DENOISING AS PROJECTION: CONSTRAINED OPTIMIZATION WITH GRADIENT-GUIDED DIFFUSION

Anonymous authors
Paper under double-blind review

ABSTRACT

Diffusion models are increasingly used not only for sampling from learned data distributions, but also for generating samples that optimize task-specific objectives. A common approach is to guide the reverse diffusion process using gradients of an external objective. However, when the data distribution is supported on a structured feasible set, such as a manifold or a constraint set, gradient guidance can move samples away from the learned data geometry. In this paper, we study a simple projected-gradient-guided diffusion update based on the observation that the Stein denoising operator can act as an approximate projection onto the data geometry. The proposed update incorporates the objective gradient inside the denoising step, yielding an inference-time method that uses only a pretrained denoiser and gradient evaluations. We analyze this update as an inexact projected-gradient method for constrained optimization over learned feasible geometries. Our theory covers three settings: linear manifolds, compact convex feasible sets, and compact Riemannian submanifolds. In all these settings, we prove descent and finite-time convergence guarantees. Numerical experiments support the theoretical interpretation and illustrate how the proposed update balances objective descent with preservation of the learned geometry.

1 INTRODUCTION

Diffusion models achieve state-of-the-art performance in image synthesis, inverse problems, scientific design, and decision-making Sohl-Dickstein et al. (2015); Ho et al. (2020); Song et al. (2021b;a). Beyond realistic sampling, many applications require samples that optimize a task-specific objective while remaining consistent with a learned distribution, including inverse problems Chung et al. (2023), trajectory optimization Janner et al. (2022), molecular generation Wu et al. (2022), and control Chi et al. (2025).

A common strategy is to modify reverse diffusion with auxiliary information, as in classifier and classifier-free guidance Dhariwal & Nichol (2021); Ho & Salimans (2021). Given a differentiable objective $f : \mathbb{R}^n \to \mathbb{R}$, its gradient can instead guide the reverse dynamics. This suggests viewing guided diffusion as solving $\min_{x\in\mathcal{X}} f(x)$, where the support or high-density region of a learned distribution implicitly represents the feasible set $\mathcal{X} \subseteq \mathbb{R}^n$. The diffusion model then provides a prior over feasible designs, trajectories, or decisions, while f specifies task performance. Although gradient guidance needs only inference-time evaluations of ∇f Guo et al. (2024), directly injecting the gradient may disrupt the learned feasibility structure.

The difficulty is geometric. The set $\mathcal{X}$ may be a low-dimensional manifold, a closed convex set, a set of physically admissible designs, or a family of dynamically feasible, collision-free trajectories. A gradient direction need not be tangent or feasible, so naive guidance may lower the objective while degrading sample quality, physical validity, or constraint satisfaction. This is especially consequential in diffusion-based robotic planning: learned trajectory distributions encode useful behavioral and dynamical priors, yet objective guidance can produce unsafe or infeasible trajectories without an additional feasibility mechanism Xiao et al. (2025); Shaoul et al. (2025); Liang et al. (2025; 2026).

Existing approaches often add sampling, computation, or modeling complexity. For example, Kharitenko et al. (2025) computes score Jacobians to construct geometric primitives, while Guo et al. (2024) uses repeated batch generation and objective linearization. Related denoiser-as-projection analysis proves recovery for linear inverse problems with quadratic objectives under an RIP-type condition, rather than general constrained optimization Leong & Traonmilin (2026). Manifold-

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.

Table 1: Comparison of theoretical scope and inference-time requirements. Sublinear rates are best-iterate guarantees; constants and finite-noise or denoising-error terms are omitted.

Method	Geometry	Guarantee	Rate	Score requirement	Sample requirement
Guo et al. (2024)	Linear–Gaussian	Regularized optimality of the distributional mean	Geometric	Jacobian of score function	Multiple reverse-diffusion trajectories
Kharitenko et al. (2025)	Compact C^3 Riemannian manifold	Riemannian stationarity	$O(T^{-1/2})$	Jacobian of score function	No auxiliary inference-time samples
DCG (ours)	Linear–Gaussian	Global optimality; strongly convex f	Geometric	Don't require Jacobian	Single reverse trajectory; no auxiliary samples
	Compact convex set	Global optimality; strongly convex f	Geometric		
		Projected stationarity; smooth nonconvex f	$O(T^{-1/2})$		
	Compact C^2 Riemannian manifold	Riemannian stationarity; smooth nonconvex f	$O(T^{-1/2})$		

aware guidance Chung et al. (2022); He et al. (2024); Chung et al. (2025); Zirvi et al. (2025), constrained diffusion Christopher et al. (2024); Li & Beeson (2025); Khalafi et al. (2024); Hadou et al. (2026), and reward alignment Clark et al. (2024); Deng et al. (2024); Potaptchik et al. (2026a); Pachebat et al. (2025); Kim et al. (2025); Azangulov et al. (2026) may instead require fine-tuning, repeated sampling, or auxiliary models. Appendix A provides detailed comparisons.

This raises a natural question: *Can an update using one gradient evaluation and one pretrained Stein denoising call per step provide optimization guarantees without outer-loop generation, score Jacobians, or model fine-tuning?*

We answer affirmatively by applying the denoiser after the gradient update. The gradient step promotes objective descent, while denoising returns the iterate toward the feasible geometry. When the denoiser approximates the true projection, the update behaves as an inexact projected-gradient method.

Contributions. First, we introduce *Denoising-Corrected Gradient* (DCG) Guidance, which applies an objective-gradient step followed by one pretrained-denoiser call at each reverse step. DCG follows a single trajectory and requires no repeated batch generation, score-model adaptation Guo et al. (2024), or denoiser differentiation Kharitenko et al. (2025).

Second, we establish quantitative bounds between the Stein posterior-mean denoiser and the relevant projection for linear-Gaussian distributions, compact convex sets, and compact smooth submanifolds. These bounds identify the denoiser as DCG's geometric correction and the reverse dynamics as a time-varying inexact projected-gradient method.

Third, we prove finite-time guarantees tailored to each geometry and objective class. For linear-Gaussian data and strongly convex objectives, the projected objective error and normal component contract geometrically. For compact convex sets, we prove geometric convergence for strongly convex objectives and an $O(1/\sqrt{T})$ nonconvex best-iterate rate. For compact smooth submanifolds, we obtain the same best-iterate rate locally. This unified approximate-projection framework covers linear subspaces, possibly nonsmooth convex sets, and smooth manifolds. Table 1 compares our theory with the closest optimization-oriented approaches.

Finally, experiments on synthetic constrained optimization, motion planning, and D4RL diffusion-based reinforcement learning benchmarks (Appendix B) show that DCG favorably balances objective value and constraint satisfaction, outperforming standard gradient guidance and matching or improving related optimization-oriented diffusion baselines.

Limitations. Our analysis assumes an exact Stein posterior-mean denoiser and learned data support that faithfully represents the feasible set. Projection-error terms can capture learned-score error,

but we leave a full treatment of misspecification to future work. We also restrict the theory to deterministic DDIM dynamics under variance-exploding diffusion with an exponentially decaying noise schedule. Other samplers, parameterizations, and schedules require tracking their time-varying relaxation and denoising errors.

