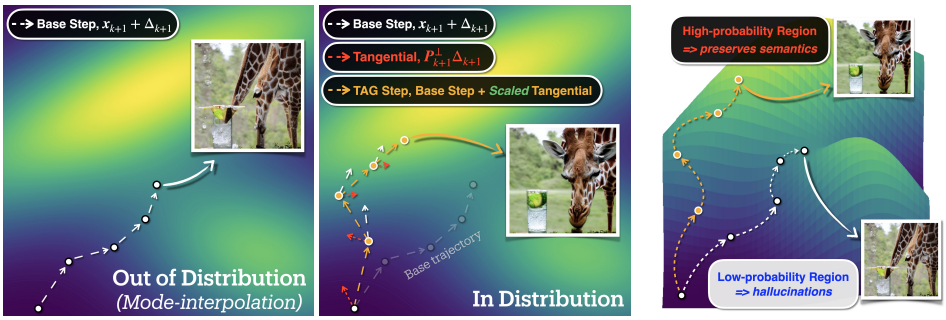
TAG: TANGENTIAL AMPLIFYING GUIDANCE FOR HALLUCINATION-RESISTANT DIFFUSION SAMPLING

Anonymous authors
Paper under double-blind review

ABSTRACT

Diffusion models achieve the state-of-the-art samples in image generation but often suffer from semantic inconsistencies or *hallucinations*. While various inference-time guidance methods can enhance generation, they often operate *indirectly* by relying on external signals or architectural modifications, which introduces additional computational overhead. In this paper, we propose **T**angential **A**mplifying **G**uidance **(TAG)**, a more efficient and *direct* guidance method that operates solely on trajectory signals without modifying the underlying diffusion model. TAG leverages an intermediate sample as a projection basis and amplifies the tangential components of the estimated scores with respect to this basis to correct the sampling trajectory. We formalize this guidance process by leveraging a first-order Taylor expansion, which demonstrates that amplifying the tangential component steers the state toward higher-probability regions, thereby reducing inconsistencies and enhancing sample quality. TAG is a plug-and-play, architecture-agnostic module that improves diffusion sampling fidelity with minimal computational addition, offering a new perspective on diffusion guidance.



(a) No Guidance (b) TAG Update

Figure 1: **Conceptual visualization of Tangential Amplifying Guidance (TAG)** from a mode-interpolation perspective (Aithal et al., 2024). Unlike **(a)** no guidance case, **(b)** TAG decomposes the base increment Δ_{k+1} on the latent sphere into parallel $\boldsymbol{P}_{k+1}\Delta_{k+1}$ and orthogonal $\boldsymbol{P}_{k+1}^{\perp}\Delta_{k+1}$ components (Eq. 6). It preserves the parallel component and amplifies *only* the tangential component by adding a tangential component with the scaling factor to the base step. This *tends to* guide trajectories toward higher-model density regions while mitigating off-manifold drift (§ 4, Eq. 15).

1 INTRODUCTION

Hallucination in diffusion models refers to the phenomenon of generating samples that violate the data distribution or contradict conditioning, thus failing to provide meaningful outputs. For example, it often manifests as mixed-up objects (Okawa et al., 2023) or anatomically implausible structures. Recent evidence suggests that the primary source of such errors lies in a failure mode known as mode interpolation. During sampling, trajectories may traverse low-density valleys between distinct modes of the data distribution, causing attribute mismatches and structural inconsistencies (Aithal et al., 2024).

A widely adopted remedy involves inference-time guidance strategies, such as classifier-free guidance (CFG) (Ho & Salimans, 2021) and their variants (Hong et al., 2023; Ahn et al., 2024; Karras et al., 2024; Rajabi et al., 2025; Kwon et al., 2025; Sadat et al., 2025; Dinh et al., 2025; Hong, 2024).

Under the assumption that deviating from low-probability regions enhances sample quality, most of these methods employ *residual scaling*, using the difference between the conditional and unconditional branches to guide the generation process away from the unconditional model's outputs. While effective, these mechanisms are fundamentally *indirect*: instead of navigating along the *underlying geometry* of the data distribution, they proceed by repeatedly moving away from an unconditional estimate at each step of the process.

In contrast, we propose a more efficient *direct* solution based on Tweedie's identity (Tweedie et al., 1984), which relates the score to the posterior mean of the clean data under Gaussian corruption. It motivates a decomposition of the model update based on its *intrinsic geometry*; a split into a *drift component* that changes the noise level and a *tangent component* to the iso-noise manifold that refines structure and semantics. We observe that the tangential component carries rich structural information (Figure 2), and amplifying it reduces out-of-distribution generations (Figure 3).

Drawing upon the principle of *amplifying the tangential component* during inference, we derive **T**angential **A**mplifying **G**uidance (**TAG**), a plug-and-play method that emphasizes the tangential component of the *score update*. TAG steers the sampling trajectory to follow the underlying data manifold closely. TAG integrates seamlessly with standard diffusion backbones—whether conditioned or not—without requiring additional denoising evaluations or retraining.

We can summarize our contributions as follows:

- We establish a concrete link between the score's intrinsic geometry and sample quality, proving that amplifying tangential components of the scores steer sampling trajectories toward in-distribution manifold.
- We introduce TAG, a computationally efficient and architecture-agnostic algorithm that realizes this geometric principle into practice.

2 PRELIMINARIES

Score-based Diffusion Model. Score-based generative models learn a time-indexed score function that approximates the gradient of the log-density of noise-perturbed data,

$$\boldsymbol{s}_\theta(\boldsymbol{x}, t_k) \approx \nabla_{\boldsymbol{x}} \log p(\boldsymbol{x}|t_k),$$

to reverse a gradual noising process for sample generation. This approach provides a continuous-time framework that unifies earlier discrete-time Denoising Diffusion Probabilistic Models (DDPMs) (Sohl-Dickstein et al., 2015; Ho et al., 2020) through the lens of stochastic differential equations (SDEs) (Song et al., 2020b). The core idea involves a forward-time SDE that transforms complex data into a simple prior distribution, given by

$$d\boldsymbol{x}_k = \mathbf{f}(\boldsymbol{x}_k, t_k)dt_k + g(t_k)d\mathbf{W}t_k.$$

Generation is then performed by a corresponding reverse-time SDE, which becomes tractable by substituting the unknown true score with the learned model $\boldsymbol{s}_\theta$ (Anderson, 1982). This score network, typically a noise-conditional U-Net, is trained efficiently via denoising score matching across various noise levels (Vincent, 2011; Song & Ermon, 2019). For sampling, one can use numerical methods like predictor-corrector schemes to simulate the stochastic reverse SDE, or solve an associated deterministic ordinary differential equation (ODE) known as the probability-flow ODE. This continuous-time framework not only provides a theoretical basis for widely used deterministic samplers like DDIM (Song et al., 2020a) but has also inspired modern refinements, such as the preconditioning and parameterization in EDM (Karras et al., 2022), which further enhance the trade-off between sample quality and efficiency.

Inference-Time Guidance. Numerous methods modify the update field during sampling to improve fidelity, typically without requiring *retraining*. Early approaches (Ho & Salimans, 2021) often rely on *residual signals*, which scale the update residual to better align samples with a desired condition. However, such isotropic manipulations can reduce sample diversity or disrupt the scheduler's calibration (Dhariwal & Nichol, 2021; Kynkäänniemi et al., 2024). These drawbacks have motivated alternatives that consider the *geometry* of the diffusion trajectory. A complementary line of work replaces external cues with model-internal signals (Hong et al., 2023; Ahn et al., 2024; Hong, 2024).

