mid x_\tau, i)$. The relationship between Q^a and π_a is given by

$$Q^{*,a}(y \mid x) = \frac{1}{|M(x)|} \cdot \pi_a(v \mid x, i) = Q^{\mathrm{base}}(y \mid x)\frac{h^a(y)}{h^a(x)},$$

for $h_\tau^a(x) = \mathbb{E}_0[e^{ar(x_1)} \mid x_\tau = x]$, which constructs the DAM's factorization of $u_\tau(y,x) = \lambda_\tau Q_\tau(y \mid x)$. We show that the total DAM loss (over the path space) is a weighted sum of fixed-time KL divergences:

Lemma 10. *For each time-space context $c = (\tau, x_\tau)$, with $u^{(\cdot)} = \lambda^{\mathrm{base}} Q^{(\cdot)}$,*

$$\sum_{y \neq x} D_{\mathrm{gKL}}(u^\theta, u^*)(y,x) = \lambda_\tau^{\mathrm{base}}(x)\mathrm{KL}(Q^\theta \| Q^*), \quad \mathbb{E}_{\mathbb{P}^{\theta,a}}\left[\int_0^1 \lambda_\tau^{\mathrm{base}}(x_\tau)\mathrm{d}t\right] = \mathfrak{N} \qquad (\#\textit{of jump}),$$

and consequently the DAM objective at level a equals the path divergence

$$\mathrm{KL}(\mathbb{P}^\theta \| \mathbb{P}^{*,a}) = \mathbb{E}_{\mathbb{P}^\theta}\left[\int_0^1 \lambda^{\mathrm{base}}\,\mathrm{KL}(Q^\theta \| Q^{*,a})\mathrm{d}t\right].$$

Proof. Summing the Bregman divergence over y, the linear terms cancel because both kernels are normalized such that

$$\sum_y u^\theta = \sum_y u^* = \lambda^{\text{base}}$$

which leaves $\lambda^{\text{base}} \sum_y Q^\theta \log(Q^\theta / Q^{*,a})$ on the generalized KL side. The path-KL identity is the CTMC Girsanov formula for matched initial laws with common intensity. □

Specifically for the Discrete Adjoint Matching, it is possible to characterize the learning dynamics as a mirror-flow.

Lemma 11 (Mirror Flow Descent of Discrete Adjoint Matching)**.** *Define the local DAM loss $\mathcal{L}_c$ as a function of the target jump law $Q_t \equiv Q^{\theta(t)}(\cdot \mid c) \in \mathcal{P}(\mathcal{Y}_c)$ where $c = (\tau, x)$ is the context, $\mathcal{P}(\mathcal{Y}_c)$ denote the probability simplex over jump neighborhood $\mathcal{Y}$ at c,*

$$\mathcal{L}_c(Q) := \mathrm{KL}(Q^* \| Q) = \sum_{y \in \mathcal{Y}_c} Q^*(y) \log \frac{Q^*(y)}{Q(y)}.$$

Note that $\mathcal{L}_c$ can also be interpreted as the Bregman divergence of ϕ from Lemma 10. The canonical choice of the metric is the respective Hessian metric for $\phi(\xi) = \xi \log \xi$ which is

$$g(\xi_1, \xi_2) := \sum_y \frac{\xi_1(y)\xi_2(y)}{Q(y)} \qquad \xi_{(\cdot)} \in T_Q\mathcal{P}(\mathcal{Y}) := \left\{ \psi : \mathcal{Y} \to \mathbb{R} \;\middle|\; \sum_y \psi(y) = 0 \right\}.$$

The gradient descent flow of $\mathcal{L}_c$ under g runs with rate a_0 has the form

$$\dot{Q}_t = -a_0(\nabla_g \mathcal{L}_c(Q_t)) = -a_0(Q_t - Q^*)$$

Proof. Treating Q as a point over the euclidean space,

$$\frac{\partial \mathcal{L}_c}{\partial Q(y)} = \partial_{Q(y)} \left(Q^*(y)[\log Q^*(y) - \log Q(y)] + \sum_{y' \neq y} \cancel{Q^*(y')[\dots]} \right) = -\frac{Q^*(y)}{Q(y)}.$$

By definition of the gradient ∇_g, there exists a unique tangent vector $v \in T_Q\mathcal{P}(\mathcal{Y})$ such that

$$g_Q(v, w) = D_w \mathcal{L}_c := \sum_y \frac{\partial \mathcal{L}_c}{\partial Q(y)} w(y) = -\sum_y \frac{Q^*(y)}{Q(y)} w(y)$$

for every tangent direction $w \in T_Q\mathcal{P}(\mathcal{Y})$. This is equivalent to

$$\left\langle \mathrm{diag}\,(1/Q)\, v + \frac{Q^*}{Q}, w \right\rangle_{\ell^2} = 0 \qquad \text{for all } w \in \mathbf{1}^\perp$$

hence we have $\mathrm{diag}\,(1/Q)\, v + \frac{Q^*}{Q} = \lambda \mathbf{1}$ for some $\lambda \in \mathbb{R}$. Subsequently, applying the inverse metric $\mathrm{diag}(Q)$ on both sides gives

$$v = \mathrm{diag}(Q) \left(\lambda \mathbf{1} - \frac{Q^*}{Q} \right) = \lambda Q - Q^*.$$