2 PRELIMINARIES

Constrained optimization over learned data geometry. We consider the constrained optimization problem

$$\min_{x\in\mathcal{X}} f(x), \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable and $\mathcal{X} \subseteq \mathbb{R}^n$ is the feasible geometry represented by the data distribution. The set $\mathcal{X}$ may be a linear subspace, compact convex set, or smooth embedded submanifold. Given a diffusion model trained on samples supported on or near $\mathcal{X}$, we use the pretrained model to approximately solve (1) at inference time.

If the Euclidean projector $\pi_{\mathcal{X}}$ were available, one could use projected gradient descent: $x^+ = \pi_{\mathcal{X}}\big(x - \eta\nabla f(x)\big)$. Here, however, the learned distribution represents $\mathcal{X}$ only implicitly, so $\pi_{\mathcal{X}}$ may be unavailable. Our premise is that diffusion denoising provides a learned approximate projection.

Gaussian diffusion and Stein denoising. We use a variance-exploding forward process

$$x_{t+1} = x_t + \sqrt{\alpha_t} w_t, \;\; w_t \sim \mathcal{N}(0, I_n),$$

where $\alpha_t > 0$ is the noise variance added at step t. Equivalently, [1]

$$x_t = x_0 + \sigma_t \overline{w}_t, \;\; \overline{w}_t \sim \mathcal{N}(0, I_n), \qquad \sigma_t^2 := \textstyle\sum_{s=0}^{t-1} \alpha_s. \tag{2}$$

Let P_t be the marginal distribution of x_t, with score $\nabla \log P_t(x)$. The Tweedie–Miyasawa identity Robbins (1992); Miyasawa et al. (1961) gives $\mathbb{E}[x_0 \mid x_t = x] = x + \sigma_t^2 \nabla \log P_t(x)$. We therefore define the Stein posterior-mean denoiser

$$\widehat{\pi}_\sigma(x) := x + \sigma^2 \nabla \log P_\sigma(x), \tag{3}$$

where P_σ is the data distribution convolved with $\mathcal{N}(0, \sigma^2 I_n)$. At step t, write

$$\widehat{\pi}_t(x) := x + \sigma_t^2 \nabla \log P_t(x). \tag{4}$$

In particular, $\widehat{\pi}_t(x_t) = \mathbb{E}[x_0 \mid x_t]$. In standard implementations, $\widehat{\pi}_t(x_t)$ is the posterior-mean or clean-sample prediction $\widehat{x}_0(x_t, t)$. Evaluating $\widehat{\pi}_t$ therefore requires no model or optimization subroutine beyond a standard sampler's denoising operation.

When clean data lie on a structured set $\mathcal{X}$, $\widehat{\pi}_t$ moves noisy inputs toward that geometry. We quantify this behavior by comparing it with Euclidean projection onto linear or convex sets and nearest-point projection onto smooth submanifolds.

Deterministic backward diffusion. Following DDIM Song et al. (2021a), we consider the deterministic reverse update

$$\begin{aligned} x_{t-1} &= \left(1 - \tfrac{\sigma_{t-1}}{\sigma_t}\right) \mathbb{E}[x_0 \mid x_t] + \tfrac{\sigma_{t-1}}{\sigma_t} x_t \\ &= \tfrac{\sigma_{t-1}}{\sigma_t} x_t + \left(1 - \tfrac{\sigma_{t-1}}{\sigma_t}\right) \widehat{\pi}_t(x_t). \end{aligned} \tag{5}$$

Thus, each reverse step interpolates between the noisy sample x_t and its denoised estimate $\widehat{\pi}_t(x_t)$. This update is also structurally similar to Delbracio & Milanfar (2023).

Gradient-guided diffusion. To bias samples toward a differentiable objective $f : \mathbb{R}^n \to \mathbb{R}$, standard gradient guidance adds a descent direction to (5) (cf. Dhariwal & Nichol (2021); Janner et al. (2022); Chung et al. (2023)):

$$x_{t-1} = \tfrac{\sigma_{t-1}}{\sigma_t} x_t + \left(1 - \tfrac{\sigma_{t-1}}{\sigma_t}\right) \widehat{\pi}_t(x_t) - \eta_t \nabla f(x_t). \tag{6}$$

[1] **Relation to variance-preserving diffusion:** The additive form (2) simplifies the analysis. A standard variance-preserving process is $\widetilde{x}_t = \sqrt{\overline{\alpha}_t} x_0 + \sqrt{1 - \overline{\alpha}_t}\, w, \;\; w \sim \mathcal{N}(0, I_n)$. Rescaling by $x_t := \frac{\widetilde{x}_t}{\sqrt{\overline{\alpha}_t}}$, we obtain $x_t = x_0 + \sigma_t w, \;\; \sigma_t^2 = \frac{1-\overline{\alpha}_t}{\overline{\alpha}_t}$. therefore converts variance-preserving diffusion into the additive-noise notation used below.

Algorithm 1 Denoising-Corrected Gradient Guidance (DCG Guidance)

Require: Initial noisy sample x_T, noise schedule $\{\sigma_t\}_{t=0}^{T}$, stepsizes $\{\eta_t\}_{t=1}^{T}$, denoisers $\{\widehat{\pi}_t\}_{t=1}^{T}$

for $t = T, T-1, \ldots, 1$ **do**
  Compute the gradient $\nabla f(x_t)$
  Set $z_t = x_t - \eta_t \nabla f(x_t)$
  Denoise/project z_t using $\widehat{\pi}_t(z_t)$
  Update $x_{t-1} = \frac{\sigma_{t-1}}{\sigma_t} x_t + \left(1 - \frac{\sigma_{t-1}}{\sigma_t}\right) \widehat{\pi}_t(z_t)$
end for
return x_0

We call (6) *post-denoising gradient guidance* (PDG) because it applies the gradient correction outside the denoiser. PDG requires only evaluating ∇f and does not retrain the diffusion model.

If the data lie on a structured feasible set, however, the gradient may violate the learned geometry and cause off-manifold drift. We instead apply the Stein denoiser, viewed as an approximate projection, to a gradient-updated point. The next section introduces this *denoising-corrected gradient (DCG) guidance* update.

3 Denoising-Corrected Gradient Guidance

The Stein denoiser $\widehat{\pi}_t(x) = x + \sigma_t^2 \nabla \log P_t(x)$ estimates the clean sample's posterior mean. When the data lie on or near a constraint set $\mathcal{X}$ or manifold $\mathcal{M}$, it maps noisy points toward the feasible geometry and thus acts as a learned approximate projection Kharitenko et al. (2025); Permenter & Yuan (2024); Chung et al. (2022).

Lemmas 1, 2, and 5 formalize this interpretation for linear manifolds, compact convex sets, and compact C^2 embedded submanifolds, respectively. It suggests reversing the order used by PDG: apply the gradient step before denoising. Specifically, we consider

$$x_{t-1} = \frac{\sigma_{t-1}}{\sigma_t} x_t + \left(1 - \frac{\sigma_{t-1}}{\sigma_t}\right) \widehat{\pi}_t\big(x_t - \eta_t \nabla f(x_t)\big), \tag{7}$$

where $\eta_t > 0$ is the guidance stepsize.