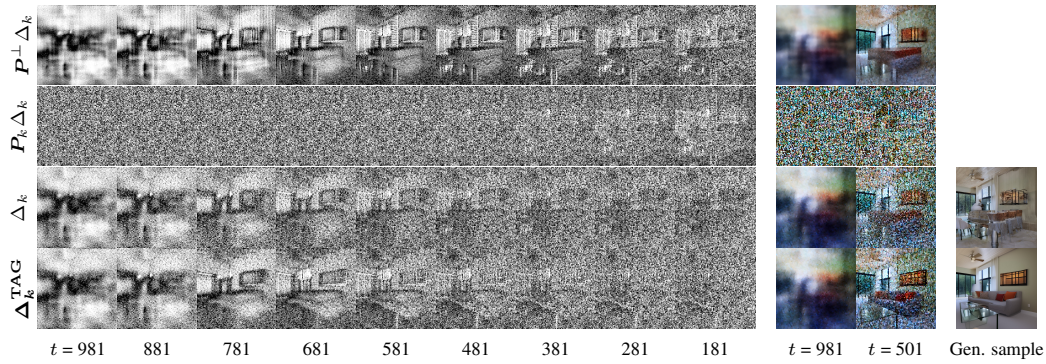


Figure 2: **Amplifying the tangential component enhances semantic content by isolating it from noise.** This figure illustrates the decomposition of the update step Δ_k into *normal* and *tangential* components. Subtracting the unstructured, noisy normal component $\boldsymbol{P}_k\Delta_k$ from the original update acts as a *denoising operation*, revealing the tangential component $\boldsymbol{P}_k^{\perp}\Delta_k$, which preserves the *principal semantic structure*. Images decoded from intermediate timesteps (t=981, 501) indicate that semantic information is most salient in the tangential component. Motivated by this observation, our method $\boldsymbol{\Delta}_{\boldsymbol{k}}^{\mathbf{TAG}}$ amplifies this semantically rich component, yielding a clearer and more coherent final sample (far right) than that obtained from the unmodified Δ_k (Please zoom-in for details).

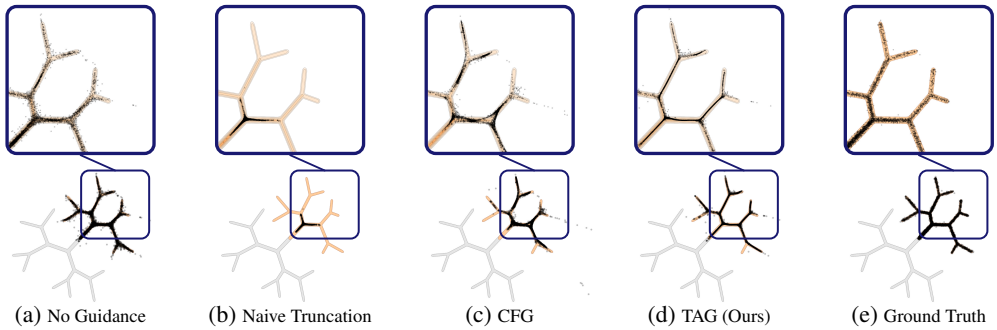


Figure 3: **Sampling on a 2D branching distribution (Karras et al., 2024) under different guidance methods.** (a) *No guidance*: probability mass drifts off the data manifold, yielding fragmented branches and OOD (Out of Distribution) points. (b) *Naive truncation*: suppresses some OOD but oversimplifies the geometry, dropping fine branches. (c) *CFG*: reduces boundary violations but also reduces diversity and can still leave OOD strays in our run. (d) ***TAG (Ours)***: trajectories are steered toward high-density regions along the branches, suppressing off-manifold outliers while retaining detail. (e) *Ground truth*. Overall, TAG achieves the highest similarity to the GT distribution **without additional #NFE**, concentrating mass on the correct branches while substantially reducing residual OOD outliers.

The common aim of these strategies is to steer the inference-time update to *suppress* directions associated with off-manifold drift while preserving the learned prior. More recent formulations make this objective explicit by distinguishing motion along iso-noise level sets from motion normal to them (Sadat et al., 2025; Kwon et al., 2025). Such geometry-aware perspectives offer a principled basis for guidance design and integrate cleanly with modern solvers.

3 Motivation and Intuition

Under Gaussian corruption, Tweedie's formula (Tweedie et al., 1984) links the posterior mean of the clean signal to the noisy observation via the score (i.e., the gradient of the log marginal density):

$$\mathbb{E}[\boldsymbol{x}_0|\boldsymbol{x}_k] = \Big(\underbrace{\boldsymbol{x}_k}_{:=\text{ drift term}} + \underbrace{\sigma_k^2 \nabla_{\boldsymbol{x}} \log p(\boldsymbol{x} \mid t_k)\big|_{\boldsymbol{x}=\boldsymbol{x}_k}}_{:=\text{ Tweedie increment } \Delta_k^{\mathrm{Tw}},\ (a.k.a.\text{ data term})} \Big) / \sqrt{\bar{\alpha}_k}. \tag{1}$$

Geometrically, the score field $\nabla_{\boldsymbol{x}} \log p(\boldsymbol{x}|t_k)\big|_{\boldsymbol{x}=\boldsymbol{x}_k}$ points in the direction of steepest increase of the marginal density. Tweedie's formula therefore adjusts $\boldsymbol{x}_k$ in this ascent direction, nudging the state

toward higher-probability regions. Therefore, the aim of modeling is to bias this movement toward *data-driven directions*.

However, naively guiding the states to chase *higher-probability regions* can disturb the scheduler's prescribed radius/SNR trajectory and may degrade sample quality(Figure 4). Accordingly, to avoid altering the radial term, we *isolate* $\boldsymbol{x}_k$ and reweight only the increment by decomposing it into *normal* and *tangential* parts with respect to $\widehat{\boldsymbol{x}}_k := \boldsymbol{x}_k/\|\boldsymbol{x}_k\|_2$: $\boldsymbol{P}_k = \widehat{\boldsymbol{x}}_k\widehat{\boldsymbol{x}}_k^\top$ and $\boldsymbol{P}_k^\perp = I - \boldsymbol{P}_k$. Guided by this separation, we form the amplified state $\boldsymbol{x}^+$, where the normal component is fixed and only the tangential component is amplified, via

$$\boldsymbol{x}^+ = \left(\boldsymbol{x}_k + \boldsymbol{P}_k\Delta_k^{\mathrm{Tw}}\right) + \eta\,\boldsymbol{P}_k^\perp\Delta_k^{\mathrm{Tw}}, \quad \text{with} \quad \eta \geq 1. \tag{2}$$

By doing so, we can preserve the radial first-order term (Eq. equation 17) while biasing the step toward *higher-probability regions*, $\nabla_{\boldsymbol{x}} \log p(\boldsymbol{x}|t_k)\big|_{\boldsymbol{x}=\boldsymbol{x}_k}$ (Empirical evidence is provided in Figure 2 and 3). In the following section (§4.1), we formalize this bias as a *constrained MLE* update that allocates first-order gain to the tangential subspace.

4 TAG: Tangential Amplifying Guidance

We introduce Tangential Amplifying Guidance (TAG), which reweights base increments along normal/tangential directions on the latent space.

Definitions & Algorithm. We work per sample on $\mathbb{R}^{C\times H\times W} \cong \mathbb{R}^d$ with Euclidean inner product $\langle\cdot,\cdot\rangle$ and norm $\|\cdot\|_2$. Let $\{t_k\}_{k=K}^0$ be *descending* timesteps with $t_K > \cdots > t_0$, and let ϵ_θ denote the denoiser. Given $\boldsymbol{x}_{k+1}$ at time t_{k+1}, the denoiser predicts

$$\boldsymbol{\varepsilon}_{k+1} \;=\; \epsilon_\theta(\boldsymbol{x}_{k+1}, t_{k+1}).$$

A base solver (e.g., DDIM) then produces a *provisional* state (Karras et al., 2022)

$$\tilde{\boldsymbol{x}}_k \;=\; a_{k+1}\,\boldsymbol{x}_{k+1} \;+\; b_{k+1}\,\boldsymbol{\varepsilon}_{k+1}, \quad \text{where} \quad a_{k+1}, b_{k+1} \quad \text{are base solver coefficients.} \tag{3}$$

Corresponding base increment at $\boldsymbol{x}_{k+1}$ as

$$\Delta_{k+1} \;:=\; \tilde{\boldsymbol{x}}_k - \boldsymbol{x}_{k+1}. \tag{4}$$

For any $\boldsymbol{x}\in\mathbb{R}^d$, we define the unit vector and orthogonal projectors

$$\widehat{\boldsymbol{x}} = \boldsymbol{x} \,/\, \|\boldsymbol{x}\|_2, \qquad \boldsymbol{P}(\boldsymbol{x}) = \widehat{\boldsymbol{x}}\widehat{\boldsymbol{x}}^\top, \qquad \boldsymbol{P}^\perp(\boldsymbol{x}) = \boldsymbol{I} - \boldsymbol{P}(\boldsymbol{x}). \tag{5}$$

Given positive scales $\eta \geq 1$, TAG *reweights* the base increment at $\boldsymbol{x}_{k+1}$:

$$\boldsymbol{x}_k \;=\; \boldsymbol{x}_{k+1} \;+\; \boldsymbol{P}_{k+1}\,\Delta_{k+1} \;+\; \eta\,\boldsymbol{P}_{k+1}^\perp\,\Delta_{k+1} \tag{6}$$

where $\boldsymbol{P}_{k+1} = \boldsymbol{P}(\boldsymbol{x}_{k+1})$ and $\boldsymbol{P}_{k+1}^\perp = \boldsymbol{P}^\perp(\boldsymbol{x}_{k+1})$.

