Greed is Good: A Unifying Perspective on Guided Generation

Zander W. Blasingame **Chen Liu**

Clarkson University

- Consider $q(\boldsymbol{X}_1)$ which models real-world data

- Consider $q(\boldsymbol{X}_1)$ which models real-world data Call it the target distribution

- Consider $q(\boldsymbol{X}_1)$ which models real-world data
- Consider some tractable $p(\boldsymbol{X}_0)$

- Consider $q(\boldsymbol{X}_1)$ which models real-world data
- Consider some tractable $p(\boldsymbol{X}_0)$ Call it the source distribution

- Consider $q(\boldsymbol{X}_1)$ which models real-world data
- Consider some tractable $p(\boldsymbol{X}_0)$
- Consider the *flow* $\Phi \in C^{1,r}(\mathbb{R} \times \mathbb{R}^d; \mathbb{R}^d)$ which satisfies

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi_{t,1}(\boldsymbol{x}) = \boldsymbol{u}_t(\Phi_{t,1}(\boldsymbol{x})) \tag{1}$$

- Consider $q(\boldsymbol{X}_1)$ which models real-world data
- Consider some tractable $p(\boldsymbol{X}_0)$
- Consider the *flow* $\Phi \in C^{1,r}(\mathbb{R} \times \mathbb{R}^d; \mathbb{R}^d)$ which satisfies

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi_{t,1}(\boldsymbol{x}) = \boldsymbol{u}_t(\Phi_{t,1}(\boldsymbol{x})) \tag{1}$$

- We want to find a flow (or vector field) such that

$$\Phi_{t,1}(\boldsymbol{X}_0) \sim q(\boldsymbol{X}_1) \tag{2}$$

- Consider $q(\boldsymbol{X}_1)$ which models real-world data
- Consider some tractable $p(\boldsymbol{X}_0)$
- Consider the *flow* $\Phi \in C^{1,r}(\mathbb{R} \times \mathbb{R}^d; \mathbb{R}^d)$ which satisfies

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi_{t,1}(\boldsymbol{x}) = \boldsymbol{u}_t(\Phi_{t,1}(\boldsymbol{x})) \tag{1}$$

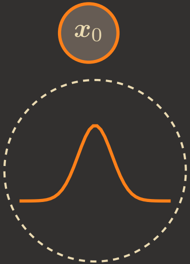
- We want to find a flow (or vector field) such that

$$\Phi_{t,1}(\boldsymbol{X}_0) \sim q(\boldsymbol{X}_1) \tag{2}$$

- Then we aim to find θ such that

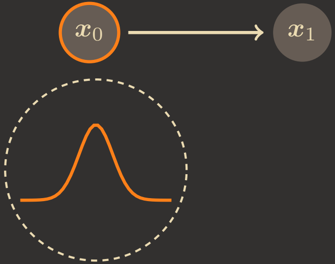
$$\boldsymbol{u}_t^{\theta}(\boldsymbol{x}) \approx \boldsymbol{u}_t(\boldsymbol{x}) \tag{3}$$

- Consider an Euler scheme with steps $\{t_n\}_{n=0}^{N}$ and step size h



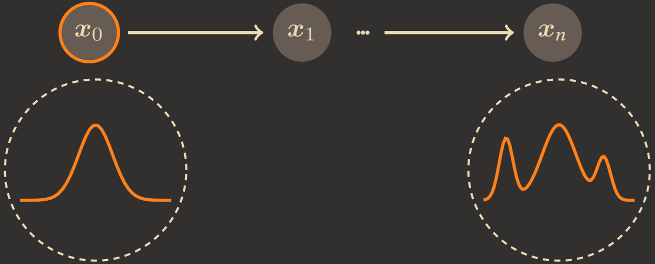
- Consider an Euler scheme with steps $\{t_n\}_{n=0}^{N}$ and step size h
- Now take Euler steps…