We require the tangent vector v to drift in an admissible direction in the probability simplex (i.e. $\sum_y v(y) = 0$). Since $\sum_y Q(y) = \sum_y Q^*(y) = 1$,

$$0 = \sum_y v(y) = \sum_y \left[\lambda Q(y) - Q^*(y) \right] = \lambda - 1$$

Hence, we obtain the value of $\lambda = 1$, which gives

$$v = \nabla_g \mathcal{L}_c(Q) = Q - Q^*,$$

and the natural descent flow at the rate of a_0, given by $\dot{Q}_t := -a_0 \nabla_g \mathcal{L}_c(Q_t)$, takes the form

$$\dot{Q}_t = -a_0(Q_t - Q^*).$$

□

While the proof has been done for forward KL instead of backward loss $\mathrm{KL}(Q^* \| Q)$, the lemma holds in the reverse case by the mirror flow where the flow contracts both orientations simultaneously. Mathematically speaking, the linear relaxation shown at the end gives the same flow contraction for the reverse-cell Lyapunov function.

SA-DAM Analysis with Lyapunov Functional Leveraging the mirror flow descent of Discrete Adjoint Matching, we present the central claim of the benefit derived from the Slowly-Annealed Discrete Adjoint Matching (SA-DAM).

proof of Theorem 5. Write the the path expectation of the time integral at the Lyapunov functional as marginal flow of $\mathbb{P}^{*,a(t)}$ which is deterministic quantity:

$$H(t) = \int_0^1 \sum_x \rho_\tau^{*,a(t)}(x)\lambda_\tau^{\text{base}}(x)k_t(\tau,x).$$

By Lemma 10, the sums are finite and convergent. Fix the time t which gives fixed $a = a(t)$ concurrently. We shall write $\pi := Q^{*,a}(\cdot \mid c)$ and $f_t := Q_t/\pi$. By Lemma 11 at a frozen level $a = a(t)$, the per-context mirror flow is given by $\dot{Q}_t = -a_0(Q_t - \pi_t)$. Therefore, for the generator defined by

$$(\mathcal{A}_{a(t)}g)(y) := \sum_{y'} \pi_t(y')\,[g(y') - g(y)] = \mathbb{E}_{\pi_t}[g] - g(y)$$

which the forward adjoint gives

$$(\mathcal{A}^*_{a(t)}Q)(y) := \sum_{y'} [Q(y')\pi_t(y) - Q(y)\pi_t(y')] = \pi_t(y) - Q(y),$$

the mirror flow returns to the KFE given by $\dot{Q}_t = a_0\mathcal{A}^*_{a(t)}Q_t$.

We can now differentiate the Lyapunov functional and analyze the contributions:

$$\dot{H}(t) = \underbrace{\int_0^1 \sum_x w_t(\tau,x)\dot{k}_t(\tau,x)\,\mathrm{d}\tau}_{(1)+(2)} + \underbrace{\int_0^1 \sum_x \dot{w}_t(\tau,x)k_t(\tau,x)\,\mathrm{d}\tau}_{(3)}$$

and especially the $\dot{k}_t$ can be expanded by freezing t to

$$\dot{k}_t(\tau,x) = \underbrace{\partial_t \mathrm{KL}(Q_t\|Q^{*,a})}_{(1)} + \underbrace{\dot{a}(t)\,\partial_a \mathrm{KL}(Q^\theta\|Q^{*,a})}_{(2)}.$$

Now, the goal is to compute each three terms.

(1) Bounding $\partial_t \mathrm{KL}(Q_t\|Q^{*,a})$ Term Direct computation shows the gradient is given by

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\,\mathrm{KL}(Q_t\|Q^{*,a}) &= \frac{\mathrm{d}}{\mathrm{d}t}\sum_y Q_t(y)\log f_t(y)\\ &= \sum_y \dot{Q}_t(y)\log f_t(y) + \sum_y Q_t(y)\frac{\dot{f}_t(y)}{f_t(y)}\end{aligned}$$