Algorithm 1 Tangential Amplifying Guidance (TAG)

Require: Denoiser $\epsilon_\theta(\cdot)$, timesteps $\{t_k\}_{k=K}^0$, base solver coefficients a_{k+1}, b_{k+1}, TAG scale $\eta \geq 1$

1: Sample $\boldsymbol{x}_K \sim \mathcal{N}(\mathbf{0}, I)$
2: **for** $k = K-1, \ldots, 0$ **do**
3: $\quad\boldsymbol{\varepsilon}_{k+1} \leftarrow \epsilon_\theta(\boldsymbol{x}_{k+1}, t_{k+1})$ ▷ noise prediction
4: $\quad\tilde{\boldsymbol{x}}_k \leftarrow a_{k+1}\boldsymbol{x}_{k+1} + b_{k+1}\,\boldsymbol{\varepsilon}_{k+1}$ ▷ e.g., `scheduler.step`
5: $\quad\Delta_{k+1} \leftarrow \tilde{\boldsymbol{x}}_k - \boldsymbol{x}_{k+1}$ ▷ base increment
6: $\quad\widehat{\boldsymbol{x}}_{k+1} \leftarrow \boldsymbol{x}_{k+1}/\|\boldsymbol{x}_{k+1}\|_2$
7: $\quad\boldsymbol{P}_{k+1} \leftarrow \widehat{\boldsymbol{x}}_{k+1}\widehat{\boldsymbol{x}}_{k+1}^\top, \quad \boldsymbol{P}_{k+1}^\perp \leftarrow I - \boldsymbol{P}_{k+1}$ ▷ projectors at $\boldsymbol{x}_{k+1}$
8: $\quad\boldsymbol{x}_k \leftarrow \boldsymbol{x}_{k+1} + \boldsymbol{P}_{k+1}\Delta_{k+1} + \eta\,(\boldsymbol{P}_{k+1}^\perp\Delta_{k+1})$ ▷ TAG amplification

4.1 Why does TAG improve Image Quality?

Log-likelihood maximization. A foundational goal of training generative models is to maximize the log-likelihood of the data, as formalized by the Maximum Likelihood Estimation (MLE) principle:

$$\max_\theta \sum_i \log p_\theta(\boldsymbol{x}_i). \tag{7}$$

This principle suggests that high-quality samples should concentrate in regions of high probability. To connect this idea to an *update rule*, we relate likelihood increase to movement along the score via a local linearization:

$$\log p_\theta(\boldsymbol{x}) = \log p_\theta(\boldsymbol{x}_0) + (\boldsymbol{x} - \boldsymbol{x}_0)^\top \nabla_{\boldsymbol{x}} \log p_\theta(\boldsymbol{x})\big|_{\boldsymbol{x}=\boldsymbol{x}_0} + \mathcal{O}(\|\cdot\|^2). \tag{8}$$

Diffusion models (Song et al., 2020b; Ho et al., 2020) are designed to predict a score function, $\nabla_{\boldsymbol{x}} \log p(\boldsymbol{x} \mid t_k)\big|_{\boldsymbol{x}=\boldsymbol{x}_k} \approx -\epsilon_\theta(\boldsymbol{x}_k, t_k)/\sigma_k$, which operates on noisy versions of the data. Because diffusion models learn this score field, optimizing the global likelihood (Eq. 7) for a sample $\boldsymbol{x}_0$ during inference is *not directly tractable*. Therefore, we propose to apply the spirit of MLE at each *local step* of the sampling trajectory.

$$\log p(\boldsymbol{x}_k \mid t_{k+1}) \approx \log p(\boldsymbol{x}_{k+1} \mid t_{k+1}) + (\boldsymbol{x}_k - \boldsymbol{x}_{k+1})^\top \nabla_{\boldsymbol{x}} \log p(\boldsymbol{x} \mid t_{k+1})\big|_{\boldsymbol{x}=\boldsymbol{x}_{k+1}} + \mathcal{O}(\|\cdot\|^2). \tag{9}$$

The idea of enhancing a pre-trained score function with inference-time guidance has proven effective. For instance, when the score function is well trained on given training sets and this leads to well-trained maximum log-likelihood, we observe that the pre-trained score function could be improved by CFG (Ho & Salimans, 2021) which linearly biases the score toward the conditional target. Inspired by this, our approach provides inference-time guidance on the score function by maximizing the following local log-likelihood term, thereby guiding the sampling trajectory towards high-likelihood regions of the data distribution and reducing off-manifold artifacts (hallucination):

$$\max_{\boldsymbol{x}_k} (\boldsymbol{x}_k - \boldsymbol{x}_{k+1})^\top \nabla_{\boldsymbol{x}} \log p(\boldsymbol{x} \mid t_{k+1})\big|_{\boldsymbol{x}=\boldsymbol{x}_{k+1}} \tag{10}$$

Single-step increment decomposition. For deterministic DDIM/ODE samplers, the *single-step score state decomposition* can be written as

$$\Delta_{k+1} := \tilde{\boldsymbol{x}}_k - \boldsymbol{x}_{k+1} \;=\; \tilde{\alpha}_k \epsilon_\theta(\boldsymbol{x}_{k+1}, t_{k+1}) + \beta_k \boldsymbol{x}_{k+1}, \tag{11}$$

with coefficients

$$\tilde{\alpha}_k := \sigma_k - \tfrac{\sqrt{\bar{\alpha}_k}}{\sqrt{\bar{\alpha}_{k+1}}}\sigma_{k+1}, \quad \beta_k := \tfrac{\sqrt{\bar{\alpha}_k}}{\sqrt{\bar{\alpha}_{k+1}}} - 1, \quad \text{with} \quad \tilde{\alpha}_k < 0,\ \beta_k > 0,$$

where $\bar{\alpha}$ is standard diffusion cumulative product term. Using the projection operators, which satisfy $\boldsymbol{P}_{k+1}^\perp \boldsymbol{x}_{k+1} = 0$ and $\boldsymbol{P}_{k+1}\boldsymbol{x}_{k+1} = \boldsymbol{x}_{k+1}$, yields the *projection-wise* identities

$$\begin{aligned} \boldsymbol{P}_{k+1}^\perp \Delta_{k+1} &= \tilde{\alpha}_k \boldsymbol{P}_{k+1}^\perp \epsilon_\theta(\boldsymbol{x}_{k+1}, t_{k+1}), \\ \boldsymbol{P}_{k+1} \Delta_{k+1} &= \tilde{\alpha}_k \boldsymbol{P}_{k+1} \epsilon_\theta(\boldsymbol{x}_{k+1}, t_{k+1}) + \beta_k \boldsymbol{x}_{k+1}. \end{aligned} \tag{12}$$

Substituting equation 12 into the equation 6 gives

$$\boldsymbol{x}_k^{\text{TAG}} = \boldsymbol{x}_{k+1} + \tilde{\alpha}_k \big[\boldsymbol{P}_{k+1} + \eta \boldsymbol{P}_{k+1}^\perp\big] \epsilon_\theta(\boldsymbol{x}_{k+1}, t_{k+1}) + \beta_k \boldsymbol{x}_{k+1}, \quad \text{with} \quad \eta \geq 1. \tag{13}$$