$$x_1 = x_0 + h u_0^{\theta}(x_0)$$



- Consider an Euler scheme with steps $\{t_n\}_{n=0}^{N}$ and step size h
- Now take Euler steps…

$$x_n = x_{n-1} + h u_{n-1}^{\theta}(x_{n-1})$$



- Consider an Euler scheme with steps $\{t_n\}_{n=0}^N$ and step size h
- Now take Euler steps…

$$x_{n+1} = x_n + h u_n^\theta(x_n)$$

- Consider an Euler scheme with steps $\{t_n\}_{n=0}^{N}$ and step size h
- Now take Euler steps…

$$x_N = x_{N-1} + h u_{N-1}^{\theta}(x_{N-1})$$

Definition 1 (Problem statement). Given some $t_1 \in [0, 1)$ and step size regime $\{t_1 < t_2 < \ldots < t_N = 1\}$ solve:

$$\begin{aligned} &\text{Find a sequence} && \{x_n\}_{n=1}^{N} && \text{which minimizes } \mathcal{L}(x_N), \\ &\text{subject to} && x_{n+1} = \Phi(t_{n+1}, t_n, x_n). \end{aligned} \tag{4}$$

Gradient descent?

Need $\nabla_{\{\boldsymbol{x},\theta\}} L\left(\varphi_1^{\theta}(\boldsymbol{x})\right)$

Discretize-then-optimize (DTO)

- Simplest approach, just backprop through the solver
- Pros: Accuracy of gradients, fast, and easy to implement.
- Cons: Memory intensive $\mathcal{O}(n)$, optimization w.r.t. discretization and not continuous ideal

Discretize-then-optimize (DTO)

- Simplest approach, just backprop through the solver
- Pros: Accuracy of gradients, fast, and easy to implement.
- Cons: Memory intensive $\mathcal{O}(n)$, optimization w.r.t. discretization and not continuous ideal

Optimize-then-discretize (OTD)

- *I.e.*, the adjoint method, numerically solve another ODE
- Pros: Memory efficiency $\mathcal{O}(1)$, flexibility
- Cons: Computational cost, truncation errors

- Recall the target prediction formula

$$\boldsymbol{x}_{1|t}(\boldsymbol{x}) = \mathbb{E}[\boldsymbol{X}_1 | \boldsymbol{X}_t = \boldsymbol{x}] \tag{4}$$

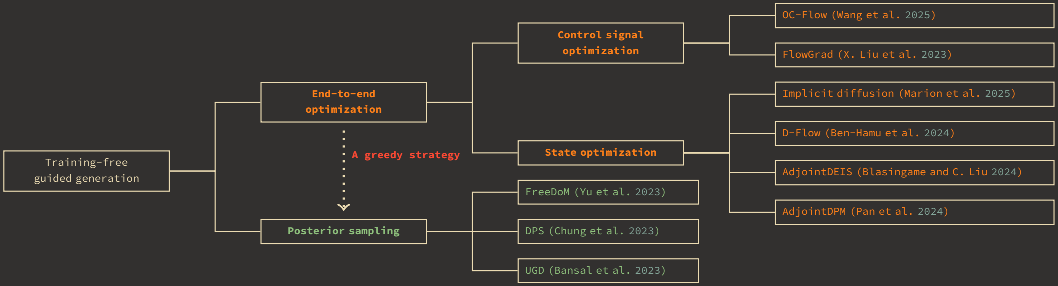
- Recall the target prediction formula

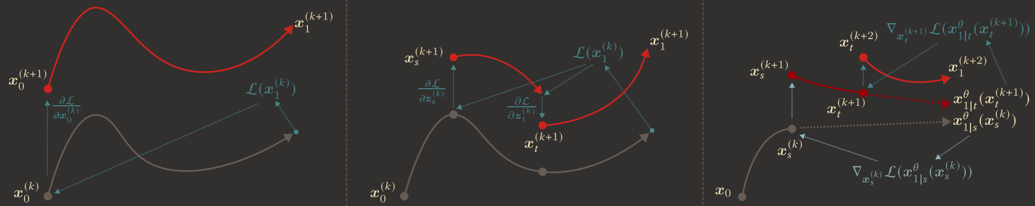
$$x_{1|t}(x) = \mathbb{E}[X_1 | X_t = x] \tag{4}$$

