Preconditioned Flow Matching

Reshaping the geometry that flow matching must regress

1 · PRECONDITION

$x_1 \to \tilde{x}_1$

learn a reversible map P

2 · MATCH

$\mathcal{N}(0, I) \to \tilde{x}_1$

train the flow in the new space

3 · MAP BACK

$\hat{x}_1 = P^{-1}(\tilde{x}_1)$

return to the data space

Shadab Ahamed[1] · Eshed Gal[1] · Md Shahriar Rahim Siddiqui[1] · Simon Ghyselincks[1] · Moshe Eliasof[2] · Eldad Haber[1]

[1]University of British Columbia, Vancouver, Canada

[2]Ben-Gurion University of the Negev, Beersheba, Israel

NEURAL INFORMATION PROCESSING SYSTEMS

arXiv 2603.02337

Flow matching, in three steps

Train a velocity field that carries noise onto data, then integrate it to sample.

STEP 1 · PICK A PATH

$$x_t = (1 - t)\, x_{\text{noise}} + t\, x_{\text{data}}$$

STEP 2 · REGRESS VELOCITY

$$v_\theta(x_t, t) \approx x_{\text{data}} - x_{\text{noise}}$$

STEP 3 · INTEGRATE

$$\frac{dx}{dt} = v_\theta(x, t)$$

INTERMEDIATE COVARIANCE $\Sigma_t = \mathrm{Cov}_{x_t \sim p_t}[\, x_t \,]$ Σ_t is the covariance of the intermediate distribution p_t — what the next slide measures

No likelihood, no adversary — just regression on paired samples. The hard part is the path itself.

Every later claim is about the geometry of that path, not about the size of the network.

The condition number

One number decides how hard the transport is to learn.

$$\kappa(\Sigma_t) = \frac{\lambda_{\max}(\Sigma_t)}{\lambda_{\min}(\Sigma_t)}$$

the condition number of Σ_t, the covariance of the intermediate distribution p_t

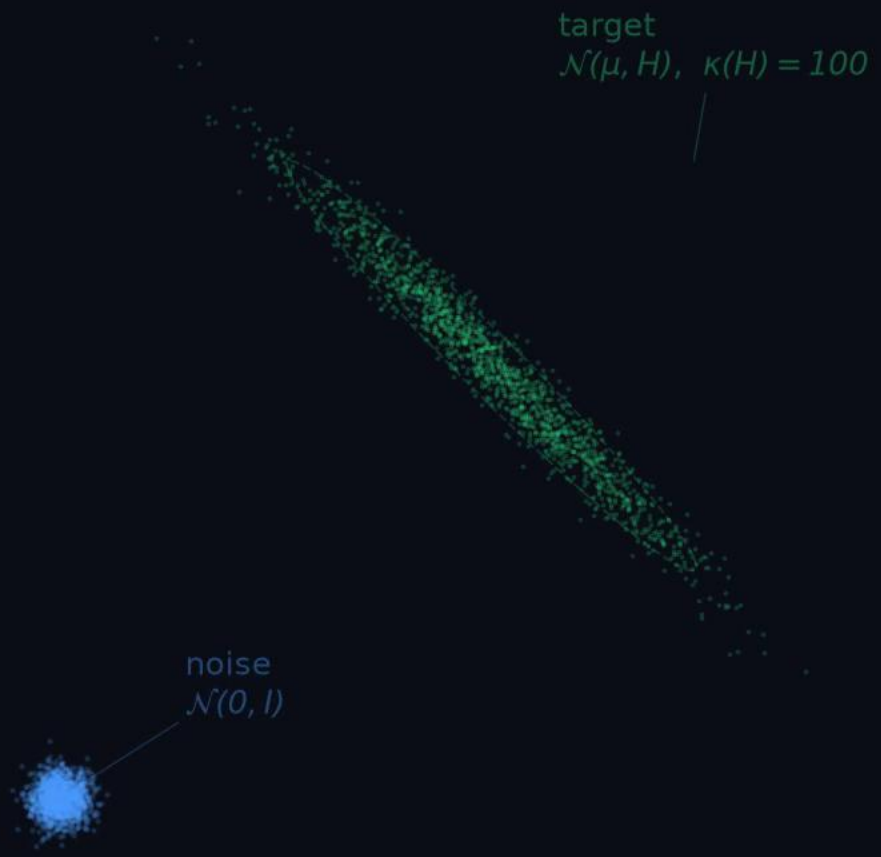
largest eigenvalue over smallest — how many times more spread out the cloud is along its longest axis than its shortest

`κ = 1` isotropic `κ = 100` a thin streak

Everything that follows is measured against this: the spread of the intermediate distributions, not the capacity of the network.