Using the definition $f_t = Q_t/\pi$,

$$\sum_y Q_t(y)\frac{\dot{f}_t(y)}{f_t(y)} = \sum_y f_t\pi(y)\frac{\dot{f}_t(y)}{f_t(y)} = \sum_y \pi(y)\dot{f}_t(y) = \frac{\mathrm{d}}{\mathrm{d}s}\sum_y \underbrace{\pi(y)f_t(y)}_{Q_t} = \frac{\mathrm{d}}{\mathrm{d}t}1 = 0,$$

which eliminates the second term. Then, for the Markov generator $\mathcal{A}^*_{a(t)}$, multiplying both sides by $\log f$ and splitting the two sums,

$$\sum_y (\mathcal{A}^*Q)(y)\log f(y) = \sum_{y,y'} Q(y')p(y',y)\log f(y) - \sum_{y,y'} Q(y)p(y,y')\log f(y). \tag{41}$$

Rewriting both sides, we have the expression of the first and second term as

$$\begin{aligned}\sum_{y,y'} Q(y')p(y',y)\log f(y) &= \sum_{y,y'} f(y')\pi(y')p(y',y)\log f(y)\\ &= \sum_{y,y'} f(y)\pi(y)p(y,y')\log f(y')\\ &= \sum_{y,y'} c(y,y')f(y)\log f(y')\end{aligned}$$

and

$$\sum_{y,y'} Q(y)p(y,y')\log f(y) = \sum_{y,y'} f(y)\pi(y)p(y,y')\log f(y)$$
$$= \sum_{y,y'} c(y,y')f(y)\log f(y).$$

Plugging into Equation (41), we obtain

$$\sum_{y} (\mathcal{A}^* Q)(y)\log f(y) = \sum_{y,y'} c(y,y')f(y)[\log f(y') - \log f(y)] = -\mathcal{I}(f).$$

By the equivalence of the entropy functional and the KL divergence, we are able to show the equivalence between the KL gradient as

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\mathrm{KL}(Q_t\|\pi) = a_0 \sum_{y} (\mathcal{A}_a^* Q_t)(y)\log f_t(y) = -a_0 \mathcal{I}_a(f_t).$$

Leveraging the mLSI along the path, for each context c and $t \in [0,T]$, there exists a map $t \mapsto \alpha_t(c) > 0$ such that

$$\mathcal{I}_t(f_t \mid c) \geq \alpha_t(c)\,\mathrm{Ent}_{\pi_t(\cdot|c)}(f_t) = \alpha_t(c)\,\mathrm{KL}\left(Q_t^\theta(\cdot \mid c)\,\|\,\pi_t(\cdot \mid c)\right),$$

which gives the gradient descent across the KL as per-context learning bound

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\mathrm{KL}(Q_t\|\pi) = -a_0\mathcal{I}_a(f_t) \leq -a_0\alpha_t k_t(\tau,x).$$

Since α_t is obtained by a measurable function over t, we can choose a uniform learning bound by $\alpha_t = \inf_c(\alpha_t(c))$ to factor out the context-independent constant. From the uniformity, we shall factor out the constant α_t term then multiply by $w_t(c)$ and integrate to obtain the inequality:

$$\int_0^1 \sum_x w_t(\tau,x)\,\partial_t k_t(\tau,x)\,\mathrm{d}\tau \leq\ -a_0\alpha_t \int_0^1 \sum_x w_t(\tau,x)k_t(\tau,x)\,\mathrm{d}\tau$$
$$= -a_0\alpha_t\, H(s).$$

(2) Bounding $\dot{a}(t)\,\partial_a \mathrm{KL}(Q_t\|Q^{*,a})$ Term Let q be an arbitrary probability on $\mathcal{Y}$, and fix a time a. Similarly with SA-DTM analysis, define $\psi_a := \partial_a \log Q^{*,a}$. We have

$$\partial_a\,\mathrm{KL}(Q_t\|Q^{*,a}) = \partial_a \sum_y q\log\left(\frac{q}{Q^{*,a}}\right) = -\sum_y q\,\psi_a = -\langle q - Q^{*,a}, \psi_a\rangle$$

since $\mathbb{E}_{Q^{*,a}}[\psi_a] = \sum_y Q^{*,a}\partial_a \log Q^{*,a} = \sum_y \partial_a Q^{*,a} = 0$. By definition, q and $Q^{*,a}$ are both probabilities resulting in the invariance under shifts of ψ_a from $\langle q - Q^{*,a}, \mathbf{1}\rangle = 0$. Then, we obtain the similar bound with respect to $\tilde{\psi}_a := \psi_a - \frac{1}{2}(\max_y \psi_a + \min_y \psi_a)$ which gives

$$\|\tilde{\psi}_a\|_\infty \leq \frac{1}{2}\,\mathrm{osc}_y(\psi_a) \leq \frac{1}{2}\,\mathrm{osc}(r).$$