- Use this estimate of the posterior to perform guidance,

$$\nabla_x L(x_{1|t}(x)) \tag{5}$$

We view posterior guidance as a *greedy strategy* of end-to-end guidance


Training-free guided generation
End-to-end optimization
A greedy strategy
Posterior sampling
Control signal optimization
State optimization
OC-Flow (Wang et al. 2025)
FlowGrad (X. Liu et al. 2023)
Implicit diffusion (Marion et al. 2025)
D-Flow (Ben-Hamu et al. 2024)
AdjointDEIS (Blasingame and C. Liu 2024)
AdjointDPM (Pan et al. 2024)
FreeDoM (Yu et al. 2023)
DPS (Chung et al. 2023)
UGD (Bansal et al. 2023)



State optimization

Control signal optimization

End-to-end guidance

Posterior guidance

Proposition 1 (Exact solution of affine probability paths). Given an initial value of $\boldsymbol{x}_s$ at time $s \in [0,1]$ the solution $\boldsymbol{x}_t$ at time $t \in [0,1]$ of an affine probability path is:

$$\boldsymbol{x}_t = \frac{\sigma_t}{\sigma_s}\boldsymbol{x}_s + \sigma_t \int_{\gamma_s}^{\gamma_t} \boldsymbol{x}^{\theta}_{1|\gamma}(\boldsymbol{x}_\gamma)\,\mathrm{d}\gamma, \tag{6}$$

where $\gamma_t = \alpha_t/\sigma_t$.

- Consider the Taylor expansion of Eq. (6)

$$\boldsymbol{x}_t = \frac{\sigma_t}{\sigma_s}\boldsymbol{x}_s + \sigma_t \sum_{n=0}^{k-1} \frac{\mathrm{d}^n}{\mathrm{d}\gamma^n}\left[\boldsymbol{x}^{\theta}_{1|\gamma}(\boldsymbol{x}_\gamma)\right]_{\gamma=\gamma_s} \frac{h^{n+1}}{n!} + \mathcal{O}(h^{k+1}) \tag{7}$$

- Consider the Taylor expansion of Eq. (6)
- Then, the first-order expansion is

$$x_t = \frac{\sigma_t}{\sigma_s} x_s + \sigma_t h x^{\theta}_{1|s}(x_s) + \mathcal{O}(h) \tag{7}$$

- Consider the Taylor expansion of Eq. (6)
- Then, the first-order expansion is
- Drop high-order error terms

$$x_t \approx \frac{\sigma_t}{\sigma_s} x_s + \sigma_t h x^{\theta}_{1|s}(x_s) \tag{7}$$

- Consider the Taylor expansion of Eq. (6)
- Then, the first-order expansion is
- Drop high-order error terms

$$\boldsymbol{x}_t \approx \frac{\sigma_t}{\sigma_s}\boldsymbol{x}_s + \sigma_t \left(\frac{\alpha_t}{\sigma_t} - \frac{\alpha_s}{\sigma_s}\right)\boldsymbol{x}^{\theta}_{1|s}(\boldsymbol{x}_s) \tag{7}$$

- Consider the Taylor expansion of Eq. (6)
- Then, the first-order expansion is
- Drop high-order error terms

$$\boldsymbol{x}_t \approx \frac{\sigma_t}{\sigma_s}\boldsymbol{x}_s + \left(\alpha_t - \alpha_s \frac{\sigma_t}{\sigma_s}\right)\boldsymbol{x}_{1|s}^{\theta}(\boldsymbol{x}_s) \tag{7}$$