$x_t = \cos(\pi t/2)\, x_0 + \sin(\pi t/2)\, x_1, \quad x_0 \sim \mathcal{N}(0, I),\ x_1 \sim \mathcal{N}(\mu, H)$

$v_t = (\pi/2)[\cos(\pi t/2)\, x_1 - \sin(\pi t/2)\, x_0]$



same MLP, same budget

Flow matching regresses

$$v_\theta(x_t, t) \approx v_t$$

$v_t = (\pi/2)[\cos(\pi t/2)\, x_1 - \sin(\pi t/2)\, x_0]$

not $x_1 - x_0$ — that is only the linear path

along a path whose intermediate distributions carry the target's anisotropy.

$\Sigma_t = \cos^2(\pi t/2)\, I + \sin^2(\pi t/2)\, H$

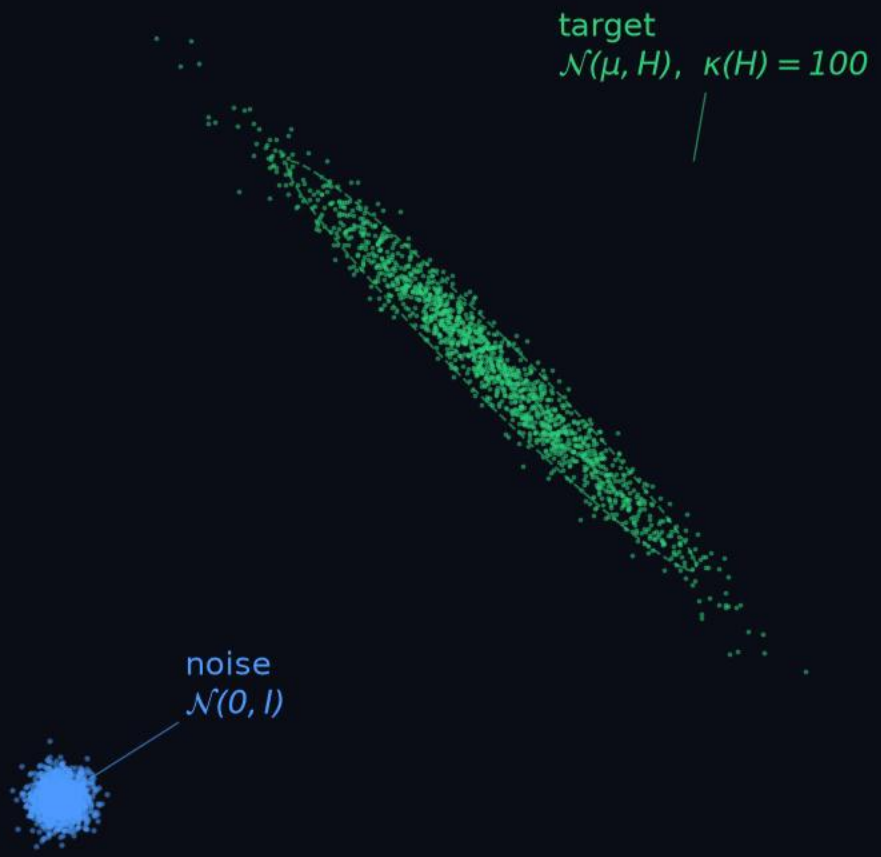
$\kappa(\Sigma_t)$: 1 → 100

Standard flow matching must transport noise onto a strongly anisotropic target.

Every intermediate distribution it regresses on inherits that anisotropy.

$x_t = \cos(\pi t/2)\, x_0 + \sin(\pi t/2)\, x_1, \quad x_0 \sim \mathcal{N}(0, I),\ x_1 \sim \mathcal{N}(\mu, H)$

$v_t = (\pi/2)[\cos(\pi t/2)\, x_1 - \sin(\pi t/2)\, x_0]$



same MLP, same budget

Flow matching regresses

$$v_\theta(x_t, t) \approx v_t$$

$v_t = (\pi/2)[\cos(\pi t/2)\, x_1 - \sin(\pi t/2)\, x_0]$

not $x_1 - x_0$ — that is only the linear path

along a path whose intermediate distributions carry the target's anisotropy.