Subsequently by Hölder's inequality, we have

$$|\langle q - Q^{*,a}, \psi_a\rangle| \leq \frac{\mathrm{osc}(r)}{2}\|q - Q^{*,a}\|_{\mathrm{TV}},$$

which Pinsker's inequality finishes the bound

$$|\partial_a\,\mathrm{KL}(q\|Q^{*,a})| = |\langle q - Q^{*,a}, \psi_a\rangle|$$
$$\leq \frac{\mathrm{osc}(r)}{2}\|q - Q^{*,a}\|_{\mathrm{TV}} \leq \frac{\mathrm{osc}(r)}{2}\sqrt{2\,\mathrm{KL}(q\|Q^{*,a})}.$$

Letting $q = Q_t$ and multiplying both sides by $\dot{a}(t) \geq 0$, the resultant upper bound is given by

$$\dot{a}(t)\,\partial_a\,\mathrm{KL}(Q_t\|Q^{*,a(t)}) \leq \dot{a}(t)\frac{\mathrm{osc}(r)}{2}\sqrt{2k_t(\tau,x)}.$$

Its contribution to the Lyapunov functional $H(t)$ is then obtained by applying per-context weights $w_t(c)$ to give

$$\int_0^1 \sum_x w_t(\tau,x)\,\dot{a}(t)\,\partial_a k_t(\tau,x)\,\mathrm{d}\tau \le \dot{a}(t)\frac{\mathrm{osc}(r)}{2}\int_0^1 \sum_x w_t(\tau,x)\sqrt{2\,k_t(\tau,x)}\,\mathrm{d}\tau$$
$$\le \dot{a}(t)\frac{\mathrm{osc}(r)}{2}\Bigg(\underbrace{\int_0^1 \sum_x w_t(\tau,x)\cdot\,\mathrm{d}\tau}_{\mathfrak{N}}\Bigg)^{1/2}\Bigg(\underbrace{\int_0^1 \sum_x w_t(\tau,x)\cdot 2k_t(\tau,x)\,\mathrm{d}\tau}_{2H(t)}\Bigg)^{1/2}.$$

Thus, the final bound is given by

$$\int_0^1 \sum_x w_t(\tau,x)\,\dot{a}(t)\,\partial_a k_t(\tau,x)\,\mathrm{d}\tau \le \dot{a}(t)\frac{\mathrm{osc}(r)}{2}\sqrt{2\mathfrak{N}}\sqrt{H(s)}$$

(3) Bounding $\int_0^1 \sum_x \dot{w}_t(\tau,x)\,k_t(\tau,x)\,\mathrm{d}\tau$ term Define an auxiliary functional

$$\mathcal{F} := \int_0^1 \lambda_\tau^{\text{base}}(x_\tau)k_t(\tau,x)d\tau \ge 0.$$

Then, $\int_0^1 \sum_x w_t(c)k_t(c)d\tau = \mathbb{E}_{\mathbb{P}^a}[\mathcal{F}]$ and it is possible to write expectation over $\mathbb{P}^a$ as an expectation over $\mathbb{P}^{\text{base}}$ as $\mathbb{E}_{\mathbb{P}^{*,a}}[\mathcal{F}] = \mathbb{E}_{\mathbb{P}^{\text{base}}}\left[\mathcal{F}\cdot\frac{e^{ar(X_1)}}{Z_a}\right]$. Differentiating with $\partial_a \log(e^{ar(X_1)}/Z_a) = r(X_1) - \partial_a \log Z_a$, we can find out for the optimal tilted marginal $\rho_t^{*,a}(x) = \rho_t^{\text{base}}(x)h_a(t,x)/Z_a$, the marginal score is given by

$$\partial_a \log Z_a = \mathbb{E}_{\rho^{1,a}}[r] = \mathbb{E}_{\rho^{*,a}}[\mathbb{E}_a[r(x_1)\mid x_\tau = x]]$$

resulting in the gradient of the functional with respect to a as

$$\partial_a \mathbb{E}_{\mathbb{P}^a}[\mathcal{F}] = \mathbb{E}_{\mathbb{P}^{\text{base}}}\left[\mathcal{F}\cdot\frac{e^{ar(X_1)}}{Z_a}\big(r(X_1) - \partial_a \log Z_a\big)\right] = \mathbb{E}_{\mathbb{P}^{*,a}}\left[\mathcal{F}\cdot(r(X_1) - \mathbb{E}_{\mathbb{P}^{*,a}}[r])\right].$$