Therefore, the TAG update $\Delta_{k+1}^{\text{TAG}}$ can be expressed in terms of the decomposed components of the original update Δ_{k+1}:

$$\Delta_{k+1}^{\text{TAG}} = \big(\boldsymbol{P}_{k+1} + \eta\, \boldsymbol{P}_{k+1}^\perp\big) \Delta_{k+1}. \tag{14}$$

Finally, assuming that the log-density is smooth (assume $\log p(\cdot|t_{k+1})$ is C^2 in a neighborhood of $\boldsymbol{x}_{k+1}$), the *first order Taylor expansion gain* for a small TAG update $\Delta_{k+1}^{\text{TAG}} \in \mathbb{R}^d$ is

$$G(\eta) := \big(\Delta_{k+1}^{\text{TAG}}\big)^\top \nabla_{\boldsymbol{x}} \log p(\boldsymbol{x} \mid t_{k+1})\big|_{\boldsymbol{x}=\boldsymbol{x}_{k+1}} \tag{15}$$

Next, we prove that increasing η provide monotonic increase this first-order gain.

Theorem 4.1 (Monotonicity of the First-order Taylor Gain)**.** *Assume a deterministic base step with $\Delta_{k+1} = \tilde{\alpha}_k \epsilon_\theta(\boldsymbol{x}_{k+1}, t_{k+1}) + \beta_k \boldsymbol{x}_{k+1}$ and $\tilde{\alpha}_k \leq 0$. Let $\boldsymbol{P}_{k+1} \succeq 0$ and $\boldsymbol{P}_{k+1}^\perp \succeq 0$ be the projectors defined above. For the TAG step $\Delta_{k+1}^{\text{TAG}} = \boldsymbol{P}_{k+1}\Delta_{k+1} + \eta\, \boldsymbol{P}_{k+1}^\perp \Delta_{k+1}$, the first-order Taylor gain $G(\eta) := \big(\Delta_{k+1}^{\text{TAG}}\big)^\top \nabla_{\boldsymbol{x}} \log p(\boldsymbol{x} \mid t_{k+1})\big|_{\boldsymbol{x}=\boldsymbol{x}_{k+1}}$ satisfies*

$$\frac{\partial G(\eta)}{\partial \eta} \approx \frac{-\tilde{\alpha}_k}{\sigma_{k+1}} \left\|\boldsymbol{P}_{k+1}^\perp \epsilon_\theta(\boldsymbol{x}_{k+1}, t_{k+1})\right\|_2^2 \;\geq\; 0,$$

and, in particular,

$$G^{\text{TAG}} - G^{\text{base}} = \underbrace{-\sigma_{k+1}^{-1} \cdot \big(\tilde{\alpha}_k(\eta - 1)\big)}_{\geq\, 0 \ \text{as} \ \tilde{\alpha}_k \leq 0} \cdot \left\|\boldsymbol{P}_{k+1}^\perp \epsilon_\theta(\boldsymbol{x}_{k+1}, t_{k+1})\right\|_2^2 \geq 0,$$

Equality holds iff $\eta = 1$. The proof is provided in Appendix A.

Table 1: **Quantitative results across previous guidance methods and +TAG sampling settings for unconditional generation.** Evaluated on the ImageNet *val* with 30K samples. All images are sampled with Stable Diffusion (SD) v1.5 using the DDIM sampler.

Methods	Guidance Scale	TAG Amp. (η)	#NFE	#Steps	FID ↓	IS ↑
DDIM (Song et al., 2020a)	–	–	50	50	76.942	14.792
DDIM + **TAG**	–	1.05	50	50	67.971	**16.620**
DDIM + **TAG**	–	1.15	50	50	**67.805**	16.487
DDIM + **TAG**	–	1.25	50	50	71.801	15.815
SAG (Hong et al., 2023)	0.2	–	50	25	71.984	15.803
SAG + **TAG**	0.2	1.15	50	25	**65.340**	**17.014**
PAG (Ahn et al., 2024)	3	–	50	25	64.595	19.30
PAG + **TAG**	3	1.15	50	25	**63.619**	**19.90**
SEG (Hong, 2024)	3	–	50	25	65.099	17.266
SEG + **TAG**	3	1.15	50	25	**60.064**	**18.606**

Table 2: **Quantitative results of TAG on various Stable Diffusion baselines.** The table presents a comparison for Stable Diffusion (SD) v2.1 and SDXL, evaluated on 10K ImageNet validation images using the DDIM sampler with 50 NFEs.

Methods	TAG Amp. (η)	#NFE	#Steps	FID ↓	IS ↑
SD v2.1 (Rombach et al., 2022)	–	50	50	100.977	11.553
SD v2.1 + **TAG**	1.15	50	50	**88.788**	**13.311**
SDXL (Podell et al., 2024)	–	50	50	124.407	9.034
SDXL + **TAG**	1.20	50	50	**113.798**	**9.716**

Table 3: **Quantitative results for *unconditional image generation*** on the ImageNet dataset. We leverage a Stable Diffusion (SD) v1.5. All metrics are calculated using 30K samples. We further demonstrate that *strong performance* is achievable ***even with fewer #NFE***. We measure the inference time using *torch.cuda.Event* and report the average over 100 runs.

	DDIM (Song et al., 2020a)			**DPM++** (Lu et al., 2025)	
	SD v1.5	+ **TAG**	+ **TAG**	SD v1.5	+ **TAG**
#NFE	50	25	50	10	10
Inference Time (s)	1.9507	1.0191	1.9674	0.4433	0.4522
FID ↓	76.942	72.535	**67.805**	85.983	74.238
IS ↑	14.794	15.528	**16.487**	13.037	14.930

Log-likelihood improvements via TAG. We cast inference-time guidance as maximizing a log-likelihood gain (equation 10). TAG simply reweights the update step by amplifying the component that is orthogonal to the current state while leaving the parallel component unchanged. By Theorem 4.1, increasing the orthogonal weight monotonically raises the first-order Taylor gain, so TAG steers the sampler toward higher-density regions of the data manifold, improving image quality.

Avoidance of normal amplification. Amplifying the tangential component monotonically increases the first-order term of a Taylor gain of $\log p(\cdot|t_{k+1})$ (Theorem 4.1), which produces samples with less hallucination. However, amplifying the normal component increases radial contraction and leads to over-smoothing (Figure 4). This radial component of the single-step is aligned with the radial part of Tweedie's correction, which links $\boldsymbol{x}_k$ to the posterior mean $\mathbb{E}[\boldsymbol{x}_0|\boldsymbol{x}_k]$ via the score function (Tweedie et al., 1984; Song et al., 2020b). Formally, rescaling the normal part by a

Figure 4: **Effectiveness of TAG**. At 50 NFEs, TAG surpasses the sample quality at 250 NFEs from baseline. In contrast, *+Normal* causes severe over-smoothing.

Table 4: **Quantitative results across guidance-only (i.e. CFG, PAG, SEG) and guidance w/ TAG sampling settings.** Evaluated on the MS-COCO 2014 *val* split with 10k random text prompts. All images are sampled with Stable Diffusion v1.5 using the DDIM sampler. *cfg_scale=2.5*, *pag_scale=2.5* and *seg_scale=2.5* is applied for each experiments.