Unlike PDG, which shifts the denoised prediction, (7) takes a gradient step before applying the Stein denoiser. When $\widehat{\pi}_t$ approximates the true projection, the update resembles relaxed projected gradient:

$$x_{t-1} \approx (1 - \epsilon_t) x_t + \epsilon_t \pi_{\mathcal{X}}\big(x_t - \eta_t \nabla f(x_t)\big).$$

This correspondence underlies our optimization analysis.

Each reverse step requires one evaluation of $\nabla f(x_t)$ and one application of the pretrained denoiser. DCG neither differentiates through $\widehat{\pi}_t$, evaluates $\nabla_x f(\widehat{\pi}_t(x))$, solves an auxiliary constrained problem, nor retrains the diffusion model. The following sections prove convergence when the Stein denoiser accurately approximates projection.

For the analysis, we use the geometric schedule $\sigma_{t-1}/\sigma_t = 1-\epsilon$ and constant guidance scale $\eta_t = \eta$, yielding

$$x_{t-1} = (1 - \epsilon) x_t + \epsilon \widehat{\pi}_t\big(x_t - \eta \nabla f(x_t)\big). \tag{8}$$

This choice isolates the core mechanism without time-dependent notation; we leave more general schedules and guidance scales to future work.

Comparison with related optimization methods. The update of Guo et al. (2024) shares PDG's post-denoising structure but evaluates the gradient at a sampled population mean: $x_{t-1}^{k+1} = (1 - \epsilon_t) x_t^{k+1} + \epsilon_t \widehat{\pi}_t(x_t^{k+1}) - \eta_t \nabla f(\bar{x}^k)$. Here, $\bar{x}^k := \frac{1}{B} \sum_{b=1}^{B} x_0^{k,b}$, where $\{x_0^{k,b}\}_{b=1}^{B}$ are terminal samples from B reverse trajectories at outer iteration k. The next batch shares a guidance direction updated after estimating $\bar{x}^k$ from new samples. DCG instead uses the current state and places the gradient step inside the denoiser, avoiding batch averaging and outer guidance updates.

The method of Kharitenko et al. (2025) uses the posterior-mean denoiser and its Jacobian: $x^{k+1} = \widehat{\pi}_\sigma\big(x^k - \eta_k D\widehat{\pi}_\sigma(x^k) \nabla f(x^k)\big)$. Here, $D\widehat{\pi}_\sigma(x^k)$ approximates tangent-space projection and $\widehat{\pi}_\sigma$ approximates retraction. DCG uses only $\widehat{\pi}_t$ and follows a time-varying reverse schedule rather than standalone iterations at a fixed noise level.

The gradient–denoiser composition also resembles plug-and-play proximal gradient Venkatakrishnan et al. (2013); Sreehari et al. (2016) and RED Romano et al. (2017). Our contribution is to identify the time-varying Stein posterior mean along a reverse trajectory as an approximate projection onto the learned feasible set and derive constrained-optimization guarantees.

4 Convergence Analysis: Linear-Gaussian Setting

We first analyze the proposed guided diffusion update in a linear setting, where the data distribution is supported on a linear subspace and the Stein denoising operator can be characterized explicitly.

Let $A \in \mathbb{R}^{d\times n}$ satisfy $AA^\top = I_d$, so that $\mathrm{range}(A^\top) \subseteq \mathbb{R}^n$ is a d-dimensional linear subspace. We assume that the clean data distribution is a Gaussian supported on this subspace. All proofs are in Appendix C.

Assumption 1 (Linear data manifold)**.** *The clean data distribution satisfies $x_0 \sim \mathcal{N}(0, A^\top \Sigma A)$, where $\Sigma \in \mathbb{R}^{d\times d}$ is positive definite and $A \in \mathbb{R}^{d\times n}$ satisfies $AA^\top = I_d$.*

Throughout this section, we use $A^\perp \in \mathbb{R}^{(n-d)\times n}$ to denote the matrix whose rows form an orthonormal basis of $\ker(A)$. We also assume that the objective is strongly convex and smooth.

Assumption 2. *The function $f : \mathbb{R}^n \to \mathbb{R}$ is μ-strongly convex and L-smooth.*

The next lemma gives an explicit expression for the Stein denoising operator in this linear-Gaussian model. It shows that $\widehat{\pi}_t$ is not exactly the Euclidean projection onto $\mathrm{range}(A^\top)$ at finite noise, but rather a noise-dependent Tikhonov-regularized projection onto the data subspace. As $\sigma_t \to 0$, this operator approaches the Euclidean projection onto $\mathrm{range}(A^\top)$.

Lemma 1 (Characterization of the Stein denoising operator $\widehat{\pi}_t$)**.** *Under Assumption 1, the Stein denoising operator $\widehat{\pi}_t$ defined in (4) satisfies*

$$\widehat{\pi}_t(x) = (A^\top \Sigma A + \sigma_t^2 I)^{-1} A^\top \Sigma A x = \arg\min_{z\in \mathrm{range}(A^\top)} \|x - z\|^2 + \sigma_t^2 \|Az\|^2_{\Sigma^{-1}}.$$

Lemma 1 shows that, in the linear-Gaussian setting, the denoiser has an exact optimization interpretation: it returns the point in the data subspace that balances proximity to the noisy input with a noise-dependent quadratic regularization induced by the data covariance. This motivates the following analysis of the guided diffusion dynamics. The key idea is that, after the gradient step, the Stein denoiser acts as a regularized projection back toward the linear data manifold.

We now state the convergence result of DCG. The proof uses a time-dependent Lyapunov function that combines objective suboptimality on the data subspace, the normal component away from the subspace, and an explicit error term induced by the finite diffusion noise.

Theorem 1. *We study the guided diffusion dynamics specified by (8) under Assumptions 1 and 2. Let $t_0 = \max\{t : \sigma_t^2 \le \lambda_{\min}(\Sigma)(1-c)/c\}$, where $c \in (0,1)$ is an arbitrary constant. We further assume a constant guidance step size satisfying $\eta_t = \eta < 1/(\mu c)$ and $\eta\epsilon \le 1/L$ for all $t \le t_0$. Define*

$$\Phi_t(x) := f(x) + \tfrac{\sigma_t^2 \eta^{-1}}{2} x^\top A^\top \Sigma^{-1} A x, \qquad x_t^\star = \arg\min_{x\in\mathrm{range}(A^\top)} \Phi_t(x) \tag{9}$$

and

$$V_t(x) := (\Phi_t(A^\top A x) - \Phi_t(x_t^\star)) + \tfrac{\eta L^2}{2(1-\eta\mu c)} \|A^\perp x\|^2 + \tfrac{\eta^{-1} D^2}{2(1-\eta\mu c)} \sigma_t^2,$$

where $D = \max_{t\le t_0} \|Ax_t^\star\|_{\Sigma^{-1}}$. Then running (8) starting from t_0 we have

$$V_{t-1}(x_{t-1}) \le (1 - \eta\epsilon\mu c) V_t(x_t), \ \ \forall t \le t_0$$

Hence, letting $x^\star := \arg\min_{x\in\mathrm{range}(A^\top)} f(x)$, the projected component satisfies

$$f(A^\top A x_0) - f(x^\star) \le (1-\eta\epsilon\mu c)^{t_0} V_{t_0}(x_{t_0}) + (1-\epsilon)^{2t_0} \tfrac{\sigma_T^2}{2\eta} \|Ax^\star\|^2_{\Sigma^{-1}},$$

while the normal component satisfies $\|A^\perp x_0\| = (1-\epsilon)^{t_0} \|A^\perp x_{t_0}\|$.