Since $\mathbb{E}_{\mathbb{P}^{*,a}}[\mathcal{F}\cdot\mathbb{E}_{\mathbb{P}^{*,a}}[r]] = \mathbb{E}_{\mathbb{P}^{*,a}}[\mathcal{F}]\cdot\mathbb{E}_{\mathbb{P}^{*,a}}[r]$, we show the equivalence with the coveriance by

$$\begin{aligned}\partial_a \mathbb{E}_{\mathbb{P}^{*,a}}[\mathcal{F}] &= \mathbb{E}_{\mathbb{P}^{*,a}}[\mathcal{F}\cdot(r(X_1) - \mathbb{E}_{\mathbb{P}^{*,a}}[r(X_1)])]\\ &= \mathrm{Cov}_{\mathbb{P}^{*,a}}(\mathcal{F}, r(X_1))\\ &\le \mathbb{E}_{\mathbb{P}^{*,a}}\left[\mathcal{F}\cdot|r(X_1) - \mathbb{E}_{\mathbb{P}^{*,a}}[r]|\right] \qquad (\mathcal{F}\ge 0)\end{aligned}$$

which gives

$$\partial_a \mathbb{E}_{\mathbb{P}^{*,a}}[\mathcal{F}] \le \mathrm{osc}(r)\cdot\mathbb{E}_{\mathbb{P}^{*,a}}[\mathcal{F}] = \mathrm{osc}(r)H(t).$$

Lastly, multiplying by $\dot{a}(t)$, we achieve the final result

$$\int_0^1 \sum_x \dot{w}_t(\tau,x)k_t(\tau,x)\,\mathrm{d}\tau = \dot{a}(t)\,\partial_a \mathbb{E}_{\mathbb{P}^{*,a}}[\mathcal{F}] \le \dot{a}(t)\,\mathrm{osc}(r)H(t).$$

Therefore, combining the three terms

$$\frac{\mathrm{d}}{\mathrm{d}t}H(t) = \int_0^1 \sum_x \dot{w}_t(\tau,x)k_t(\tau,x)\,\mathrm{d}\tau + \int_0^1 \sum_x w_t(\tau,x)\left(\partial_t \mathrm{KL}(Q_t\|Q^{*,a}) + \dot{a}(t)\,\partial_a \mathrm{KL}(Q_t\|Q^{*,a})\right)\mathrm{d}\tau$$
$$\le \dot{a}(t)\,\mathrm{osc}(r)H(t) - a_0\alpha_t\,H(t) + \dot{a}(t)\frac{\mathrm{osc}(r)}{2}\sqrt{2\mathfrak{N}}\sqrt{H(t)}$$
$$\le -(a_0\alpha_t - \dot{a}(t)\,\mathrm{osc}(r))H(t) + \dot{a}(t)\,\mathrm{osc}(r)\sqrt{\frac{\mathfrak{N}}{2}H(t)}.$$

Applying the Grönwall gives the desired result. □

Remark 8. *Looking at the conclusion from Theorem 5, the gradient of the Lyapunov functional across the time s is increased when $\dot{a}(t) \ge 0$ is decreased. Conversely, increasing the mirror flow rate a_0 (flow descent decays faster) and increasing the mLSI constant α_t (stronger mLSI) condition makes the Lyapnuv functional to decay faster.*

F EXPERIMENTAL DETAILS

In this section we provide the experiment details of the Discrete Tilt Matching and Discrete Adjoint Matching that were used in the evaluation. For most learning rate and reward we follow or slightly modify the original recipe.

F.1 REWARD FUNCTION

Maze Planning

CountDown. The reward is computed by `countdown_reward_func` (`dtm/reward_func.py`), which extracts the content of the last `<answer>...</answer>` tag and assigns

$$r_{\text{CountDown}} = \begin{cases} 1.0, & \text{if the equation evaluates to the target and uses each given number exactly once,} \\ 0.1, & \text{if an answer is present but invalid (wrong numbers, non-evaluable, or wrong value),} \\ 0.0, & \text{if no <answer> tag is present.} \end{cases}$$

The maximum attainable reward is max_rwd $= 1.0$.

Sudoku. The reward is computed by `sudoku_reward_func`, which extracts the digits from the last `<answer>` tag and pads/truncates the string to 16 cells. The reward is dense rather than near-binary:

$$r_{\text{Sudoku}} = 2.0 \times \frac{\text{\# correctly filled empty cells}}{\text{\# empty cells}},$$

with $r_{\text{Sudoku}} = 0.0$ if no answer is given. The maximum attainable reward is max_rwd $= 2.0$, so we report Sudoku "accuracy" as reward$/2$ to place it on the same $[0, 1]$ scale as CountDown's correctness rate.

F.2 HYPERPARAMETERS AND IMPLEMENTATION DETAILS

For both tasks, we use the AdamW optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.99$, $\epsilon = 1 \times 10^{-8}$, and weight decay of 0.1, with a linear learning-rate schedule (warmup over the first 5% and decay over the final 10% of each h-phase). We use a fixed seed of 42.

For **Sudoku**, we use a learning rate of 3×10^{-6}, annealed to a floor of 1×10^{-6}, with gradient clip at 1.0. We use 3-shot prompting. The base annealing configuration uses $h = 5$ with 256 gradient steps per h-phase; the EMA of the student adapter is disabled.