Methods	TAG Amp. (η)	#NFE	#Steps	FID ↓	CLIPScore ↑
Condition-Only	–	30	30	85.145	19.77 ±3.43
Condition-Only **+ TAG**	1.2	30	30	**58.438**	**21.88** ±**2.99**
CFG (Ho & Salimans, 2021)	–	100	50	26.266	22.60 ± 3.28
CFG + **C-TAG**	2.5	*30*	*15*	**23.414**	**22.82** ± **3.21**
PAG (Ahn et al., 2024)	–	50	25	24.280	**22.72** ± **3.25**
PAG + **C-TAG**	1.25	50	25	**22.109**	22.07 ± 3.49
SEG (Hong, 2024)	–	50	25	29.215	**18.17** ± **3.55**
SEG + **C-TAG**	1.25	50	25	**23.446**	16.94 ± 3.96

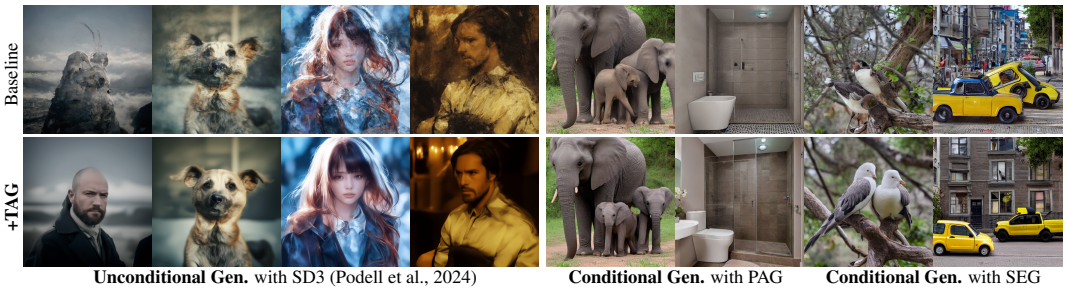


Figure 5: **Qualitative comparison of TAG across unconditional and conditional generation settings.** The left four columns demonstrate that for unconditional generation, TAG enhances the detail and coherence of samples from the SD3 (Podell et al., 2024). The right four columns show that for conditional generation, TAG can be applied on top of existing guidance methods (e.g., PAG (Ahn et al., 2024), SEG (Hong, 2024)) to further improve their outputs.

κ (> 1), the radial first–order change is multiplied by κ:

$$\langle \widehat{\boldsymbol{x}}_{k+1}, \Delta_{k+1}^{(\kappa)} \rangle = \kappa \, \langle \widehat{\boldsymbol{x}}_{k+1}, \Delta_{k+1} \rangle. \tag{16}$$

Therefore, a value of κ (> 1) excessively strengthens this contraction under the VP/DDIM schedule, leading to over-smoothing. In contrast, tangential scaling by η (> 1) preserves the radial first–order term:

$$\langle \widehat{\boldsymbol{x}}_{k+1}, \Delta_{k+1}^{\mathrm{TAG}} \rangle = \langle \widehat{\boldsymbol{x}}_{k+1}, \Delta_{k+1} \rangle. \tag{17}$$

To summarize, normal amplification breaks one–step calibration and *induces over-smoothing*, whereas tangential boosting improves alignment without disturbing the radial schedule.

4.2 Tangential Amplifying Guidance for Conditional Generation

Our analysis (§3, §4) shows that the *tangential* component of the score field encodes manifold-consistent, content-selective directions whose amplification improves image quality. In CFG (Ho & Salimans, 2021), the guided score combines conditional and unconditional branches:

$$\widetilde{\boldsymbol{\varepsilon}}_k = \epsilon_\theta(\boldsymbol{x}_k, \boldsymbol{c}) + \omega(\epsilon_\theta(\boldsymbol{x}_k, \boldsymbol{c}) - \epsilon_\theta(\boldsymbol{x}_k, \emptyset)). \tag{18}$$

Because these two scores follow distinct trajectories, an incoherence between them can arise, and such an effect can degrade generation quality, an issue recently highlighted by Kwon et al. (2025). Motivated by this established score mismatch, and informed by our core intuition that the tangential field encodes data geometry (equation 1), we posit that this incoherence is fundamentally tangential in nature; that is, a persistent *mismatch* exists primarily between the conditional and unconditional tangential components.

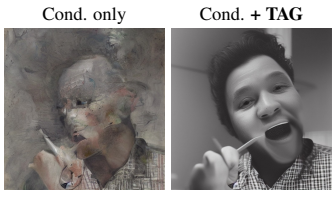


Figure 6: **Conditional generation without CFG.** Compared to the condition-only baseline, adding TAG produces more faithful semantics for the prompt at matched NFE.

Algorithm 2 Conditional TAG (C-TAG)

Require: Denoiser $\epsilon_\theta(\cdot)$, timesteps $\{t_k\}_{k=K}^{0}$, CFG scale ω, TAG scale $\eta \geq 0$

Sample $\boldsymbol{x}_K \sim \mathcal{N}(\mathbf{0}, I)$ ▷ initialize from prior
for $k = K-1, \ldots, 0$ **do**
  $(\boldsymbol{\varepsilon}_u, \boldsymbol{\varepsilon}_c) \leftarrow \epsilon_\theta(\boldsymbol{x}_{k+1}, t_{k+1}, \cdot)$ ▷ uncond / cond noise
  $\boldsymbol{g}_k \leftarrow \boldsymbol{\varepsilon}_c - \boldsymbol{\varepsilon}_u$ ▷ CFG direction in ε-space
  $\widehat{\boldsymbol{x}}_{k+1} \leftarrow \boldsymbol{x}_{k+1} / \|\boldsymbol{x}_{k+1}\|_2$
  $\boldsymbol{P}_{k+1}^{\perp} \leftarrow I - \widehat{\boldsymbol{x}}_{k+1} \widehat{\boldsymbol{x}}_{k+1}^{\top}$ ▷ projector at $\boldsymbol{x}_{k+1}$
  $\boldsymbol{g}_k^{\perp} \leftarrow \boldsymbol{P}_{k+1}^{\perp} \boldsymbol{g}_k$ ▷ tangential component
  $\tilde{\boldsymbol{\varepsilon}}_k \leftarrow \boldsymbol{\varepsilon}_u + \omega \boldsymbol{g}_k + \eta \left(\frac{\langle \boldsymbol{\varepsilon}_c, \boldsymbol{g}_k^{\perp} \rangle}{\|\boldsymbol{g}_k^{\perp}\|_2^2} \boldsymbol{g}_k^{\perp} \right)$ ▷ TAG-augmented CFG
  $\boldsymbol{x}_k \leftarrow \textsc{Step}(\tilde{\boldsymbol{\varepsilon}}_k, t_{k+1}, \boldsymbol{x}_{k+1})$ ▷ scheduler step

Conditional–unconditional tangent reconciliation. Let $\boldsymbol{g}_k := \epsilon_\theta(\boldsymbol{x}_k, \boldsymbol{c}) - \epsilon_\theta(\boldsymbol{x}_k, \emptyset)$ be the score from a CFG guidance where $\epsilon_\theta(\cdot, \boldsymbol{c}), \epsilon_\theta(\cdot, \emptyset)$ denote the conditional and unconditional scores. We form a *conditional-relative tangent* by removing the unconditional tangent from the conditional one,

$$\boldsymbol{g}_k^{\perp} = \boldsymbol{P}^{\perp}(\boldsymbol{x}_k)\big(\epsilon_\theta(\boldsymbol{x}_k, \boldsymbol{c}) \; - \; \epsilon_\theta(\boldsymbol{x}_k, \emptyset)\big) = \boldsymbol{P}^{\perp}(\boldsymbol{x}_k)\boldsymbol{g}_k, \tag{19}$$

and *project* the conditional score $\epsilon_\theta(\boldsymbol{x}_k, \boldsymbol{c})$ onto this tangent subspace. We then amplify this condition relative tangent:

$$\tilde{\boldsymbol{\varepsilon}}_k \; = \; \epsilon_\theta(\boldsymbol{x}_k, \boldsymbol{c}) \; + \; \omega \boldsymbol{g}_k + \eta \left(\sigma_k^{-1} \boldsymbol{P}(\boldsymbol{g}_k^{\perp}) \epsilon_\theta(\boldsymbol{x}_k, \boldsymbol{c})\right), \tag{20}$$

where ω is the usual CFG scale and η controls the extra tangential emphasis.