Comparison with prior linear-Gaussian theory. The closest linear-Gaussian analysis is Guo et al. (2024), which proves convergence of the generated-distribution mean to a data-regularized optimum, with the unregularized target requiring adaptive score updates. Algorithmically, it relies on repeated batch generation and objective linearization, whereas our method uses a single reverse trajectory with one objective-gradient evaluation per step, making it substantially more sample efficient.

5 Convergence Analysis: Convex Constraint Setting

We next consider the case where the learned data geometry is a compact convex feasible region $\mathcal{X} \subset \mathbb{R}^n$. This is a natural intermediate case between the linear-Gaussian model and general data geometries. The exact projection $\pi_{\mathcal{X}}$ is well-defined and nonexpansive, but the geometry may be nonsmooth. All proofs are in Appendix D.

Assumption 3 (Data model on a compact convex set in an affine subspace)**.** *Let $\mathcal{A} \subseteq \mathbb{R}^n$ be a d-dimensional affine subspace, and let $\mathcal{X} \subseteq \mathcal{A}$ be compact and convex with nonempty relative interior in $\mathcal{A}$. Suppose that X is supported on $\mathcal{X}$ with law μ satisfying*

$$d\mu(q) = \vartheta(q)\, d\mathcal{H}^d(q),$$

where $\mathcal{H}^d$ denotes the d-dimensional Hausdorff measure restricted to $\mathcal{A}$. Assume that there exist constants $0 < \vartheta_- \le \vartheta_+ < \infty$ such that $\vartheta_- \le \vartheta(q) \le \vartheta_+, \qquad q \in \mathcal{X}$.

We first state the projection-approximation property of the Stein denoiser in this convex setting. Under Assumption 3, the denoiser is uniformly close to the Euclidean projection onto $\mathcal{X}$, with an error proportional to the noise level. This result is the key ingredient that allows us to transfer projected-gradient arguments to the guided diffusion dynamics.

Lemma 2 (Stein score approximates projection onto a compact convex set)**.** *Under Assumption 3, there exists a constant $C_{\mathcal{X}} > 0$, depending only on d and ϑ_+/ϑ_-, such that, for every $x \in \mathbb{R}^n$ and every $\sigma > 0$, the Stein denoising operator $\widehat{\pi}_\sigma$ defined in (3) satisfies*

$$\left\|\widehat{\pi}_\sigma(x) - \pi_{\mathcal{X}}(x)\right\| \le C_{\mathcal{X}}\sigma, \qquad \widehat{\pi}_\sigma(x) \in \mathcal{X}.$$

Lemma 2 shows that, in the small-noise regime, replacing the exact projection $\pi_{\mathcal{X}}$ by the Stein denoiser introduces a controlled projection error. We now use this error bound to prove convergence guarantees for the proposed update. We first consider strongly convex objectives, and then turn to nonconvex objectives.

5.1 Convergence for Strongly Convex Objectives

For strongly convex objectives, the analysis uses a Lyapunov function that combines the projected iterate with control of the distance to the feasible region. Specifically, the following one-step lemma tracks both the projected component $p_t = \pi_{\mathcal{X}}(x_t)$ and the normal displacement $n_t = x_t - p_t$.

Lemma 3. *Let $\mathcal{X} \subseteq \mathbb{R}^n$ be nonempty, closed, convex, and bounded, and let $D_{\mathcal{X}} := \sup_{u,v\in\mathcal{X}} \|u - v\| < \infty$. Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is μ-strongly convex and L-smooth on a convex neighborhood containing the iterates and $\mathcal{X}$. Let $x^\star := \arg\min_{x\in\mathcal{X}} f(x)$, $G_\star := \|\nabla f(x^\star)\|$. Consider the update in (8) with $\epsilon(1 + L\eta) \le 2, \eta\mu < 1$. Assume that, for every z, $\widehat{\pi}_t(z) \in \mathcal{X}, \|\widehat{\pi}_t(z) - \pi_{\mathcal{X}}(z)\| \le E_t$.*

Let $p_t := \pi_{\mathcal{X}}(x_t)$, $n_t := x_t - p_t$, $\mathcal{V}_t := f(p_t) - f(x^\star) + \frac{1}{2\eta}\|p_t - x^\star\|^2$. Define

$$\rho := 1 - \epsilon\eta\mu, \quad A_{\eta,\epsilon} := 1 + \epsilon\eta L, \quad M_\eta := G_\star + \left(L + \frac{1}{\eta}\right) D_{\mathcal{X}}, \quad C_\eta := \frac{L}{2} + \frac{1}{2\eta},$$
$$B_1 := M_\eta A_{\eta,\epsilon}, \quad B_2 := 2C_\eta A_{\eta,\epsilon}^2.$$

Then

$$\mathcal{V}_{t-1} \le \rho\mathcal{V}_t + B_1\|n_t\| + B_2\|n_t\|^2 + \epsilon M_\eta E_t + 2\epsilon^2 C_\eta E_t^2.$$

Moreover, the normal component satisfies $\|n_{t-1}\| \le (1 - \epsilon)\|n_t\|$.

The lemma shows that the Lyapunov function $\mathcal{V}_t$ contracts by the factor $\rho = 1 - \epsilon\eta\mu$, up to errors arising from the normal component and the denoising approximation. Since both $\|n_t\|$ and E_t decay geometrically along the reverse trajectory, iterating the recursion yields the finite-time bound below.

Theorem 2 (Convergence for strongly convex objectives)**.** *Let Assumption 3 and the assumptions of Lemma 3 hold. Consider the update in (8). Then, with ρ, B_1, B_2, M_η, and C_η defined as in Lemma 3, we have*

$$\mathcal{V}_0 \le \rho^T\mathcal{V}_T + (B_1\|n_T\| + \epsilon M_\eta C_{\mathcal{X}}\sigma_T)\,\frac{\rho^T - (1-\epsilon)^T}{\rho - (1-\epsilon)} + \left(B_2\|n_T\|^2 + 2\epsilon^2 C_\eta C_{\mathcal{X}}^2\sigma_T^2\right)\frac{\rho^T - (1-\epsilon)^{2T}}{\rho - (1-\epsilon)^2}.$$

Consequently,

$$\|p_0 - x^\star\|^2 \le 2\eta \left[\rho^T \mathcal{V}_T + (B_1 \|n_T\| + \epsilon M_\eta C_{\mathcal{X}} \sigma_T) \, \tfrac{\rho^T - (1-\epsilon)^T}{\rho - (1-\epsilon)} + \left(B_2 \|n_T\|^2 + 2\epsilon^2 C_\eta C_{\mathcal{X}}^2 \sigma_T^2 \right) \tfrac{\rho^T - (1-\epsilon)^{2T}}{\rho - (1-\epsilon)^2} \right].$$

Moreover, $\|n_0\| \le (1-\epsilon)^T \|n_T\|$.

We use T in this section because the convex-set guarantee holds for an arbitrary number of reverse steps. This differs from the preceding linear analysis, where t_0 denotes the largest index satisfying the required noise threshold. We use t_0 again in the subsequent local analysis for the same reason, namely that the guarantee applies only after the reverse process enters the required low-noise regime.