For **CountDown**, we use a learning rate of 4×10^{-6}, annealed to a floor of 2×10^{-6}, with gradient clip at 2.0. We use zero-shot prompting. The base annealing configuration uses $h = 6$ with 1200 gradient steps per h-phase; the EMA of the student adapter is enabled with decay 0.95.

c-DTM annealing schedule. Both tasks use $a_{\text{end}} = 100$ (the terminal tilt A) and control-variate coefficient $c = 1.0$, with SAR-aligned active-block-norm weighting. In the step-size sweeps, the base $(h, \text{steps_per_h})$ pairs above are overridden by $h \in \{2.5, 5, 10, 20, 25\}$, with steps_per_h $= 5120/N$ and $N = a_{\text{end}}/h = 100/h$ (e.g. for $h = 10$, $N = 10$ and 512 steps per stage).

Replay buffer. For Sudoku, we maintain a buffer of 80 distinct prompts, refreshing 20 every 32 steps in chunks of 16, with 4 prompts per batch and 1 completion per prompt (grad_accum $= 16$). For CountDown, we maintain a buffer of 40 distinct prompts, refreshing 10 every 25 steps in chunks of 30, with 12 prompts per batch and 6 completions per prompt (grad_accum $= 8$).

G RESULT DETAILS

In this section, we present the detailed result of the reward fine-tuning experiments that were conducted in the section 5. The numerical values discussed in experiment sections refer to the values here.

Table 2: Slowdown performance across varying stages N at three tilt strengths alongside rollout count and wall-clock time. KL indicates the KL divergence between the model distribution and the tractable target distribution. Target entropies: $H_8^\star = 2.15$, $H_{16}^\star = 1.58$, $H_{24}^\star = 1.50$; the idealized flow gives KL $\approx 10^{-16}$ for all N at every tilt.

		$A \cdot \mathrm{osc} = 8$		$A \cdot \mathrm{osc} = 16$		$A \cdot \mathrm{osc} = 24$		
N	steps/stage	KL	H_8	KL	H_{16}	KL	H_{24}	wall (s)
1	32768	1.84×10^{-1}	2.10	2.75×10^{0}	1.17	1.50×10^{1}	0.75	17.29
2	16384	1.64×10^{-1}	2.10	2.69×10^{0}	1.19	1.49×10^{1}	0.76	17.65
4	8192	1.41×10^{-1}	2.10	2.40×10^{0}	1.18	1.35×10^{1}	0.78	17.94
8	4096	1.10×10^{-1}	2.09	2.43×10^{0}	1.17	1.38×10^{1}	0.75	18.32
16	2048	7.61×10^{-2}	2.12	2.14×10^{0}	1.19	1.40×10^{1}	0.76	18.27
32	1024	5.22×10^{-2}	2.13	1.63×10^{0}	1.24	1.11×10^{1}	0.87	18.52
64	512	3.50×10^{-2}	2.13	1.06×10^{0}	1.34	9.28×10^{0}	0.98	19.00
128	256	1.60×10^{-2}	2.13	8.03×10^{-1}	1.34	7.94×10^{0}	1.03	20.24

Table 3: Compute-fixed DAM slowdown across stages N at three integer tilt strengths. Target entropies: $H_8^\star = 1.99$, $H_{16}^\star = 1.61$, $H_{24}^\star = 1.12$.

		$A \cdot \mathrm{osc} = 8$		$A \cdot \mathrm{osc} = 16$		$A \cdot \mathrm{osc} = 24$		
N	steps/stage	KL	H_8	KL	H_{16}	KL	H_{24}	wall (s)
1	32768	1.27×10^{-2}	1.99	2.70×10^{-2}	1.61	3.31×10^{-2}	1.11	134.4
2	16384	1.41×10^{-3}	1.99	1.62×10^{-3}	1.61	1.85×10^{-3}	1.11	135.4
4	8192	2.42×10^{-4}	1.99	5.27×10^{-4}	1.62	3.68×10^{-4}	1.12	138.2
8	4096	1.31×10^{-4}	2.00	2.68×10^{-4}	1.62	3.26×10^{-4}	1.12	141.7
16	2048	8.68×10^{-5}	1.99	2.66×10^{-4}	1.62	3.11×10^{-4}	1.12	150.0
32	1024	9.58×10^{-5}	1.99	3.59×10^{-4}	1.62	3.74×10^{-4}	1.12	167.2
64	512	7.33×10^{-5}	2.00	2.14×10^{-4}	1.62	2.55×10^{-4}	1.12	200.1
128	256	9.18×10^{-5}	1.99	3.47×10^{-4}	1.61	3.55×10^{-4}	1.12	267.0