$\Sigma_t = \cos^2(\pi t/2)\, I + \sin^2(\pi t/2)\, H$

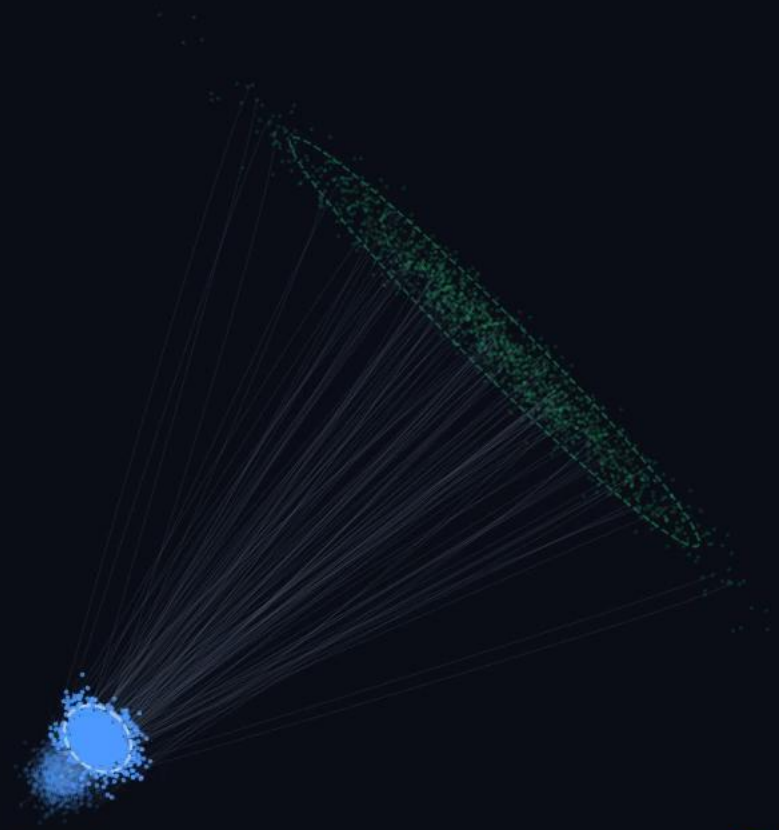
$\kappa(\Sigma_t)$: 1 → 100

Standard flow matching must transport noise onto a strongly anisotropic target.

Every intermediate distribution it regresses on inherits that anisotropy.

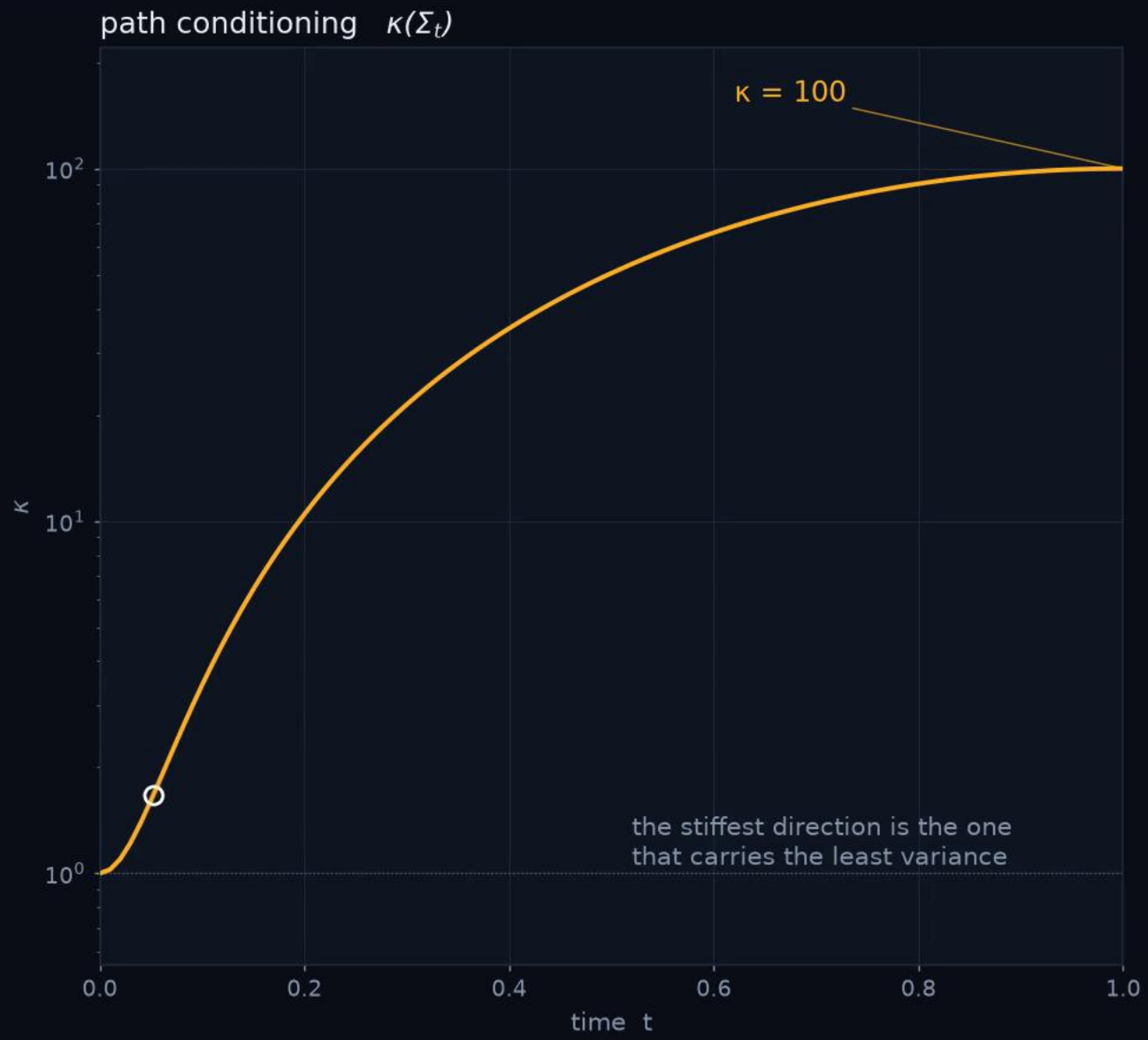
$x_t = \cos(\pi t/2)\, x_0 + \sin(\pi t/2)\, x_1, \quad x_0 \sim \mathcal{N}(0, I), \; x_1 \sim \mathcal{N}(\mu, H)$