5.2 Convergence for Nonconvex Objectives

We next consider nonconvex objectives over the same compact convex feasible region. In this case, global optimality is not expected. The appropriate stationarity measure is the projected-gradient mapping

$$\mathcal{G}_\eta(p) := \tfrac{1}{\eta} \left(p - \pi_{\mathcal{X}}(p - \eta \nabla f(p)) \right), \tag{10}$$

which vanishes exactly at first-order stationary points of the constrained problem over $\mathcal{X}$.

Lemma 4 (One-step descent for nonconvex objectives over a convex region)**.** *Let $\mathcal{X} \subseteq \mathbb{R}^n$ be nonempty, compact, and convex. Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is L-smooth on a convex neighborhood containing the iterates and $\mathcal{X}$. Consider the backward update as in (8) with $\epsilon\eta L < 2$. Assume that, for every z, $\widehat{\pi}_t(z) \in \mathcal{X}$, $\|\widehat{\pi}_t(z) - \pi_{\mathcal{X}}(z)\| \le E_t$. Let $p_t := \pi_{\mathcal{X}}(x_t), n_t := x_t - p_t$. Define the following constants*

$$\alpha_{\eta,\epsilon} := \epsilon\eta \left(1 - \tfrac{\epsilon\eta L}{2}\right), \;\; A_{\eta,\epsilon} := 1 + \epsilon\eta L, \;\; G := \sup_{x \in \mathcal{X}} \|\nabla f(x)\|, \;\; \Gamma_1 := G A_{\eta,\epsilon}, \;\; \Gamma_2 := L A_{\eta,\epsilon}^2.$$

Then, with $\mathcal{G}_\eta(p_t)$ defined in (10), we have

$$f(p_{t-1}) \le f(p_t) - \alpha_{\eta,\epsilon} \|\mathcal{G}_\eta(p_t)\|^2 + \Gamma_1 \|n_t\| + \Gamma_2 \|n_t\|^2 \; + \epsilon G E_t + \epsilon^2 L E_t^2.$$

Moreover, the normal component satisfies $\|n_{t-1}\| \le (1-\epsilon)\|n_t\|$.

The next theorem sums the one-step descent inequality over the backward diffusion trajectory. The result gives a standard nonconvex convergence guarantee: at least one iterate has small projected-gradient mapping, with explicit error terms due to the denoising approximation.

Theorem 3 (Convergence to stationarity)**.** *Let Assumption 3 and the conditions of Lemma 4 hold. Consider the backward update as in (8). Then, with $\alpha_{\eta,\epsilon}$, Γ_1, Γ_2, G, and L defined as in Lemma 4, define*

$$\Delta_T := \max_{x \in \mathcal{X}} f(x) - \min_{x \in \mathcal{X}} f(x) + \frac{\Gamma_1 \|n_T\|}{\epsilon} + \frac{\Gamma_2 \|n_T\|^2}{\epsilon} + G C_{\mathcal{X}} \sigma_T + \epsilon L C_{\mathcal{X}}^2 \sigma_T^2,$$

where $C_{\mathcal{X}}$ is the constant from Lemma 2. Then

$$\alpha_{\eta,\epsilon} \textstyle\sum_{t=1}^T \|\mathcal{G}_\eta(p_t)\|^2 \le \Delta_T, \quad \Longrightarrow \quad \min_{1 \le t \le T} \|\mathcal{G}_\eta(p_t)\|^2 \le \tfrac{\Delta_T}{\alpha_{\eta,\epsilon} T}.$$

Moreover, the normal component satisfies $\|n_0\| \le (1-\epsilon)^T \|n_T\|$.

Comparison with prior convex theory. Together, Theorems 2 and 3 show that DCG recovers the standard first-order convergence behavior of projected-gradient methods. For strongly convex objectives, the projected optimization error converges geometrically, with the overall rate determined by the optimization contraction, the normal contraction, and the decay of the denoising error. For smooth nonconvex objectives, DCG achieves an $O(1/T)$ best-iterate bound on the squared projected-gradient mapping, equivalently an $O(1/\sqrt{T})$ rate in gradient norm. In both cases, the normal component contracts geometrically, so the analysis controls feasibility together with optimization progress.

The closest convex-support analysis is Leong & Traonmilin (2026), which likewise interprets diffusion denoisers as time-varying approximate projections, but analyzes a direct projected-gradient-style iteration for linear inverse recovery under a restricted-isometry condition, corresponding to a specialized quadratic objective. In contrast, our method retains the DDIM reverse-update structure and analyzes the general problem $\min_{x \in \mathcal{X}} f(x)$, establishing standard first-order rates for general strongly convex and smooth nonconvex objectives.

6 CONVERGENCE ANALYSIS: RIEMANNIAN MANIFOLD SETTING

We finally consider the case where the learned data geometry is a smooth embedded submanifold $\mathcal{M} \subseteq \mathbb{R}^n$. Compared with the convex setting, the projection geometry is local: the nearest-point projection is well-defined only in a tubular neighborhood of $\mathcal{M}$, and descent is measured by the Riemannian gradient along the manifold. All proofs are in Appendix E.

Problem setup. Let $\mathcal{M} \subseteq \mathbb{R}^n$ be a d-dimensional C^2 embedded submanifold, equipped with the Riemannian metric induced by the ambient Euclidean inner product. For each $p \in \mathcal{M}$, let $T_p\mathcal{M}$ denote the tangent space of $\mathcal{M}$ at p, and let $N_p\mathcal{M} := (T_p\mathcal{M})^\perp$ denote the corresponding normal space in $\mathbb{R}^n$. Let $U_p \in \mathbb{R}^{d\times n}$ and $V_p \in \mathbb{R}^{(n-d)\times n}$ be row-orthonormal matrices whose rows form orthonormal bases of $T_p\mathcal{M}$ and $N_p\mathcal{M}$, respectively. Thus $U_pU_p^\top = I_d$, $V_pV_p^\top = I_{n-d}$, and $U_pV_p^\top = 0$. By the local graph theorem for embedded submanifolds, there exist an open neighborhood $\mathcal{A}_p \subseteq \mathbb{R}^d$ of the origin and a C^2 map $\phi_p : \mathcal{A}_p \to \mathbb{R}^{n-d}$ such that every point of $\mathcal{M}$ sufficiently close to p can be uniquely written as $p + U_p^\top\alpha + V_p^\top\phi_p(\alpha)$ for some $\alpha \in \mathcal{A}_p$. We refer to ϕ_p as the local graph function, or Monge parametrization, of $\mathcal{M}$ over $T_p\mathcal{M}$. Since this parametrization is centered at p and tangent to $T_p\mathcal{M}$ at $\alpha = 0$, it satisfies $\phi_p(0) = 0$ and $D\phi_p(0) = 0$. We denote the corresponding local parametrization by $\Psi_p(\alpha) := p+U_p^\top\alpha+V_p^\top\phi_p(\alpha)$. Finally, let $r_{\mathrm{tub}} > 0$ be a uniform tubular radius of $\mathcal{M}$, meaning that every point within distance r_{tub} of $\mathcal{M}$ has a unique nearest point on $\mathcal{M}$. We define the corresponding tubular neighborhood by $\mathcal{T} := \{x \in \mathbb{R}^n : \mathrm{dist}(x, \mathcal{M}) < r_{\mathrm{tub}}\}$ and let $\pi_{\mathcal{M}}(x) := \mathrm{argmin}_{q\in\mathcal{M}} \|x - q\|$ denote the nearest-point projection onto $\mathcal{M}$. The projection $\pi_{\mathcal{M}}$ is well-defined and smooth on $\mathcal{T}$. We measure stationarity using the Riemannian gradient. For $p \in \mathcal{M}$, the Riemannian gradient is the tangent projection of the ambient Euclidean gradient: $\mathrm{grad}_{\mathcal{M}} f(p) = \mathrm{Proj}_{T_p\mathcal{M}}\nabla f(p) = U_p^\top U_p\nabla f(p)$.