- Consider the Taylor expansion of Eq. (6)
- Then, the first-order expansion is
- Drop high-order error terms
- In the limit as $t \longrightarrow 1$, $\alpha_t \longrightarrow 1$, $\sigma_t \longrightarrow 0$

$$\boldsymbol{x}_1 \approx \boldsymbol{x}^{\theta}_{1|s}(\boldsymbol{x}_s) \tag{7}$$

- Consider the Taylor expansion of Eq. (6)
- Then, the first-order expansion is
- Drop high-order error terms
- In the limit as $t \to 1$, $\alpha_t \to 1$, $\sigma_t \to 0$

$$\boldsymbol{x}_1 \approx \boldsymbol{x}^{\theta}_{1|s}(\boldsymbol{x}_s) \tag{7}$$

- Hence, the greedy gradient $\nabla_{\boldsymbol{x}} L(\boldsymbol{x}^{\theta}_{1|t}(\boldsymbol{x}_t))$, can be viewed as a DTO scheme with a large explicit Euler step.

- Now consider an OTD scheme

- Now consider an OTD scheme
- Continuous adjoint equations have a form of

$$\boldsymbol{a}_{\boldsymbol{x}}(s) = \frac{\sigma_t}{\sigma_s}\boldsymbol{a}_{\boldsymbol{x}}(t) + \sigma_t \int_{\gamma_s}^{\gamma_t} \boldsymbol{a}_{\boldsymbol{x}}(\gamma)^\top \frac{\partial \boldsymbol{x}_{1|\gamma}^{\theta}(\boldsymbol{x}_\gamma)}{\partial \boldsymbol{x}_\gamma}\,\mathrm{d}\gamma \tag{8}$$

- Now consider an OTD scheme
- Continuous adjoint equations have a form of

$$\boldsymbol{a}_{\boldsymbol{x}}(s) = \frac{\sigma_t}{\sigma_s}\boldsymbol{a}_{\boldsymbol{x}}(t) + \sigma_t \int_{\gamma_s}^{\gamma_t} \boldsymbol{a}_{\boldsymbol{x}}(\gamma)^\top \frac{\partial \boldsymbol{x}_{1|\gamma}^\theta(\boldsymbol{x}_\gamma)}{\partial \boldsymbol{x}_\gamma}\,\mathrm{d}\gamma \tag{8}$$

- Then in the limit as $t \to 1$ the first iteration of a fixed-point iteration scheme yields

$$\boldsymbol{a}_{\boldsymbol{x}}(s) \approx \boldsymbol{a}_{\boldsymbol{x}}(1)^\top \frac{\partial \boldsymbol{x}_{1|t}^\theta(\boldsymbol{x}_s)}{\partial \boldsymbol{x}_s} = \nabla_{\boldsymbol{x}} L(\boldsymbol{x}_{1|s}^\theta(\boldsymbol{x}_s)) \tag{9}$$

Proposition 2 (Dynamics of greedy gradient guidance). Consider the standard affine Gaussian probability paths model trained to zero loss. The Gateaux differential of $\boldsymbol{x}$ at some time $t \in [0,1]$ in the direction of the gradient $\nabla_{\boldsymbol{x}}\mathcal{L}\left(\boldsymbol{x}_{1|t}^{\theta}(\boldsymbol{x})\right)$ is given by

$$\delta_{\boldsymbol{x}}^{\mathcal{G}}\Phi_{t,1}^{\theta}(\boldsymbol{x}) = -\nabla_{\boldsymbol{x}}\Phi_{t,1}^{\theta}(\boldsymbol{x})\nabla_{\boldsymbol{x}}\boldsymbol{x}_{1|t}^{\theta}(\boldsymbol{x})^{\top}\nabla_{\boldsymbol{x}_1}\mathcal{L}(\boldsymbol{x}_1). \tag{10}$$