$v_t = (\pi/2)[\cos(\pi t/2)\, x_1 - \sin(\pi t/2)\, x_0]$



t = 0.05

$\kappa(\Sigma_t) = 1.7$

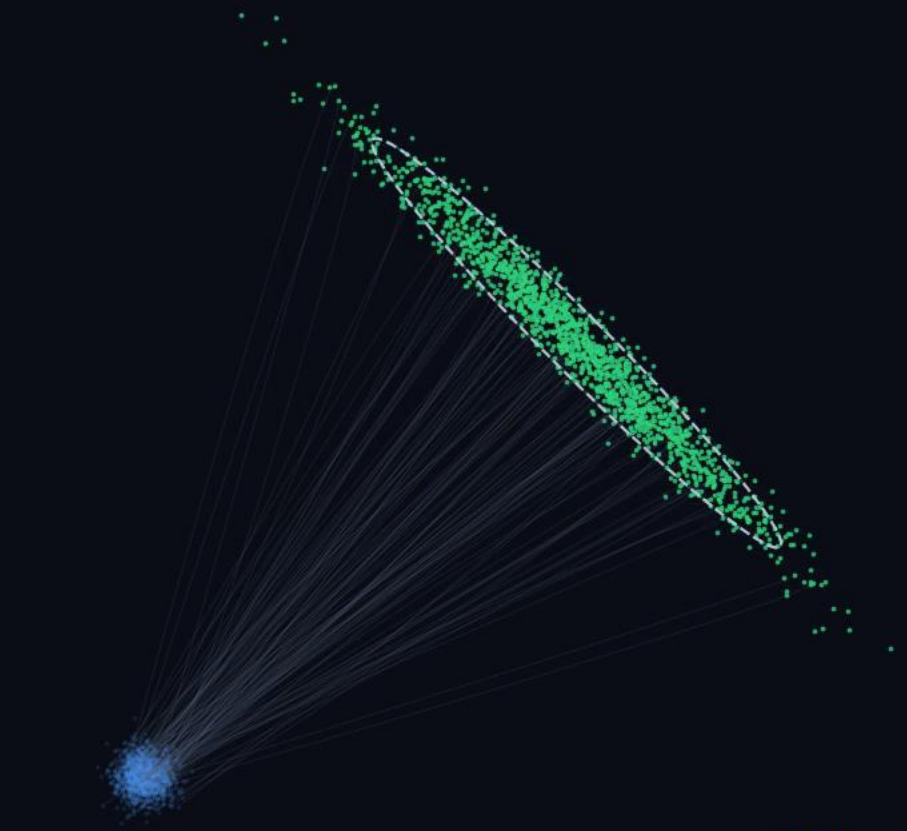


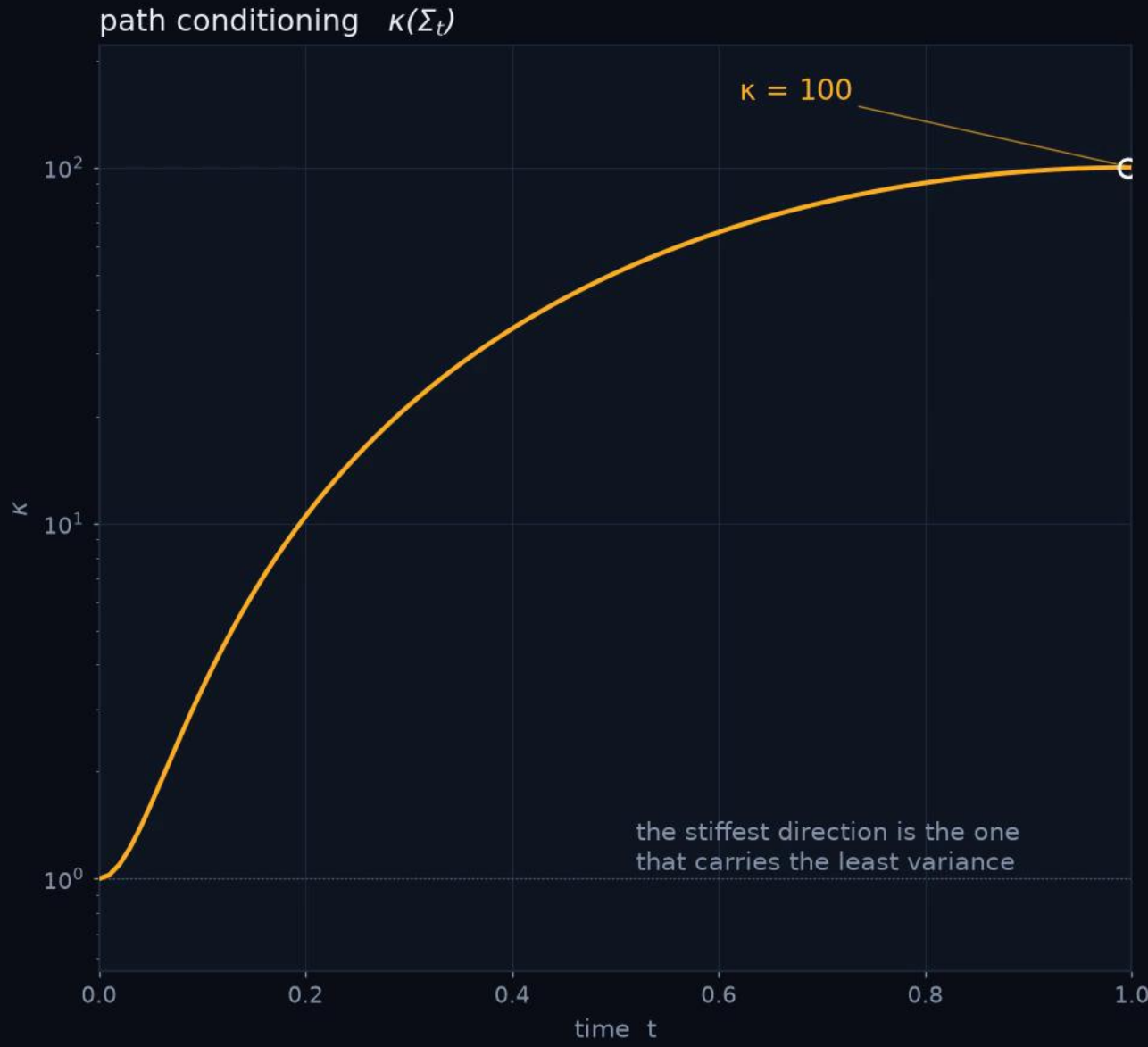
As t → 1 the regression problem becomes 100× stiffer in the narrow direction.

Gradient steps fit the wide direction first; the narrow one keeps lagging.