Stein projection lemma. We first state the data and regularity assumptions needed to control the Stein denoiser near the manifold.

Assumption 4 (Data model for the Stein score)**.** *Let X be supported on $\mathcal{M}$ with density*

$$d\mu(x) \;=\; \vartheta(x)\, d\mathrm{vol}_{\mathcal{M}}(x),$$

where $\vartheta \in C^2(\mathcal{M})$ and there exist $0 < \vartheta_- \le \vartheta_+ < \infty$ such that $\vartheta_- \le \vartheta(x) \le \vartheta_+$ for all $x \in \mathcal{M}$.

Assumption 5 (Compact uniform local graph regularity)**.** *Assume that $\mathcal{M} \subseteq \mathbb{R}^n$ is compact and that there exist constants $\rho_0 > 0$ and $\kappa > 0$, such that, for every $p \in \mathcal{M}$, $\mathcal{M}$ admits a Monge chart*

$$\Psi_p(\alpha) = p + U_p^\top\alpha + V_p^\top\phi_p(\alpha), \qquad \alpha \in B_{\rho_0}(0) \subseteq \mathbb{R}^d,$$

with $\phi_p(0) = 0$, $D\phi_p(0) = 0$. Moreover, uniformly over $p \in \mathcal{M}$ and $\alpha \in B_{\rho_0}(0)$,

$$\|D^2\phi_p(\alpha)[u, v]\| \le \kappa\|u\|\|v\|.$$

Consequently, whenever $\|\alpha\| \le \rho_0$, $\|D\phi_p(\alpha)\| \le \kappa\|\alpha\|$. In particular, if $\|\alpha\| \le 1/(6\kappa)$, then $\|D\phi_p(\alpha)\| \le \frac{1}{6}$.

Under these assumptions, the Stein denoiser approximates the nearest-point projection onto $\mathcal{M}$ inside a tubular neighborhood.

Lemma 5. *Assume that Assumption 4, 5 hold. Let*

$$r := \min\left\{r_{\mathrm{tub}}, \rho_0, \tfrac{1}{6\kappa}\right\}, \qquad \mathcal{U} := \{x : \mathrm{dist}(x, \mathcal{M}) \le r\}. \tag{11}$$

Then there exist constants $\bar\sigma > 0$ and $C > 0$, depending only on the constants in the assumptions above, such that, for all $0 < \sigma \le \bar\sigma$ and all $x \in \mathcal{U}$, we have $\left\|\widehat{\pi}_\sigma(x) - \pi_{\mathcal{M}}(x)\right\| \le C\sigma|\log\sigma|$.

Convergence. We now analyze the guided diffusion update on the manifold. Since the nearest-point projection is only locally well-defined and the Stein projection bound holds only in $\mathcal{U}$, we impose a local trajectory condition ensuring that the iterates and the gradient-updated points remain in the region where the projection approximation is valid.

Assumption 6. *The objective function f is L-smooth and satisfies $\|\nabla f(x)\| \le G$ on $\mathcal{U}$.*

Similar to the analysis in the previous sections, the descent analysis uses a Lyapunov function that combines the objective evaluated at the projected point $p_t := \pi_{\mathcal{M}}(x)$ with a quadratic penalty on the normal displacement $n_t := x - \pi_{\mathcal{M}}(x)$.

Lemma 6 (Lyapunov descent with inexact projection)**.** *Suppose Assumption 5 and 6 hold. Consider the update (8). Let $z_t := x_t - \eta\nabla f(x_t)$ and suppose that $\|\widehat{\pi}_t(z_t) - \pi_{\mathcal{M}}(z_t)\| \le E_t$. Assume that $\mathrm{dist}(x_t, \mathcal{M}) \le \frac{r}{2}$, where r is defined in (11). Let $p_t := \pi_{\mathcal{M}}(x_t), n_t := x_t - p_t$. If $\eta \le \min\{\frac{1}{50L}, \frac{r}{4G}\}, \epsilon \le \min\{\frac{1}{512\eta(G\kappa+L)}, \frac{r}{4E_t}\}$, then we have that $\mathrm{dist}(x_{t-1}, \mathcal{M}) \le \frac{r}{2}$, and with*

$$V(x) := f(\pi_{\mathcal{M}}(x)) + \tfrac{1}{160\eta}\|x - \pi_{\mathcal{M}}(x)\|^2,$$

we have
$$V(x_{t-1}) \le V(x_t) - \tfrac{\epsilon\eta}{480}\|\mathrm{grad}_{\mathcal{M}} f(p_t)\|^2 - \tfrac{\epsilon}{2000\eta}\|n_t\|^2 + \tfrac{6\epsilon}{\eta}E_t^2.$$

The descent inequality contains two negative terms. The term involving $\|\mathrm{grad}_{\mathcal{M}} f(p_t)\|^2$ controls the Riemannian stationarity of the projected iterate $p_t = \pi_{\mathcal{M}}(x_t)$, while the term involving $\|n_t\|^2$ controls the normal distance to the manifold. Thus, the update simultaneously drives the projected point toward Riemannian stationarity and keeps the ambient iterate close to $\mathcal{M}$, up to the denoising error. Moreover, the argument depends on the denoiser only through its pointwise projection error, so it extends directly to an imperfectly learned denoiser. In particular, if the implemented denoiser $\widetilde{\pi}_t$ satisfies $\|\widetilde{\pi}_t(z_t) - \widehat{\pi}_t(z_t)\| \le \delta_t$, then the same result holds with E_t replaced by $E_t + \delta_t$. Thus, finite-noise and denoiser-approximation errors enter the Lyapunov bound through the single additive term $\frac{6\epsilon}{\eta}(E_t + \delta_t)^2$.

Combining Lemma 6 with the Stein projection approximation in Lemma 5 gives the following finite-time convergence result.