Theorem 3 (Dynamics of gradient vs greedy guidance). The difference between the dynamics of gradient guidance and greedy gradient guidance in Proposition 2 for a point $\boldsymbol{x}$ at time t with guidance function $\mathcal{L} \in C^1(\mathbb{R}^d)$ is bounded by $\mathcal{O}(h^2)$ where $h \coloneqq \gamma_1 - \gamma_t$, i.e.,

$$\left\|\nabla_{\boldsymbol{x}}\Phi_{t,1}^{\theta}(\boldsymbol{x}) - \nabla_{\boldsymbol{x}}\boldsymbol{x}_{1|t}^{\theta}(\boldsymbol{x})\right\| = \mathcal{O}(h^2). \tag{11}$$

Theorem 4 (Greedy convergence). For affine probability paths, if there exists a sequence of states $x_t^{(n)}$ at time t such that it converges to the locally optimal solution $x_{1|t}^{\theta}(x_t^{(n)}) \rightarrow x_1^*$. Then the solution, $\Phi_{1|t}^{\theta}(x_t^{(n)})$, converges to a neighborhood of size $\mathcal{O}(h^2)$ centered at x_1^*.

If *greedy* is Euler...

If *greedy* is Euler...
What if we went beyond Euler?

Theorem 5 (Truncation error of single-step gradients). Let Φ be an explict Runge-Kutta solver of order $\alpha > 0$ of a flow model with flow $\Phi_{s,t}^{\theta}(\boldsymbol{x})$. Then for any $t \in [0,1]$,

$$\left\|\nabla_{\boldsymbol{x}} \Phi_{t,1}^{\theta}(\boldsymbol{x}) - \nabla_{\boldsymbol{x}} \Phi_{t,1}(\boldsymbol{x})\right\| = \mathcal{O}(h^{\alpha+1}), \tag{12}$$

where $h = 1 - t$.

Corollary 5.1 (Convergence of a α-th order posterior gradient). For affine probability paths, if there exists a sequence of states $x_t^{(n)}$ at time t such that it converges to the locally optimal solution $\Phi_{t,1}^{\theta}(x_t^{(n)}) \to x_1^*$. Then solution, $\Phi_{1|t}^{\theta}(x_t^{(n)})$, converges to a neighborhood of size $\mathcal{O}(h^{\alpha+1})$ centered at x_1^*.

Corollary 5.2 (Dynamics of α-th order posterior gradient). Consider the standard affine Gaussian probability paths model trained to zero loss. Let Φ be an explicit Runga-Kutta solver of order $\alpha > 0$ of a flow model with flow $\Phi_{s,t}^{\theta}(x)$. The Gateaux differential of x at some time $t \in [0,1]$ in the direction of the gradient $\nabla_x \mathcal{L}\left(\Phi_{t,1}(x)\right)$ is given by

$$\delta_x^{\Phi}(x) = -\nabla_x \Phi_{t,1}^{\theta}(x) \nabla_x \Phi_{t,1}(x)^{\top} \nabla_{x_1} \mathcal{L}(x_1). \tag{13}$$

Figure 1: Qualitative visualization of using greedy guidance to solve the HDR inverse problem. Top row is the ground truth, middle row is the measurement, and the bottom row is the reconstruction.

Table 1: Further ablations on the number of discretization steps on the non-linear HDR inverse problem.