$x_t = \cos(\pi t/2)\, x_0 + \sin(\pi t/2)\, x_1, \quad x_0 \sim \mathcal{N}(0, I), \; x_1 \sim \mathcal{N}(\mu, H)$

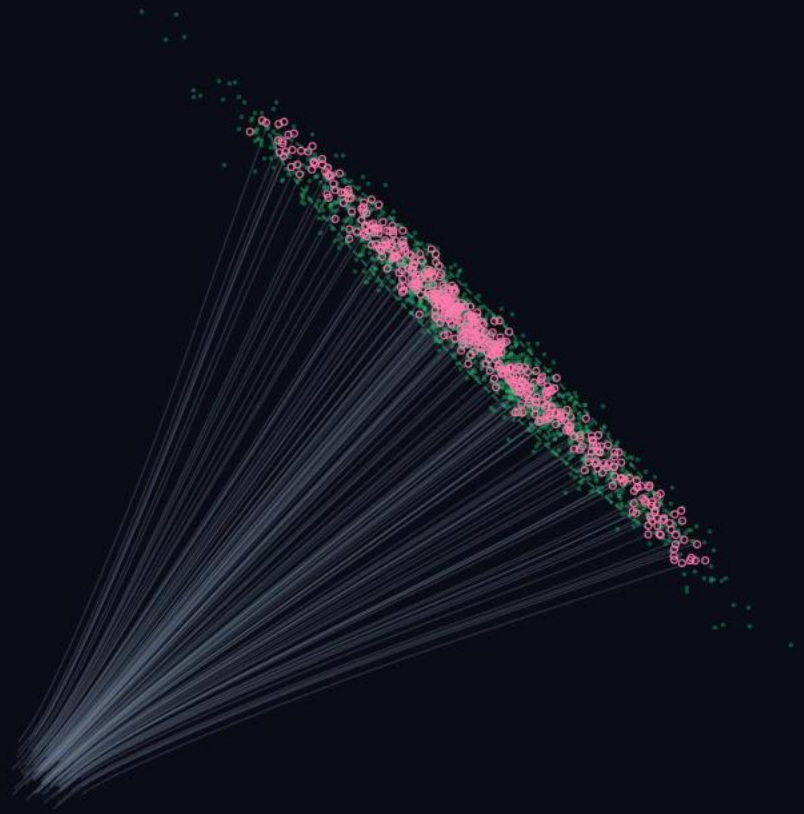
$v_t = (\pi/2)[\cos(\pi t/2)\, x_1 - \sin(\pi t/2)\, x_0]$





As t → 1 the regression problem becomes 100× stiffer in the narrow direction.