5 Experiments

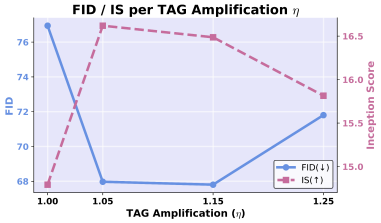
Backbones and inference setup. We apply TAG at inference on pretrained backbones, using Stable Diffusion v1.5 (Rombach et al., 2022) for major experiments and SD3 (Esser et al., 2024) for flow matching. Unconditional results are reported on ImageNet-1K *val* (Deng et al., 2009) with 30K samples per setting. Text-conditional results use MS-COCO 2014 *val* (Lin et al., 2015) with 10K prompts. The number of function evaluations (#NFE) follows each table. TAG is inserted after every solver update with amplification η. Metrics include FID (Heusel et al., 2017), IS (Salimans et al., 2016), CLIPScore (Hessel et al., 2021), and NFE. FID is computed with *pytorch-fid* (Seitzer, 2020), IS with Inception-V3 (Szegedy et al., 2016), and CLIPScore is computed with OpenAI CLIP ViT-L/14. All runs use fixed seeds and identical preprocessing to the corresponding baselines.

Improvements on conditional generation. Table 4 presents quantitative results on the MS-COCO, demonstrating that augmenting existing guidance samplers with TAG consistently yields substantial improvements in sample fidelity while largely preserving text-image alignment. Notably, TAG enables a 30-step sampling process to outperform the 100-step CFG baseline. Even in a condition only setting, TAG dramatically reduces FID and increases CLIPScore, confirming its *foundational benefits independent of a guidance signal*. Furthermore, this trend extends to other guidance techniques such as PAG and SEG, where TAG again reduces FID at the same computational cost. The qualitative improvements are visualized in Figure 7, which demonstrates TAG's ability to produce higher-fidelity images with fewer artifacts.

$\omega = 2.5, \eta = 0.0$ $\omega = 5.0, \eta = 0.0$ $\omega = 2.5, \eta = 1.0$

Figure 7: **Qualitative Results with CFG**. TAG produces higher-fidelity samples with fewer hallucinations, outperforming even baselines with a higher CFG scale ω.

Improvements on unconditional generation. For unconditional generation, TAG consistently improves sample quality across a range of models and samplers. As shown in Table 1, it reduces FID and increases IS at a matched NFEs. Notably, TAG acts as a 'plug-and-play' module for existing guidance methods (e.g., SAG, PAG, SEG), enhancing their performance without architectural changes or additional model evaluations. Moreover, TAG significantly pushes the compute–quality



(a) Impact of TAG amplification η on FID (↓) and IS (↑) for unconditional SD v1.5 generation.

$\eta = 1.0$ $\eta = 1.1$ $\eta = 1.2$ $\eta = 1.3$

(b) Qualitative comparison across amplification levels η for SD3 unconditional generation: moderate tangential amplification enhances detail and coherence, while excessive amplification degrades fidelity.

Figure 8: **Ablation on TAG amplification η.** Figure 8a and Table 1 show gains at moderate η and degradation when amplification is *excessive*. Figure 8b confirms the same trend for Flow-matching, underscoring the need to select an appropriate η.

frontier by enabling both faster inference and higher quality. With samplers like DDIM and DPM++, TAG can achieve superior results with as few as half the NFEs (Table 3). Concurrently, it substantially boosts performance on foundational models like SD v2.1 and SDXL at a fixed computational cost (Table 2). This dual benefit provides a practical path to faster inference and extends to SOTA models like SD3 (Table 5), with qualitative improvements visualized in Figures 5 and 9.

Improvements on Flow Matching. Figure 5 and Table 5 demonstrate that TAG transfers seamlessly to flow-matching backbones (Esser et al., 2024). Inserted as a lightweight tangential reweighting after each solver step, without architectural changes or additional function evaluations. TAG yields a modest but consistent FID improvement at matched compute and visibly reduces artifacts in unconditional samples. These results show TAG's potential to be model-agnostic across diverse architectures, including modern large-scale models such as SD3.

Table 5: **Quantitative results** for flow matching–based generator. TAG is directly applicable to flow matching. Evaluations are conducted on ImageNet *val* with 1K random seeds. FID is computed on five independently sampled ImageNet-1K subsets and reported as the mean.

Methods w/ SD3	Esser et al. (2024)	**+ TAG**
TAG Amp. (η)	–	1.1
#NFE	50	50
FID↓	153.35	**150.97**

6 Limitation & Future Work

An ablation of η reveals that moderate tangential amplification improves quality, whereas performance degrades for larger values (Fig. 8a, Tab. 1; see also Fig. 8b for flow matching). Analytically, the post-step state norm under TAG satisfies

$$\|\boldsymbol{x}_{k+1} + \Delta_{k+1}^{\mathrm{TAG}}\|_2^2 = \|\boldsymbol{x}_{k+1} + \Delta_{k+1}\|_2^2 + \underbrace{(\eta^2 - 1)\,\|\boldsymbol{P}_{k+1}^{\perp}\Delta_{k+1}\|_2^2}_{\text{additive term}}. \tag{21}$$

Therefore, for $\eta \approx 1$, the increase is small, thus the first-order *radial* (Tweedie) term is unchanged. As η grows, however, the *additive term* increasingly perturbs the scheduler's radial calibration, which explains the observed degradation. A promising direction is to model these higher-order effects and design *adaptive* gains η_k, potentially yielding a hyperparameter-free variant.

7 Conclusion

This paper introduces a new perspective for addressing the problem of hallucinations in diffusion models, demonstrating that the tangential component of the sampling update encodes critical semantic structure. Based on this geometric insight, we propose **T**angential **A**mplifying **G**uidance **(TAG)**, a practical, architecture-agnostic method that amplifies the tangential component. By doing so, TAG effectively steers the sampling trajectory toward higher-density regions of the data manifold, generating samples with fewer hallucinations and improved fidelity. Our method achieved good samples without requiring retraining or incurring any additional heavy computational overhead, offering a practical, plug-and-play solution for enhancing existing diffusion model backbones.

REFERENCES

Donghoon Ahn, Hyoungwon Cho, Jaewon Min, Wooseok Jang, Jungwoo Kim, SeonHwa Kim, Hyun Hee Park, Kyong Hwan Jin, and Seungryong Kim. Self-rectifying diffusion sampling with perturbed-attention guidance. In *European Conference on Computer Vision*, pp. 1–17. Springer, 2024.

Sumukh K Aithal, Pratyush Maini, Zachary C. Lipton, and J. Zico Kolter. Understanding hallucinations in diffusion models through mode interpolation. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 134614–134644. Curran Associates, Inc., 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/f29369d192b13184b65c6d2515474d78-Paper-Conference.pdf.

Brian DO Anderson. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3):313–326, 1982.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, 2009. doi: 10.1109/CVPR.2009.5206848.

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

Anh-Dung Dinh, Daochang Liu, and Chang Xu. Representative guidance: Diffusion model sampling with coherence. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=gWgaypDBs8.

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

Jack Hessel, Ari Holtzman, Maxwell Forbes, Ronan Le Bras, and Yejin Choi. Clipscore: A reference-free evaluation metric for image captioning. *arXiv preprint arXiv:2104.08718*, 2021.

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

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. In *NeurIPS 2021 Workshop on Deep Generative Models and Downstream Applications*, 2021. URL https://openreview.net/forum?id=qw8AKxfYbI.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Susung Hong. Smoothed energy guidance: Guiding diffusion models with reduced energy curvature of attention. *Advances in Neural Information Processing Systems*, 37:66743–66772, 2024.

Susung Hong, Gyuseong Lee, Wooseok Jang, and Seungryong Kim. Improving sample quality of diffusion models using self-attention guidance. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 7462–7471, 2023.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. *Advances in neural information processing systems*, 35:26565–26577, 2022.

Tero Karras, Miika Aittala, Tuomas Kynkäänniemi, Jaakko Lehtinen, Timo Aila, and Samuli Laine. Guiding a diffusion model with a bad version of itself. *Advances in Neural Information Processing Systems*, 37:52996–53021, 2024.

Mingi Kwon, Jaeseok Jeong, Yi Ting Hsiao, Youngjung Uh, et al. Tcfg: Tangential damping classifier-free guidance. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 2620–2629, 2025.