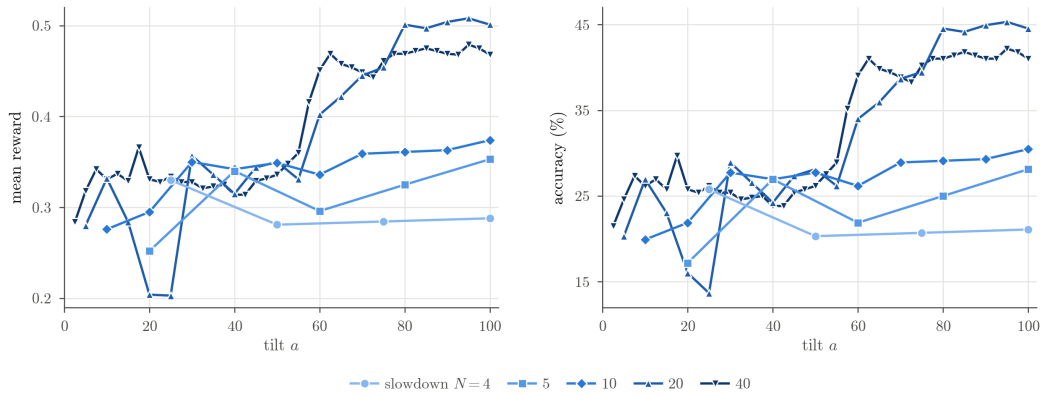

Figure 2: Fine-tuning performance as a increases.

In the main section, we have explained the phenomenon of U-shape bounce of the KL divergence at $N = 128$ in toy distribution fine-tuned with Discrete Adjoint Matching. We further hypothesized that this is largely due to the insufficient training within the annealing grid. Therefore, we validate DAM on small tractable distribution for more fine-grained annealing steps after increasing the total budget. Table 4 shows the KL divergence across multiple runs at $A \cdot \text{osc} = 8, 16$. In contrast to the $32768 = 2^{15}$ gradient step which the U-shaped KL divergence, the DAM with $524288 = 2^{19}$ gradient steps demonstrates a monotonic decrease in the KL divergence with the increasing annealing steps (decreasing annealing speed). Therefore, it can be interpreted that the U-shaped increase of the performance with respect to the gradient step is largely from the insufficient training budget especially the number of gradient steps.

Table 4: Increased budget experiment on the DAM experiment on the same tractable distribution $A \cdot \text{osc} = 8, 16$. The entropy has converged to the target similarly in this case.

N	steps/stage	KL $\pm$ sd ($A \cdot \text{osc} = 8$)	KL $\pm$ sd ($A \cdot \text{osc} = 16$)
1	524288	$6.30 \times 10^{-4} \pm 2.9 \times 10^{-4}$	$3.02 \times 10^{-3} \pm 2.1 \times 10^{-4}$
4	131072	$2.06 \times 10^{-5} \pm 3.9 \times 10^{-6}$	$1.22 \times 10^{-3} \pm 1.2 \times 10^{-4}$
16	32768	$1.01 \times 10^{-5} \pm 4.8 \times 10^{-6}$	$5.43 \times 10^{-4} \pm 7.8 \times 10^{-6}$
32	16384	$6.30 \times 10^{-6} \pm 1.7 \times 10^{-6}$	$1.60 \times 10^{-4} \pm 9.4 \times 10^{-7}$
64	8192	$5.57 \times 10^{-6} \pm 2.7 \times 10^{-6}$	$8.56 \times 10^{-5} \pm 4.3 \times 10^{-7}$
128	4096	$\mathbf{3.94 \times 10^{-6} \pm 1.5 \times 10^{-6}}$	$\mathbf{5.44 \times 10^{-5} \pm 5.1 \times 10^{-6}}$

At a large scale, we present the visualization of the paths generated on a Maze Planning task. The reward is set such that it deviates from the center region. It is possible to observe that with small reward $A = 5$ and fast annealing $N = 4$, the paths are deviating from the center. In the easy instances (columns 1, 3), $N = 4$ achieves higher mean reward than $N = 40$ (e.g. $+0.350$ vs. $+0.319$ in column (1) but collapses to a single distinct path, whereas the base model already had 1–2 distinct paths. This suggests fast annealing with a large effective tilt doesn't just optimize reward — it concentrates the sampler's mass onto one trajectory, sacrificing diversity for reward. Slower annealing partially preserves diversity even while nearly matching the reward gain.

H Relationship Between Slowly-Annealed Jumped-Process and Masked Diffusion Model Fine-tuning

The difference between slowly-annealed Glauber dynamics and MDM fine-tuning differ from the number of clocks: Glauber dynamics has 2 clocks (sampler time $t \in [0, T]$ and the slow annealing time $s = t/T$) while MDM has 3 (training time $t \in [0, T]$, the tilt $a(t) \in [0, A]$ and the MDM generation time $\tau \in [0, 1]$) that is internal and independent of t and a.