Gradient steps fit the wide direction first; the narrow one keeps lagging.







The model reaches the wide direction but under-fills the narrow one.

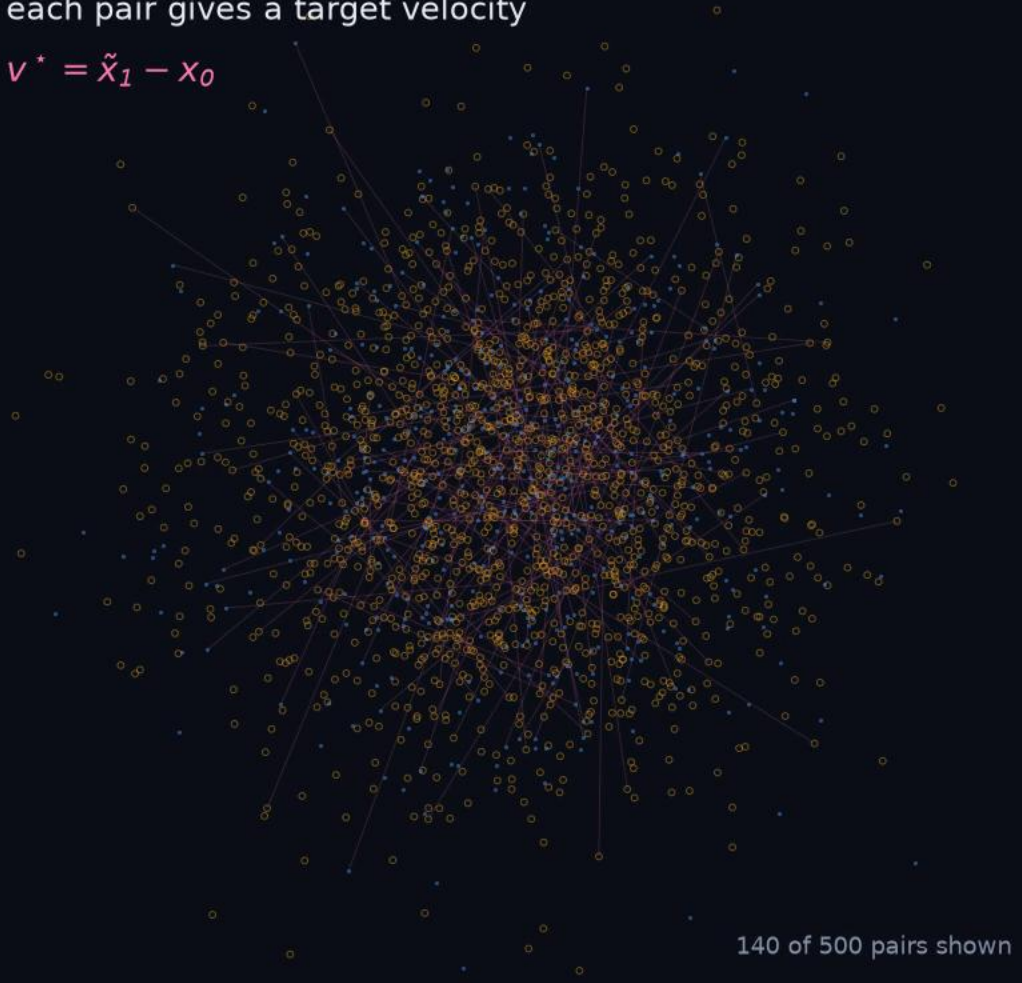
A small FM loss still hides a large velocity error where $\sigma_i(t)$ is small.

$\tilde{x}_t = \cos(\pi t/2)\, x_0 + \sin(\pi t/2)\, \tilde{x}_1, \quad \tilde{x}_1 = P(x_1), \quad \mathrm{Cov}(\tilde{x}_1) \approx I$

$v_t = (\pi/2)[\cos(\pi t/2)\, \tilde{x}_1 - \sin(\pi t/2)\, x_0]$

each pair gives a target velocity

$v^\star = \tilde{x}_1 - x_0$



140 of 500 pairs shown

what the network has to fit

every sample is paired with a target sample; the velocity between them is the label.

the pairs cross each other — no single deterministic map explains them.

what is learnable is their conditional mean: a field.

Noise samples are paired with preconditioned target samples.