Theorem 4. *Suppose Assumptions 4 and 5 hold, and define t_0 as the largest index such that $\sigma_{t_0} \le \bar{\sigma}$. Assume that $\mathrm{dist}(x_{t_0}, \mathcal{M}) \le \frac{r}{2}$. Consider the backward denoising scheme with gradient guidance as in (8) with $\eta \le \min\left\{\frac{1}{50L}, \frac{r}{4G}\right\}$, $\epsilon \le \min\left\{\frac{1}{2}, \frac{1}{512\eta(G\kappa+L)}, \frac{r}{4C\sigma_{t_0}|\log\sigma_{t_0}|}\right\}$. Let $p_t := \pi_{\mathcal{M}}(x_t)$ and $n_t := x_t - p_t$. Define*

$$\Delta_{t_0} := V(x_{t_0}) - \min_{x\in\mathcal{M}} f(x) + \tfrac{6C^2\sigma_{t_0}^2}{\eta}\left(|\log\sigma_{t_0}|^2 + 4|\log\sigma_{t_0}| + 8\right),$$

where V is defined in Lemma 6 and C is defined in Lemma 5. Then

$$\textstyle\sum_{t=1}^{t_0}\left(\tfrac{\epsilon\eta}{480}\|\mathrm{grad}_{\mathcal{M}} f(p_t)\|^2 + \tfrac{\epsilon}{2000\eta}\|n_t\|^2\right) \le \Delta_{t_0}.$$

Consequently,
$$\min_{1\le t\le t_0}\left(\tfrac{\epsilon\eta}{480}\|\mathrm{grad}_{\mathcal{M}} f(p_t)\|^2 + \tfrac{\epsilon}{2000\eta}\|n_t\|^2\right) \le \tfrac{\Delta_{t_0}}{t_0}.$$

Comparison with prior Riemannian theory. The closest theoretical comparison is Kharitenko et al. (2025). Both analyses are local and establish an $O(1/\sqrt{T})$ best-iterate rate for the Riemannian gradient norm, up to a finite-noise bias. However, Kharitenko et al. (2025) controls both the denoiser and its Jacobian to approximate retraction and tangent projection, requiring derivative-level bounds and C^3 regularity. Our analysis requires only value-level approximation of the nearest-point projection, not Jacobian accuracy, which is a weaker requirement because accurate denoiser values do not generally imply accurate Jacobians. Moreover, Kharitenko et al. (2025) constructs a standalone optimizer that applies a full approximate retraction at each iteration, whereas DCG follows a single reverse-diffusion trajectory and interpolates between the current iterate and the denoised projected-gradient candidate. This relaxed update yields weaker per-step contraction and consequently more conservative constants.

7 Conclusion

We introduced Denoising-Corrected Gradient Guidance, an inference-time method that applies a gradient step followed by ordinary pretrained diffusion denoising. By interpreting the Stein posterior-mean denoiser as an approximate projection onto the data support, we connected DCG to inexact projected-gradient optimization and established projection bounds and finite-time guarantees for linear subspaces, compact convex sets, and compact smooth submanifolds. Experiments on synthetic constrained problems, trajectory tracking, and D4RL benchmarks show that DCG improves the balance between objective value and feasibility over direct gradient guidance while remaining competitive with more geometry-intensive baselines.

8 AI USE STATEMENT

In this work, we used generative AI tools to aid and polish writing, assist with retrieval and discovery of related work, support research execution through coding assistance, draft portions of the manuscript, and assist with completing details of mathematical proofs. The research ideas, problem formulation, methodological design, and proof strategies were developed by the authors. For mathematical results, the authors provided the proof structure and key arguments, while generative AI was used to help fill in intermediate derivations and presentation details.

All AI-assisted material was reviewed and verified by the authors. In particular, references and related work identified with AI assistance were checked against the original sources; AI-assisted code was inspected and tested by the authors; mathematical derivations and proof details produced with AI assistance were independently checked for correctness; and AI-assisted text was reviewed and revised for accuracy, appropriate attribution, and consistency with the authors' intended technical content.

We take responsibility for the final content of this work, including text, claims, code, or other artifacts produced with the aid of generative AI.

9 REPRODUCIBILITY STATEMENT

We provide an anonymous implementation of our method, together with code for reproducing the experiments, at `https://anonymous.4open.science/r/denoising-as-projection-3352`. The experimental settings and implementation details are described in the appendix. For the theoretical results, all assumptions are stated explicitly and complete proofs are provided in the appendix.

REFERENCES

Iskander Azangulov, Peter Potaptchik, Qinyu Li, Eddie Aamari, George Deligiannidis, and Judith Rousseau. Adaptive diffusion guidance via stochastic optimal control. In *The 29th International Conference on Artificial Intelligence and Statistics*, 2026. URL `https://openreview.net/forum?id=mef54yiI1L`.

Joao Carvalho, An Thai Le, Piotr Kicki, Dorothea Koert, and Jan Peters. Motion planning diffusion: Learning and adapting robot motion planning with diffusion models. *IEEE Transactions on Robotics*, 41:4881–4901, 2025.

Cheng Chi, Zhenjia Xu, Siyuan Feng, Eric Cousineau, Yilun Du, Benjamin Burchfiel, Russ Tedrake, and Shuran Song. Diffusion policy: Visuomotor policy learning via action diffusion. *The International Journal of Robotics Research*, 44(10-11):1684–1704, 2025.

Jacob K. Christopher, Stephen Baek, and Ferdinando Fioretto. Constrained synthesis with projected diffusion models. In *Advances in Neural Information Processing Systems*, 2024.

Hyungjin Chung, Byeongsu Sim, Dohoon Ryu, and Jong Chul Ye. Improving diffusion models for inverse problems using manifold constraints. *Advances in Neural Information Processing Systems*, 35:25683–25696, 2022.

Hyungjin Chung, Jeongsol Kim, Michael T. McCann, Marc L. Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *International Conference on Learning Representations*, 2023.

Hyungjin Chung, Jeongsol Kim, Geon Yeong Park, Hyelin Nam, and Jong Chul Ye. Cfg++: Manifold-constrained classifier free guidance for diffusion models. In *International Conference on Learning Representations*, volume 2025, pp. 30824–30850, 2025.

Kevin Clark, Paul Vicol, Kevin Swersky, and David J. Fleet. Directly fine-tuning diffusion models on differentiable rewards. In *International Conference on Learning Representations*, 2024.

Mauricio Delbracio and Peyman Milanfar. Inversion by direct iteration: An alternative to denoising diffusion for image restoration. *Transactions on Machine Learning Research*, 2023. URL https://arxiv.org/abs/2303.11435.

Fei Deng, Qifei Wang, Wei Wei, Tingbo Hou, and Matthias Grundmann. PRDP: Proximal reward difference prediction for large-scale reward finetuning of diffusion models. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7423–7433. IEEE, 2024.

Prafulla Dhariwal and Alexander Quinn Nichol. Diffusion models beat gans on image synthesis. In *Advances in Neural Information Processing Systems*, 2021.

Paolo Giaretta, Zeyang Li, and Navid Azizan. DiRecT: Safe diffusion-based planning via receding-horizon denoising. *arXiv preprint arXiv:2606.15359*, 2026. URL https://arxiv.org/abs/2606.15359.

Yingqing Guo, Hui Yuan, Yukang Yang, Minshuo Chen, and Mengdi Wang. Gradient guidance for diffusion models: An optimization perspective. *Advances in Neural Information Processing Systems*, 37:90736–90770, 2024.

Samar Hadou, Yigit Berkay Uslu, and Alejandro Ribeiro. Constrained diffusion models with primal-dual inference. *arXiv preprint arXiv:2606.17192*, 2026.

Yutong He, Naoki Murata, Chieh-Hsin Lai, Yuhta Takida, Toshimitsu Uesaka, Dongjun Kim, WeiHsiang Liao, Yuki Mitsufuji, Zico Kolter, Ruslan Salakhutdinov, et al. Manifold preserving guided diffusion. In *International Conference on Learning Representations*, volume 2024, pp. 44819–44850, 2024.