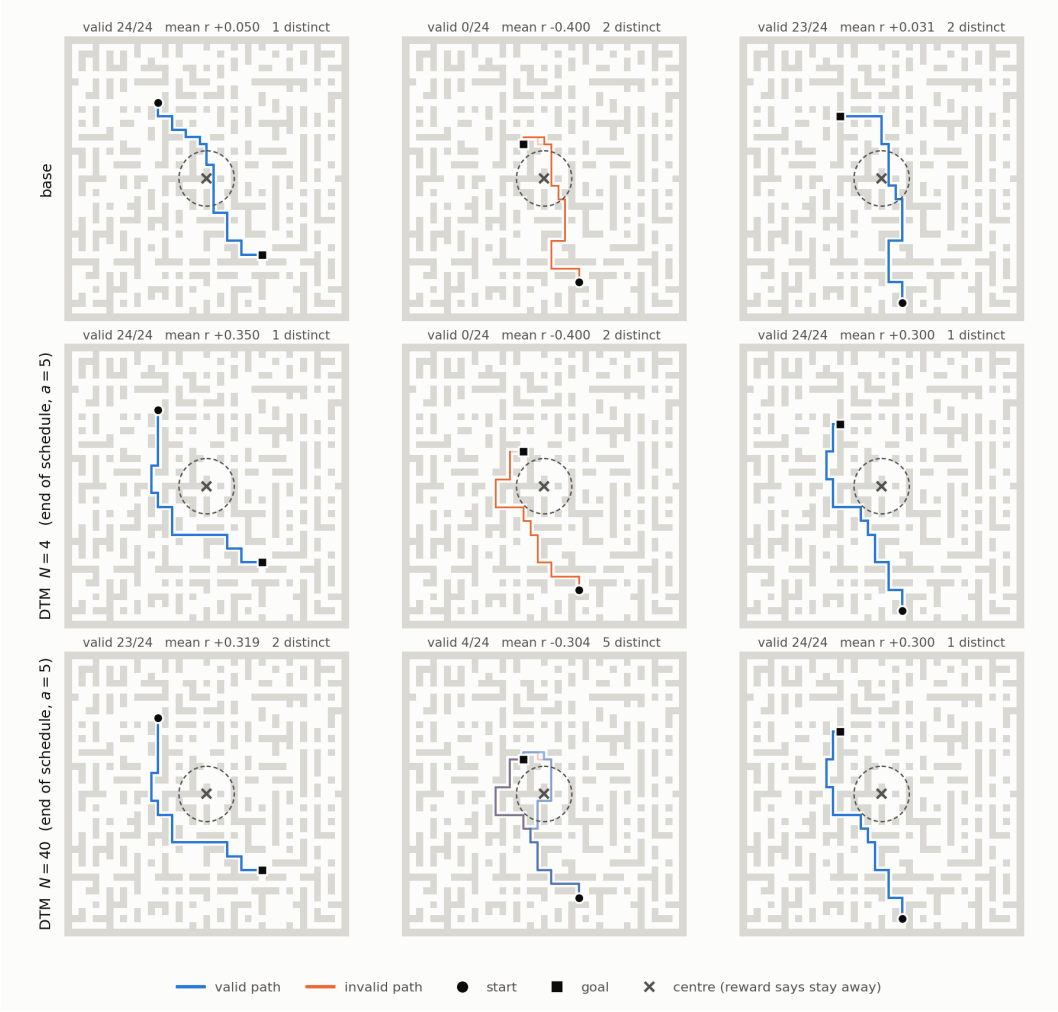

Figure 3: Generated paths by pretrained base and reward fine-tuned on a maze.

The relationship between slowly-annealed Glauber dynamics and MDM fine-tuning such that the sampler time t represent the same dynamics which is the base annealing time. Subsequently, the slowdown is applied linearly on both cases by $s(t) = t/T$ in the case of Glauber dynamics and by $a(t) = At/T$ in the case of MDM fine-tuning. Despite the two notations differ from each other, they represent the same quantity such that at the base case $T = 1$, there is no slowdown applied to the original annealing schedule. Only in the MDM case which is naturally a learned generator that iteratively unmasks the masked token with the learned probability, there exist the third clock τ, which in our analysis is not discussed as it does not contribute to the convergence analysis of the slow-annealing.

Furthermore, the annealing is applied on the annealing path denoted as $(\pi_t)_{t\in[0,T]}$ and respective slowdown is applied to annealing to follow $(\pi_s)_{s\in[0,1]}$ with $s = t/T$. Similarly, the annealing in the MDM fine-tuning is applied to the tilted target path $(\rho_{1,t})_{t\in[0,T]}$, which becomes meaningful with respect to the slowdonw $a(t) = At/T$ such that $(\rho_{1,a})_{a(t)\in[0,A]}$. For implementation, this is achieved by considering the annealing grid such that the a gradually increases by step size h instead of controlling the implicitly designed sampler time t.