Inverse Design for Conditional Distribution Matching

Ori Meidler Shaul Tolkovsky Prof. Or Zuk

Department of Statistics and Data Science, Hebrew University of Jerusalem

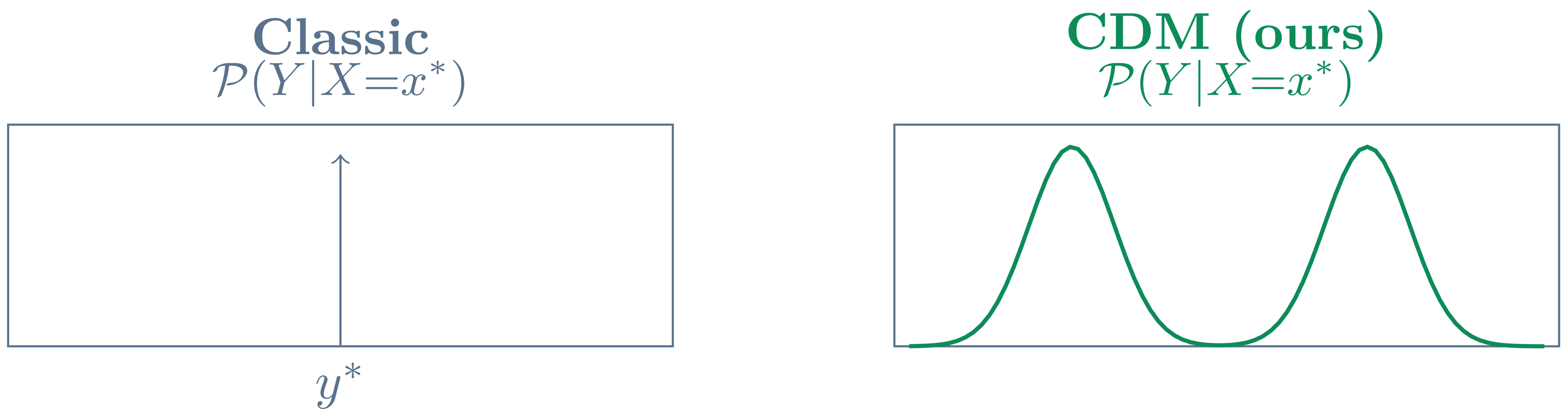
THE HEBREW UNIVERSITY OF JERUSALEM

1. Motivation

Beyond Point Targets

Standard inverse design finds x^* such that $f(x^*) \approx y^*$. Many real design goals are **distributional**:

> Generative pipeline with outputs *balanced across demographic groups*



2. Problem Formulation

Conditional Distribution Matching (CDM)

Given joint distribution $\mathcal{P}(X,Y)$ and a user-specified target $\mathcal{G}(Y)$

Problem 1 — CDMS (Sampling)

Sample from the:

$$\mathcal{Q}_\beta(x) \propto \mathcal{P}(x)\, e^{-\beta\,\mathcal{L}(x)}, \quad \mathcal{L}(x) = \|\mathcal{P}(Y|X{=}x) - \mathcal{G}(Y)\|$$

When $\beta \geq 0$ trades off prior faithfulness vs. minimizing $\mathcal{L}$.

Problem 2 — CDMO (Optimization)

$$x^* = \arg\min_x \|\mathcal{P}(Y \mid X{=}x) - \mathcal{G}(Y)\|$$

Problem 2 is the $\beta{\to}\infty$ limit of Problem 1.

3. Method: MLGD-F

Matching-Loss Guided Diffusion with a Fast Sampler

Given pretrained model for $\mathcal{P}(X)$ (diffusion model with score s_θ) and few-step sampler $f_\phi(x,\eta)$ approximating $\mathcal{P}(Y \mid X{=}x)$.