Tuomas Kynkäänniemi, Miika Aittala, Tero Karras, Samuli Laine, Timo Aila, and Jaakko Lehtinen. Applying guidance in a limited interval improves sample and distribution quality in diffusion models. *Advances in Neural Information Processing Systems*, 37:122458–122483, 2024.

Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, and Piotr Dollár. Microsoft coco: Common objects in context, 2015. URL https://arxiv.org/abs/1405.0312.

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. Dpm-solver++: Fast solver for guided sampling of diffusion probabilistic models. *Machine Intelligence Research*, pp. 1–22, 2025.

Maya Okawa, Ekdeep S Lubana, Robert Dick, and Hidenori Tanaka. Compositional abilities emerge multiplicatively: Exploring diffusion models on a synthetic task. *Advances in Neural Information Processing Systems*, 36:50173–50195, 2023.

Dustin Podell, Zion English, Kyle Lacey, Andreas Blattmann, Tim Dockhorn, Jonas Müller, Joe Penna, and Robin Rombach. SDXL: Improving latent diffusion models for high-resolution image synthesis. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=di52zR8xgf.

Javad Rajabi, Soroush Mehraban, Seyedmorteza Sadat, and Babak Taati. Token perturbation guidance for diffusion models. *arXiv preprint arXiv:2506.10036*, 2025.

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

Seyedmorteza Sadat, Otmar Hilliges, and Romann M. Weber. Eliminating oversaturation and artifacts of high guidance scales in diffusion models, 2025. URL https://arxiv.org/abs/2410.02416.

Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training gans. *Advances in neural information processing systems*, 29, 2016.

Maximilian Seitzer. pytorch-fid: FID Score for PyTorch. https://github.com/mseitzer/pytorch-fid, August 2020. Version 0.3.0.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, pp. 2256–2265. pmlr, 2015.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020a.

Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. *Advances in neural information processing systems*, 32, 2019.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020b.

Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826, 2016.

Maurice CK Tweedie et al. An index which distinguishes between some important exponential families. In *Statistics: Applications and new directions: Proc. Indian statistical institute golden Jubilee International conference*, volume 579, pp. 579–604, 1984.

Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural Computation*, 23(7):1661–1674, 2011. doi: 10.1162/NECO_a_00142.

APPENDIX

Symbol	Meaning
$\boldsymbol{x}_k \in \mathbb{R}^d$	latent at step k (time t_k).
$\widehat{\boldsymbol{x}} := \boldsymbol{x}/\|\boldsymbol{x}\|_2$	Unit vector in the direction of $\boldsymbol{x}$.
$\boldsymbol{P}_k := \widehat{\boldsymbol{x}}_k \widehat{\boldsymbol{x}}_k^\top$	Projector onto $\mathrm{span}(\boldsymbol{x}_k)$ ('normal' to the iso-noise surface at $\boldsymbol{x}$).
$\boldsymbol{P}_k^\perp := I - \boldsymbol{P}_k$	Tangential projector orthogonal to $\boldsymbol{x}_k$.
$\langle \boldsymbol{u}, \boldsymbol{v} \rangle,\ \|\boldsymbol{u}\|_2$	Euclidean inner product and norm.
$\boldsymbol{s}_\theta(x, t)$	Model score.
$\epsilon_\theta(x, t)$	Model noise prediction; $\boldsymbol{s}_\theta(x,t) = -\epsilon_\theta(x,t)/\sigma_t$.
$\boldsymbol{g}_k$	Guidance residual direction.
Δ_k	Base solver increment without TAG at step k.
Δ_k^{TAG}	TAG-modified increment at step k.
$\eta \geq 1$	TAG tangential amplification factor (scales $\boldsymbol{P}_k^\perp \Delta$).
NFE	Number of function evaluations.

A PROOF & DERIVATION

Proof for Theorem 4.1

Proof. Assume the deterministic base step $\Delta_{k+1} = \tilde{\alpha}_k\, \epsilon_\theta(\boldsymbol{x}_{k+1}, t_{k+1}) + \beta_k\, \boldsymbol{x}_{k+1}$, with $\tilde{\alpha}_k \leq 0$, and let $\boldsymbol{P}_{k+1}, \boldsymbol{P}_{k+1}^\perp$ be the orthogonal projectors with $\boldsymbol{P}_{k+1}\boldsymbol{x}_{k+1} = \boldsymbol{x}_{k+1}$ and $\boldsymbol{P}_{k+1}^\perp \boldsymbol{x}_{k+1} = \boldsymbol{0}$. Applying the projectors to the base decomposition gives

$$\boldsymbol{P}_{k+1}^\perp \Delta_{k+1} = \tilde{\alpha}_k\, \boldsymbol{P}_{k+1}^\perp \epsilon_\theta(\boldsymbol{x}_{k+1}, t_{k+1}), \qquad \boldsymbol{P}_{k+1}\Delta_{k+1} = \tilde{\alpha}_k\, \boldsymbol{P}_{k+1}\epsilon_\theta(\boldsymbol{x}_{k+1}, t_{k+1}) + \beta_k\, \boldsymbol{x}_{k+1}. \tag{22}$$

Therefore, the TAG update rule step is

$$\Delta_{k+1}^{\mathrm{TAG}} = \big(\boldsymbol{P}_{k+1} + \eta\, \boldsymbol{P}_{k+1}^\perp\big)\Delta_{k+1} = \tilde{\alpha}_k \big[\boldsymbol{P}_{k+1} + \eta \boldsymbol{P}_{k+1}^\perp\big]\epsilon_\theta(\boldsymbol{x}_{k+1}, t_{k+1}) + \beta_k \boldsymbol{x}_{k+1}. \tag{23}$$

The first-order Taylor gain with respect to TAG update at t_{k+1} is defined as:

$$\begin{aligned} G(\eta) &:= \big(\Delta_{k+1}^{\mathrm{TAG}}\big)^\top \nabla_{\boldsymbol{x}} \log p(\boldsymbol{x} \mid t_{k+1})\big|_{\boldsymbol{x}=\boldsymbol{x}_{k+1}} \\ &= \Big(\big(\boldsymbol{P}_{k+1} + \eta\, \boldsymbol{P}_{k+1}^\perp\big)\Delta_{k+1}\Big)^\top \nabla_{\boldsymbol{x}} \log p(\boldsymbol{x} \mid t_{k+1})\big|_{\boldsymbol{x}=\boldsymbol{x}_{k+1}} \end{aligned} \tag{24}$$

We analyze this gain by approximating the true score with the model's score function

$$\boldsymbol{s}_\theta(\boldsymbol{x}_{k+1}, t_{k+1}) = -\sigma_{k+1}^{-1} \epsilon_\theta(\boldsymbol{x}_{k+1}, t_{k+1}), \tag{25}$$

thus:

$$\begin{aligned} G(\eta) &= \Big(\big(\boldsymbol{P}_{k+1} + \eta\, \boldsymbol{P}_{k+1}^\perp\big)\Delta_{k+1}\Big)^\top \nabla_{\boldsymbol{x}} \log p(\boldsymbol{x} \mid t_{k+1})\big|_{\boldsymbol{x}=\boldsymbol{x}_{k+1}} \\ &\approx \Big(\big(\boldsymbol{P}_{k+1} + \eta\, \boldsymbol{P}_{k+1}^\perp\big)\Delta_{k+1}\Big)^\top \boldsymbol{s}_\theta(\boldsymbol{x}_{k+1}, t_{k+1}) \\ &= -\sigma_{k+1}^{-1} \cdot \Big(\big(\boldsymbol{P}_{k+1} + \eta\, \boldsymbol{P}_{k+1}^\perp\big)\Delta_{k+1}\Big)^\top \epsilon_\theta\big(\boldsymbol{x}_{k+1}, t_{k+1}\big) \end{aligned} \tag{26}$$