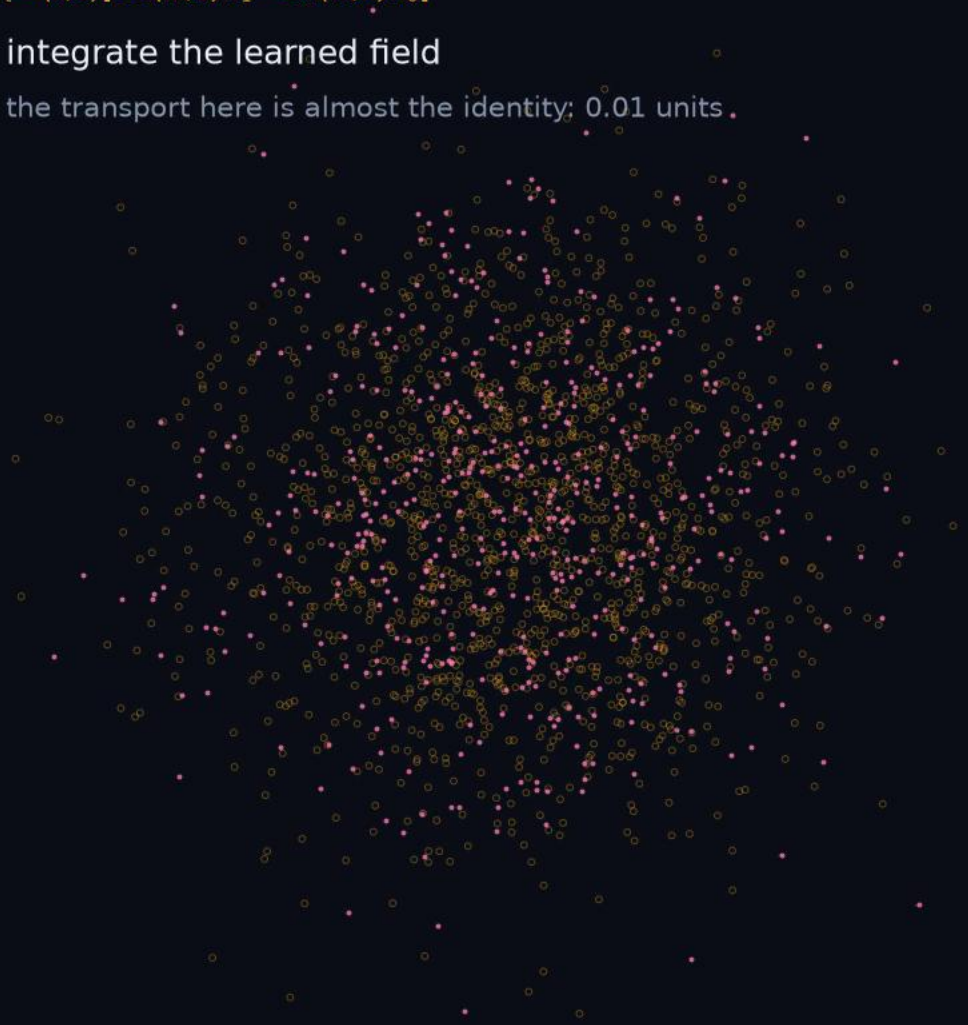
Their difference is the label the network fits — the same objective as before.

$\tilde{x}_t = \cos(\pi t/2)\, x_0 + \sin(\pi t/2)\, \tilde{x}_1, \quad \tilde{x}_1 = P(x_1), \quad \mathrm{Cov}(\tilde{x}_1) \approx I$

$v_t = (\pi/2)[\cos(\pi t/2)\, \tilde{x}_1 - \sin(\pi t/2)\, x_0]$

integrate the learned field

the transport here is almost the identity: 0.01 units



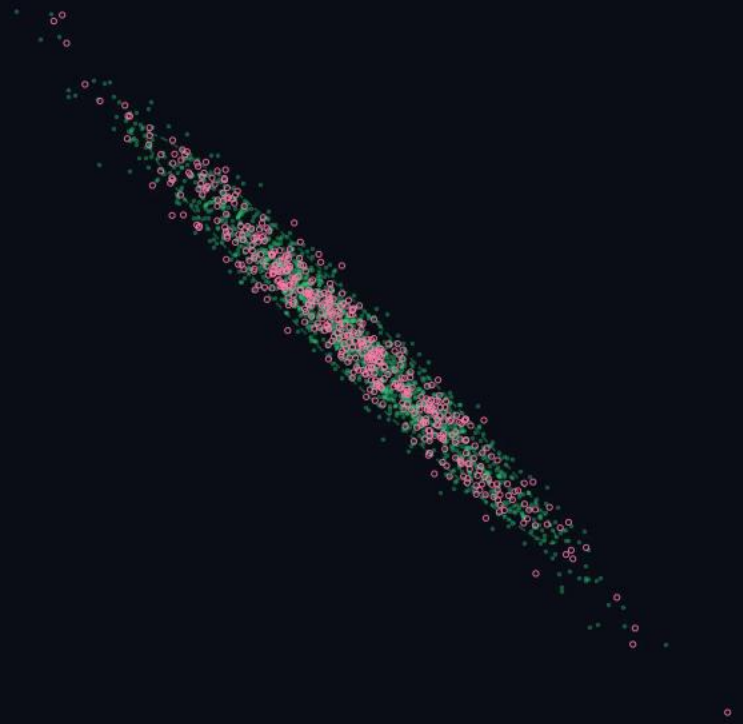
In z-space the model learns a near-identity field — 0.01 units of motion,

against 43 units in the data space. That is the conditioning doing its job.