Alg. 1 — Outer Loop (MLGD)

Input: $x_T \sim \mathcal{N}(0,I)$, score s_θ, sampler f_ϕ, targets $\mathcal{S}_\mathcal{G}$, step sizes ζ_t, noise schedule $\{\bar\alpha_t\}$

Output: optimized x_0^*

for $t = T, T{-}1, \ldots, 1$:

- $\hat{x}_0 \leftarrow \frac{1}{\sqrt{\bar\alpha_t}}(x_t + (1{-}\bar\alpha_t)\, s_\theta(x_t,t))$ — predict $\hat{x}_0$ (Tweedie)
- $\hat{\mathcal{L}} \leftarrow$ Alg.2$(\hat{x}_0, f_\phi, \mathcal{S}_\mathcal{G})$ — matching loss estimator
- $\nabla_{x_t}\hat{\mathcal{L}} \leftarrow$ autograd$(\hat{\mathcal{L}}, x_t)$ — backprop
- $x_{t-1} \leftarrow$ denoise$(x_t, s_\theta, t) - \zeta_t\, \nabla_{x_t}\hat{\mathcal{L}}$ — denoise step + correction

Alg. 2 — Inner Estimator (distribution matching loss)

Input: $\hat{x}_0$, sampler f_ϕ, targets $\mathcal{S}_\mathcal{G}$

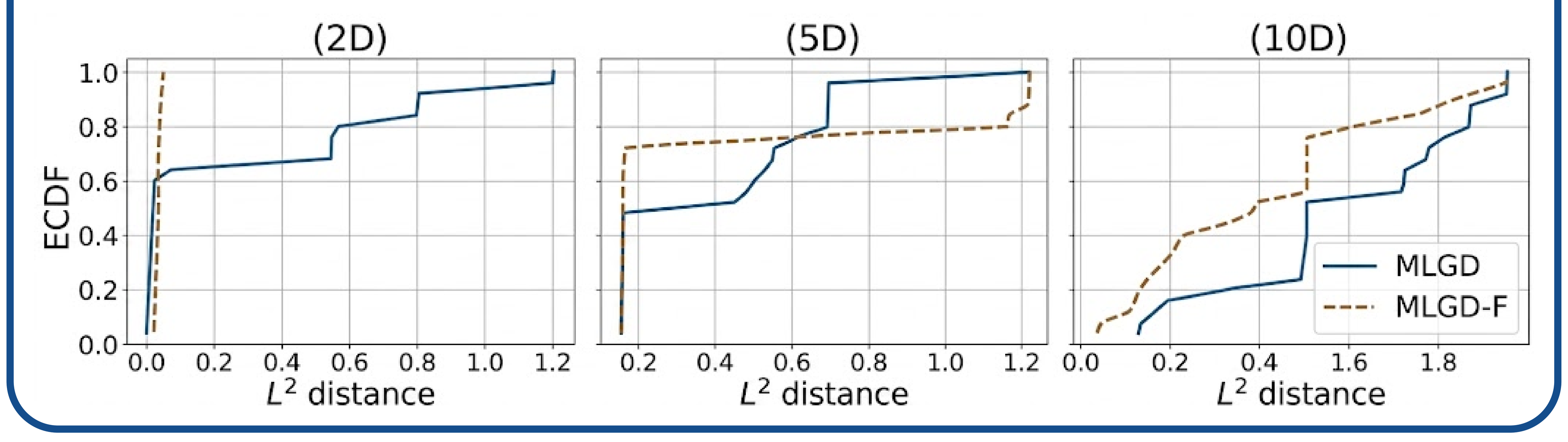
Output: $\hat{\mathcal{L}}$

- Draw $\eta_1,\ldots,\eta_{n_c} \sim \pi$ — n_c noise samples
- $\mathcal{S}_c \leftarrow \{f_\phi(\hat{x}_0,\eta_i)\}$ — conditional samples via f_ϕ
- **return** $\mathcal{L}(\mathcal{S}_c, \mathcal{S}_\mathcal{G})$ — e.g. MMD

When Alg. 2 uses a full diffusion chain the method is **MLGD**; when it uses a fast sampler - **MLGD-F**, where F stands for fast.