Substitute equation 23 into equation 26, then:

$$\begin{aligned} G(\eta) &\approx -\sigma_{k+1}^{-1}\Big(\tilde{\alpha}_k \boldsymbol{P}_{k+1}\epsilon_\theta + \beta_k \boldsymbol{x}_{k+1} + \eta \tilde{\alpha}_k \boldsymbol{P}_{k+1}^\perp \epsilon_\theta\Big)^\top \epsilon_\theta \\ &= -\sigma_{k+1}^{-1}\Big(\tilde{\alpha}_k (\boldsymbol{P}_{k+1}\epsilon_\theta)^\top \epsilon_\theta + \beta_k \boldsymbol{x}_{k+1}^\top \epsilon_\theta + \eta \tilde{\alpha}_k (\boldsymbol{P}_{k+1}^\perp \epsilon_\theta)^\top \epsilon_\theta\Big). \end{aligned} \tag{27}$$

Since $\boldsymbol{P}$ and $\boldsymbol{P}^\perp$ are symmetric and idempotent, thus

$$\boldsymbol{v}^\top \boldsymbol{P}\boldsymbol{v} = \|\boldsymbol{P}\boldsymbol{v}\|_2^2 \tag{28}$$

is established. Therefore,

$$G(\eta) \approx -\sigma_{k+1}^{-1}\left(\tilde{\alpha}_k \left\|\boldsymbol{P}_{k+1}\epsilon_\theta\right\|_2^2 + \beta_k \boldsymbol{x}_{k+1}^\top \epsilon_\theta + \eta\tilde{\alpha}_k \left\|\boldsymbol{P}_{k+1}^\perp \epsilon_\theta\right\|_2^2\right). \tag{29}$$

Differentiating the gain $G(\eta)$ in equation 29 with respect to η yields:

$$\frac{\partial G(\eta)}{\partial \eta} \approx \frac{-\tilde{\alpha}_k}{\sigma_{k+1}} \left|\boldsymbol{P}_{k+1}^\perp \epsilon_\theta(\boldsymbol{x}_{k+1}, t_{k+1})\right|_2^2 \;\geq\; 0. \tag{30}$$

This derivative is guaranteed to be non-negative, since the DDIM sampler coefficient $\tilde{\alpha}_k \leq 0$ by definition, while σ_{k+1} and the squared L2-norm are strictly non-negative. This proves that the first-order gain $G(\eta)$ is a monotonically non-decreasing function of η. Consequently, amplifying the tangential component of the update step via TAG is guaranteed to improve the first-order log-likelihood gain compared to the base update step.

Analysis on pure TAG gain. Subtracting each gain $G^{\text{base}} \triangleq G(\eta = 1)$ and $G^{\text{TAG}} \triangleq G(\eta > 1)$,

$$\begin{aligned}
&\overbrace{\Big(-\sigma_{k+1}^{-1}\cdot\Big(\Delta_{k+1}^{\text{TAG}}\Big)^\top \epsilon_\theta\big(\boldsymbol{x}_{k+1}, t_{k+1}\big)\Big)}^{\text{TAG update gain, } G^{\text{TAG}}} \;-\; \overbrace{\Big(-\sigma_{k+1}^{-1}\cdot\Big(\Delta_{k+1}\Big)^\top \epsilon_\theta\big(\boldsymbol{x}_{k+1}, t_{k+1}\big)\Big)}^{\text{base update gain, } G^{\text{base}}} \\
&\qquad = -\sigma_{k+1}^{-1}\cdot\big(\Delta_{k+1}^{\text{TAG}} - \Delta_{k+1}\big)^\top \epsilon_\theta\big(\boldsymbol{x}_{k+1}, t_{k+1}\big) \\
&\qquad = -\sigma_{k+1}^{-1}\cdot\big((\eta-1)\,\boldsymbol{P}_{k+1}^\perp \Delta_{k+1}\big)^\top \epsilon_\theta\big(\boldsymbol{x}_{k+1}, t_{k+1}\big).
\end{aligned} \tag{31}$$

Using $\Delta_{k+1} = \tilde{\alpha}_k\, \epsilon_\theta(\boldsymbol{x}_{k+1}, t_{k+1}) + \beta_k\, \boldsymbol{x}_{k+1}$, $\boldsymbol{P}_{k+1}^\perp$ be:

$$\boldsymbol{P}_{k+1}^\perp \Delta_{k+1} = \boldsymbol{P}_{k+1}^\perp \tilde{\alpha}_k\, \epsilon_\theta(\boldsymbol{x}_{k+1}, t_{k+1}). \tag{32}$$

Thus, substitute equation 32 into equation 31 then:

$$G^{\text{TAG}} - G^{\text{base}} = \underbrace{-\sigma_{k+1}^{-1}\cdot\big(\tilde{\alpha}_k(\eta-1)\big)}_{\text{scalar}} \cdot \big(\boldsymbol{P}_{k+1}^\perp\, \epsilon_\theta(\boldsymbol{x}_{k+1}, t_{k+1})\big)^\top \epsilon_\theta\big(\boldsymbol{x}_{k+1}, t_{k+1}\big). \tag{33}$$

This simplifies to the final quadratic form:

$$G^{\text{TAG}} - G^{\text{base}} = \underbrace{-\sigma_{k+1}^{-1}\cdot\big(\tilde{\alpha}_k(\eta-1)\big)}_{\geq\, \mathbf{0} \text{ as } \tilde{\alpha}_k \leq 0} \cdot \left\|\boldsymbol{P}_{k+1}^\perp \epsilon_\theta(\boldsymbol{x}_{k+1}, t_{k+1})\right\|_2^2, \tag{34}$$

This proves that the difference in gain is non-negative for any $\eta \geq 1$. Therefore, the first-order log-likelihood gain of the TAG update is always greater than or equal to that of the base update, with equality holding if and only if $\eta = 1$ or the tangential component of the score is zero. □

B Implementation of the Tangential Amplifying Guidance

Code: Tangential Amplifying Guidance (TAG)

```
output = scheduler.step(noise_pred, t, latents, return_dict=False)

if apply_tag:
    post = latents
    eta_v, eta_n = t_guidance_scale, 1

    v_t = post / (post.norm(p=2, dim=(1,2,3), keepdim=True) + 1e-8)

    latents = output
    delta  = latents - post
    a      = (delta * v_t).sum(dim=(1,2,3), keepdim=True)

    u_n = a * v_t
    u_t = delta - u_n
    latents = post + eta_v * u_t + eta_n * u_n
else:
    latents = output
```

Code: Conditional Tangential Amplifying Guidance (C-TAG)

```
def proj_par(z, n):
    return (z * n).sum(dim=(1,2,3), keepdim=True) * n

def proj(z, v):
    v = v / (v.norm(p=2, dim=(1,2,3), keepdim=True) + 1e-8)
    return (z * v).sum(dim=(1,2,3), keepdim=True) * v

eps_u, eps_c = HeadToEps(noise_pred, latents, t, scheduler, do_cfg)

s_u = -eps_u / (sigma + 1e-12)
s_c = -eps_c / (sigma + 1e-12)

n = latents / (latents.norm(p=2, dim=(1,2,3), keepdim=True) + 1e-8)

g         = s_c - s_u
t_c       = s_c - proj_par(s_c, n)
t_u       = s_u - proj_par(s_u, n)
g_aligned = proj(s_c, t_c - t_u)
g         = g + t_guidance_scale * g_aligned

s_star    = s_u + guidance_scale * g
eps       = -sigma * s_star

model_out = EpsToHead(eps, latents, t, scheduler)
latents   = scheduler.step(model_out, t, latents, return_dict=False)
```

C Additional Qualitative Results

Stable Diffusion 1.5

Stable Diffusion 2.1

Stable Diffusion XL

Stable Diffusion 3

Figure 9: **Qualitative results for unconditional generation across backbones.** For each model (SD1.5/2.1 (Rombach et al., 2022), SDXL (Podell et al., 2024), SD3 (Esser et al., 2024)), the top row shows baseline sampling and the bottom row shows +TAG at matched NFE. TAG yields sharper, more coherent structure with fewer artifacts while preserving diversity.

D LLM USAGE DISCLOSURE

During the preparation of this paper, the authors made limited use of large language models (LLMs) for polishing the writing, grammar refinement and LaTeX formatting. LLMs were not used for generating research ideas, designing or conducting experiments, analyzing results, or formulating conclusions. All scientific content and contributions are entirely the responsibility of the authors, and any LLM-assisted text was carefully reviewed and revised before inclusion.