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

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, 2020.

Michael Janner, Yilun Du, Joshua Tenenbaum, and Sergey Levine. Planning with diffusion for flexible behavior synthesis. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato (eds.), *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 9902–9915. PMLR, 17–23 Jul 2022. URL https://proceedings.mlr.press/v162/janner22a.html.

Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. In *Advances in Neural Information Processing Systems*, 2022.

Shervin Khalafi, Dongsheng Ding, and Alejandro Ribeiro. Constrained diffusion models via dual training. *Advances in Neural Information Processing Systems*, 37:26543–26576, 2024.

Andrey Kharitenko, Zebang Shen, Riccardo de Santi, Niao He, and Florian Doerfler. Landing with the score: Riemannian optimization through denoising. *arXiv preprint arXiv:2509.23357*, 2025.

Sunwoo Kim, Minkyu Kim, and Dongmin Park. Test-time alignment of diffusion models without reward over-optimization. In *International Conference on Learning Representations*, volume 2025, pp. 87238–87272, 2025.

Lingkai Kong, Yuanqi Du, Wenhao Mu, Kirill Neklyudov, Valentin De Bortoli, Dongxia Wu, Haorui Wang, Aaron M Ferber, Yian Ma, Carla P Gomes, and Chao Zhang. Diffusion models as constrained samplers for optimization with unknown constraints. In Yingzhen Li, Stephan Mandt, Shipra Agrawal, and Emtiyaz Khan (eds.), *Proceedings of The 28th International Conference on Artificial Intelligence and Statistics*, volume 258 of *Proceedings of Machine Learning Research*, pp. 4582–4590. PMLR, 03–05 May 2025. URL https://proceedings.mlr.press/v258/kong25b.html.

Oscar Leong and Yann Traonmilin. A recovery theory for diffusion priors: Deterministic analysis of the implicit prior algorithm. In *The 29th International Conference on Artificial Intelligence and Statistics*, 2026. URL https://openreview.net/forum?id=FwSlWVkW7i.

Anjian Li and Ryne Beeson. Aligning diffusion model with problem constraints for trajectory optimization. *arXiv preprint arXiv:2504.00342*, 2025.

Zeyang Li, Kaveh Alim, and Navid Azizan. HardFlow: Hard-constrained sampling for flow-matching models via trajectory optimization. *arXiv preprint arXiv:2511.08425*, 2025. URL https://arxiv.org/abs/2511.08425.

Jinhao Liang, Jacob K Christopher, Sven Koenig, and Ferdinando Fioretto. Simultaneous multi-robot motion planning with projected diffusion models. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=Sp7jclUwkV.

Jinhao Liang, Sven Koenig, and Ferdinando Fioretto. Discrete-guided diffusion for scalable and safe multi-robot motion planning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 23417–23424, 2026.

Guan-Horng Liu, Tianrong Chen, Evangelos Theodorou, and Molei Tao. Mirror diffusion models for constrained and watermarked generation. *Advances in Neural Information Processing Systems*, 36:42898–42917, 2023.

Peyman Milanfar and Mauricio Delbracio. Denoising: A powerful building block for imaging, inverse problems and machine learning. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 383(2299):20240326, 2025. doi: 10.1098/rsta.2024.0326. URL https://doi.org/10.1098/rsta.2024.0326.

Koichi Miyasawa et al. An empirical bayes estimator of the mean of a normal population. *Bull. Inst. Internat. Statist*, 38(181-188):1–2, 1961.

Kazuki Mizuta and Karen Leung. Cobl-diffusion: Diffusion-based conditional robot planning in dynamic environments using control barrier and lyapunov functions. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 13801–13808. IEEE, 2024.

Jean Pachebat, Giovanni Conforti, Alain Durmus, and Yazid Janati. Iterative tilting for diffusion fine-tuning. *arXiv preprint arXiv:2512.03234*, 2025.

Frank Permenter and Chenyang Yuan. Interpreting and improving diffusion models from an optimization perspective. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 40461–40483. PMLR, 21–27 Jul 2024. URL https://proceedings.mlr.press/v235/permenter24a.html.

Peter Potaptchik, Lee Cheuk Kit, and Michael Samuel Albergo. Tilt matching for scalable sampling and fine-tuning. In *Forty-third International Conference on Machine Learning*, 2026a. URL https://openreview.net/forum?id=dQA4Gjt4KU.

Peter Potaptchik, Adhi Saravanan, Abbas Mammadov, Alvaro Prat, Michael Samuel Albergo, and Yee Whye Teh. Meta flow maps enable scalable reward alignment. In *Forty-third International Conference on Machine Learning*, 2026b.

Herbert E Robbins. An empirical bayes approach to statistics. In *Breakthroughs in Statistics: Foundations and basic theory*, pp. 388–394. Springer, 1992.

Yaniv Romano, Michael Elad, and Peyman Milanfar. The little engine that could: Regularization by denoising (RED). *SIAM Journal on Imaging Sciences*, 10(4):1804–1844, 2017. doi: 10.1137/16M1102884. URL https://doi.org/10.1137/16M1102884.

Yorai Shaoul, Itamar Mishani, Shivam Vats, Jiaoyang Li, and Maxim Likhachev. Multi-robot motion planning with diffusion models. In *International Conference on Learning Representations*, volume 2025, pp. 95791–95811, 2025.

Jascha Sohl-Dickstein, Eric A. Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*, 2015.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2021a.

Jiaming Song, Arash Vahdat, Morteza Mardani, and Jan Kautz. Pseudoinverse-guided diffusion models for inverse problems. In *International Conference on Learning Representations*, 2023.

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

Yang Song, Liyue Shen, Lei Xing, and Stefano Ermon. Solving inverse problems in medical imaging with score-based generative models. In *International Conference on Learning Representations*, 2022.

Suhas Sreehari, S Venkat Venkatakrishnan, Brendt Wohlberg, Gregery T Buzzard, Lawrence F Drummy, Jeffrey P Simmons, and Charles A Bouman. Plug-and-play priors for bright field electron tomography and sparse interpolation. *IEEE Transactions on Computational Imaging*, 2(4): 408–423, 2016.

Singanallur V Venkatakrishnan, Charles A Bouman, and Brendt Wohlberg. Plug-and-play priors for model based reconstruction. In *2013 IEEE global conference on signal and information processing*, pp. 945–948. IEEE, 2013.

Hengkang Wang, Xu Zhang, Taihui Li, Yuxiang Wan, Tiancong Chen, and Ju Sun. Dmplug: A plug-in method for solving inverse problems with diffusion models. *Advances in Neural Information Processing Systems*, 37:117881–117916, 2024.

Lemeng Wu, Chengyue Gong, Xingchao Liu, Mao Ye, and Qiang Liu. Diffusion-based molecule generation with informative prior bridges. *Advances in neural information processing systems*, 35:36533–36545, 2022.

Wei Xiao, Tsun-Hsuan Wang, Chuang Gan, Ramin Hasani, Mathias Lechner, and Daniela Rus. Safediffuser: Safe planning with diffusion probabilistic models. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=ig2wk7kK9J.

Rayhan Zirvi, Bahareh Tolooshams, and Anima Anandkumar. Diffusion state-guided projected gradient for inverse problems. In *International Conference on Learning Representations*, 2025.