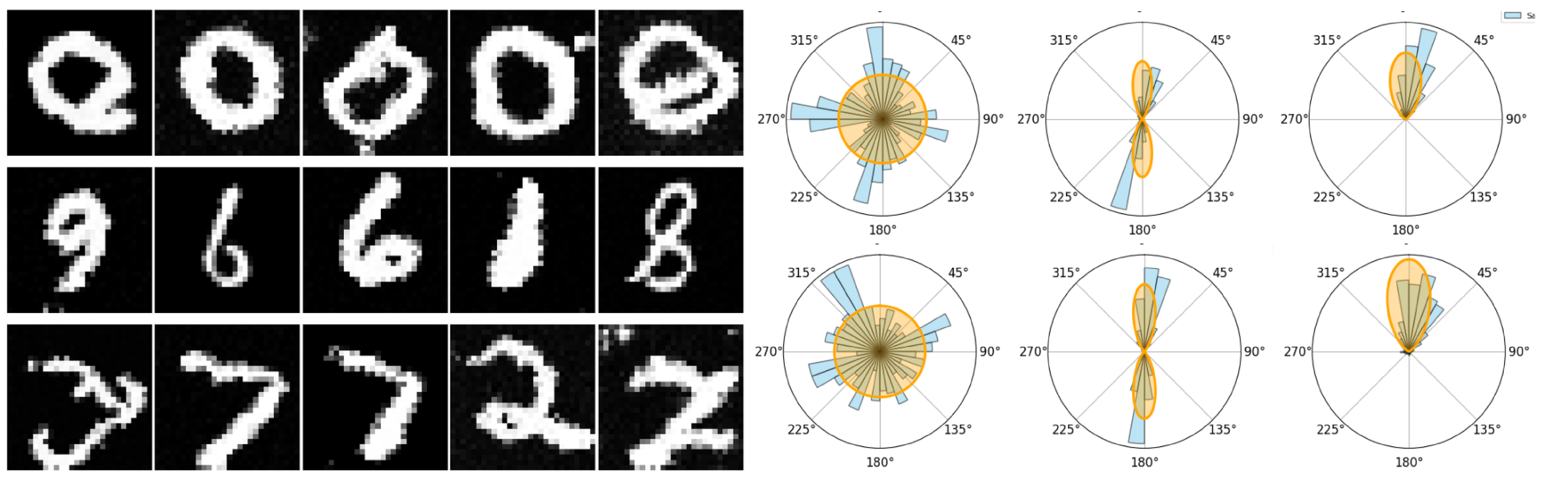
4. Scaling with Dimension: MLGD vs. MLGD-F

Despite MLGD's higher fidelity, its full chain accumulates gradient variance; MLGD-F yields better signal (see plots), 9×–14.8× speedup (2D–10D), and cuts peak VRAM to **43 GB** from projected 375 GB (Exp. 2).



5. Experiment 1: MNIST Rotation

$X \in \mathbb{R}^{784}$: rotated MNIST images; $Y \in \mathbb{R}^2$: rotation angle (polar); $\mathcal{G}$: digit-valid angle distribution. DDPM as $\mathcal{P}(X)$, Consistency model as f_ϕ.