Method	**PSNR** ($\uparrow$)	**SSIM** ($\uparrow$)	**LPIPS** ($\downarrow$)	**FID** ($\downarrow$)
DAPS	$27.12_{\pm 3.53}$	$0.752_{\pm 0.041}$	$0.162_{\pm 0.072}$	42.97
DPS	$22.73_{\pm 6.07}$	$0.591_{\pm 0.141}$	$0.264_{\pm 0.156}$	112.82
RED-diff	$22.16_{\pm 3.41}$	$0.512_{\pm 0.083}$	$0.258_{\pm 0.089}$	108.32
Greedy (Euler)	$25.07_{\pm 4.25}$	$0.776_{\pm 0.126}$	$0.173_{\pm 0.070}$	43.25
Greedy (2-step Euler)	$26.32_{\pm 4.34}$	$0.802_{\pm 0.111}$	$0.173_{\pm 0.065}$	38.64
Greedy (3-step Euler)	$27.17_{\pm 4.21}$	$0.820_{\pm 0.096}$	$0.154_{\pm 0.062}$	36.07
Greedy (4-step Euler)	$27.89_{\pm 4.10}$	$0.828_{\pm 0.092}$	$0.151_{\pm 0.061}$	36.94
Greedy (5-step Euler)	$28.27_{\pm 4.01}$	$0.831_{\pm 0.088}$	$0.149_{\pm 0.059}$	35.35

Summary

- End-to-end guidance and posterior guidance are two sides of the same coin
- Greedy guidance is reasonable up to $\mathcal{O}(h^2)$
- This can be improved with high-order solvers $\mathcal{O}(h^{\alpha+1})$ or multiple steps $\mathcal{O}(h/n)$

References

Bansal, A., Chu, H.-M., Schwarzschild, A., Sengupta, S., Goldblum, M., Geiping, J., and Goldstein, T. (2023). “Universal guidance for diffusion models”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 843–852 (cit. on p. 22).

Chung, H., Kim, J., Mccann, M. T., Klasky, M. L., and Ye, J. C. (2023). “Diffusion Posterior Sampling for General Noisy Inverse Problems”. In: *The Eleventh International Conference on Learning Representations, ICLR 2023*. The International Conference on Learning Representations (cit. on p. 22).

Liu, X., Wu, L., Zhang, S., Gong, C., Ping, W., and Liu, Q. (2023). “Flowgrad: Controlling the output of generative odes with gradients”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 24335–24344 (cit. on p. 22).

Yu, J., Wang, Y., Zhao, C., Ghanem, B., and Zhang, J. (2023). “Freedom: Training-free energy-guided conditional diffusion model”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 23174–23184 (cit. on p. 22).

Ben-Hamu, H., Puny, O., Gat, I., Karrer, B., Singer, U., and Lipman, Y. (2024). “D-Flow: Differentiating through Flows for Controlled Generation”. In: *Forty-first International Conference on Machine Learning*. URL: https://openreview.net/forum?id=SE20BFqj6J (cit. on p. 22).

Blasingame and Liu, C. (2024). “AdjointDEIS: Efficient Gradients for Diffusion Models”. In: *Advances in Neural Information Processing Systems*. Ed. by A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang. Vol. 37. Curran Associates, Inc., pp. 2449–2483. URL: https://proceedings.neurips.cc/paper_files/paper/2024/file/04badd3b048315c8c3a0ca17eff723d7-Paper-Conference.pdf (cit. on p. 22).

Pan, J., Liew, J. H., Tan, V., Feng, J., and Yan, H. (2024). "AdjointDPM: Adjoint Sensitivity Method for Gradient Backpropagation of Diffusion Probabilistic Models". In: *The Twelfth International Conference on Learning Representations*. URL: https://openreview.net/forum?id=y33lDRBgWI (cit. on p. 22).

Marion, P., Korba, A., Bartlett, P., Blondel, M., Bortoli, V. D., Doucet, A., Llinares-López, F., Paquette, C., and Berthet, Q. (2025). "Implicit Diffusion: Efficient optimization through stochastic sampling". In: *The 28th International Conference on Artificial Intelligence and Statistics*. URL: https://openreview.net/forum?id=r5F7Z8s0Qk (cit. on p. 22).

Wang, L., Cheng, C., Liao, Y., Qu, Y., and Liu, G. (2025). "Training Free Guided Flow-Matching with Optimal Control". In: *The Thirteenth International Conference on Learning Representations*. URL: https://openreview.net/forum?id=61ss5RA1MM (cit. on p. 22).