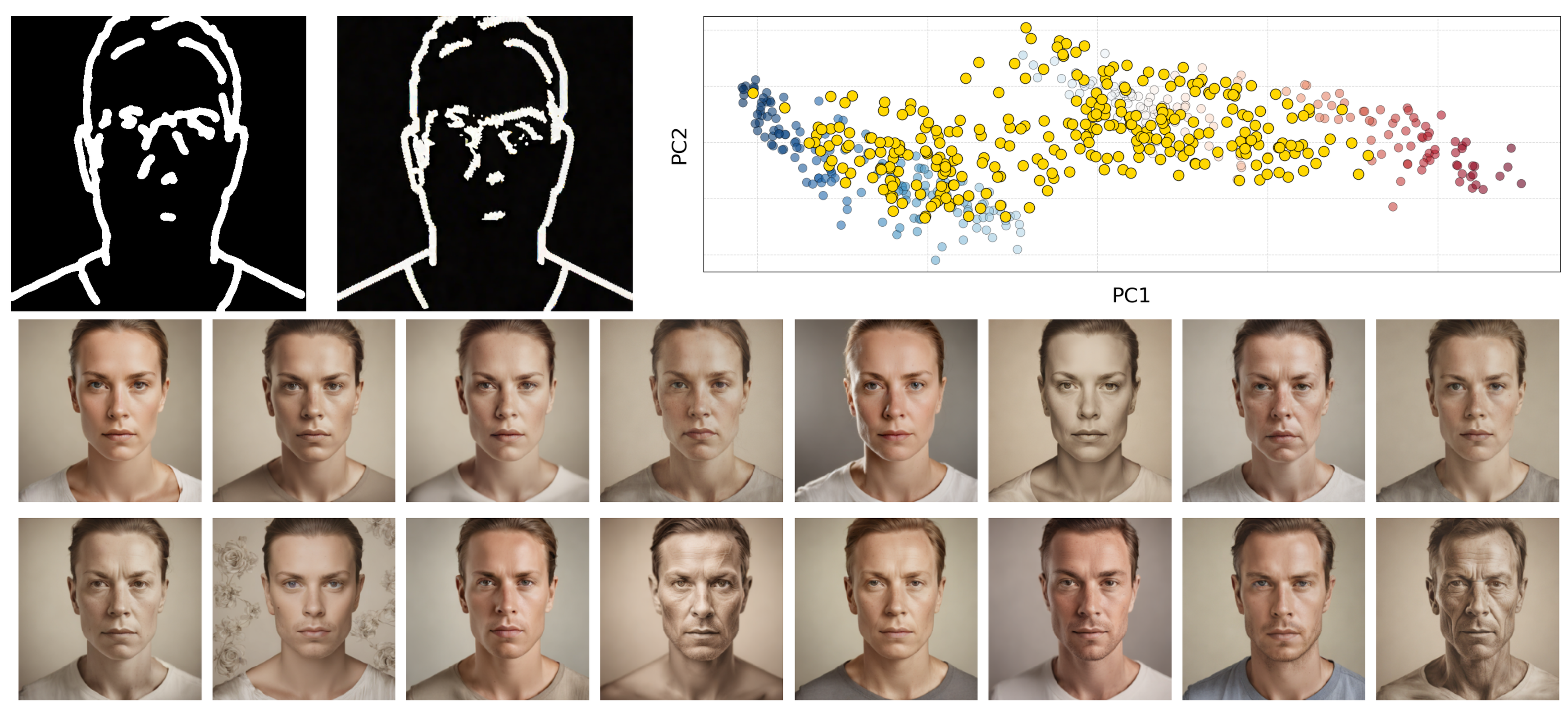


Left: Top-5 x^* by final denoise step $\hat{\mathcal{L}}$ (Alg. 2) per $\mathcal{G}$: uniform, bimodal, unimodal (top-bottom). **Right:** Top-2 x^*-f_ϕ vs. $\mathcal{G}$ (left-right, same order).

MLGD-F recovers **semantically meaningful** digits from rotation geometry: e.g., a uniform target yields **0**s, the only digit valid in all angles.

6. Experiment 2: Scribble-Conditioned Image

$X \in \mathbb{R}^{512\times512}$: scribble image; $Y \in \mathbb{R}^{768}$: CLIP embedding of scribble-conditioned image; $\mathcal{G}$: gender/age distribution. Stable Diffusion XL (SDXL) as $\mathcal{P}(X)$, SDXL-Turbo conditioned on scribble $\hat{x}_0$ as f_ϕ



Top: source scribble (left); optimized x^* (middle); CLIP PCA (right): • male→• female = target $\mathcal{S}_\mathcal{G}$ (gender interpolation); • = interpolated images generated conditioned on x^*. **Bottom:** • images ordered female→male (PC1).

Scenario	MMD Improvement (↑)
MMD: kernel-based distribution distance (↓ better)	
Balanced (50% male)	+32.5%
Skewed (25% male)	+27.1%
Gender interpolation	+22.1%
Age interpolation	+28.5%

Improvements over the **source** (male), where conditioned images are ∼ 100% male. For balanced/skewed targets, MLGD-F achieves 47.4% male (target 50%, CI [45.3, 49.5]) and 26.4% male (target 25%, CI [24.5, 28.3]).

7. Conclusion

Summary

- ✓ First problem class to generalise inverse design to **distributional targets**
- ✓ **MLGD-F** operates at inference time with **no retraining or fine-tuning**
- ✓ Quality improves as stronger pretrained models become available

Limitations

- – Quality bounded by model f_ϕ fidelity
- – Requires end-to-end differentiable f_ϕ
- – Runtime may be prohibitive for latency-sensitive applications