$\tilde{x}_t = \cos(\pi t/2)\, x_0 + \sin(\pi t/2)\, \tilde{x}_1, \quad \tilde{x}_1 = P(x_1), \quad \mathrm{Cov}(\tilde{x}_1) \approx I$

$v_t = (\pi/2)[\cos(\pi t/2)\, \tilde{x}_1 - \sin(\pi t/2)\, x_0]$

$$\hat{x}_1 = P^{-1}(\hat{\tilde{x}}_1)$$



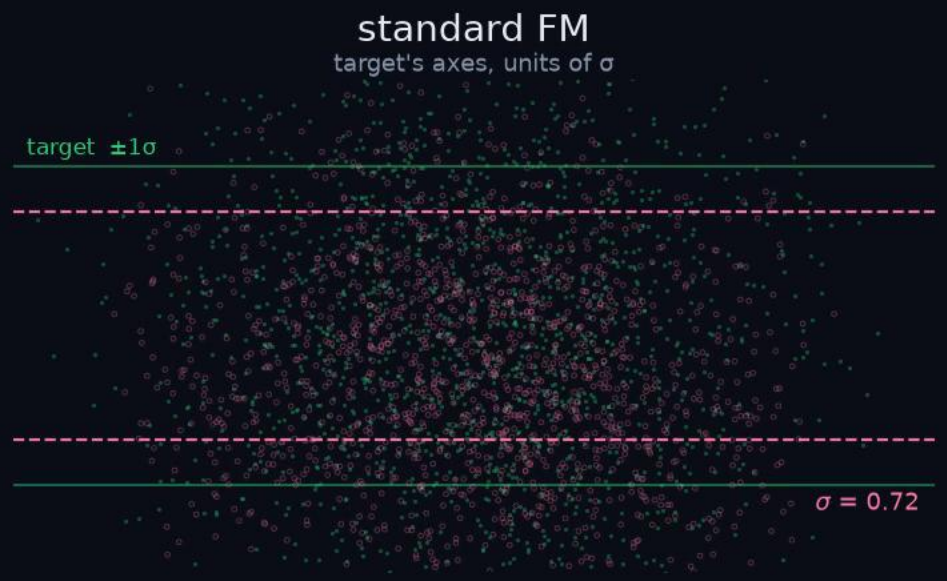


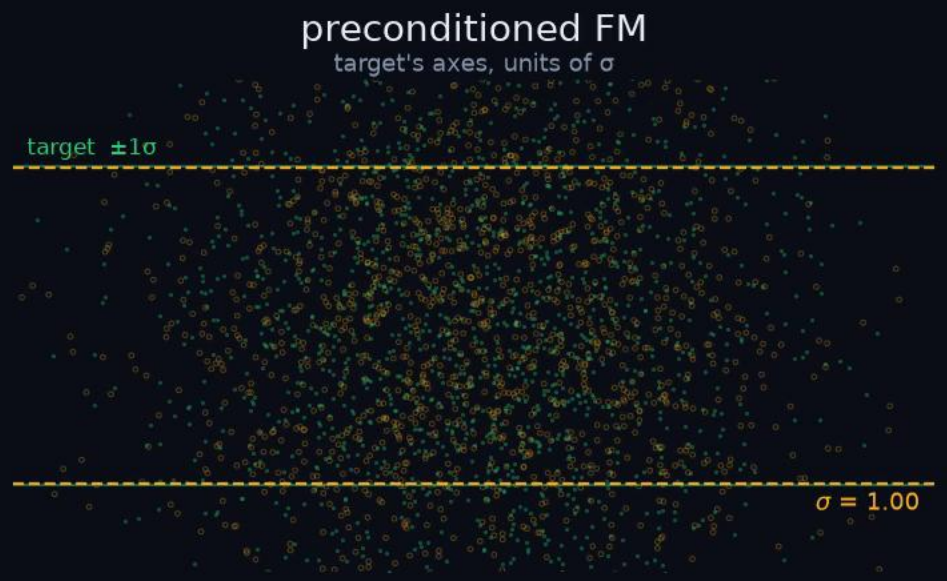
the samples land on the elongated target with the right width — the part standard FM under-fits.

MMD 0.001 vs 0.004

Mapping back with P^{-1} restores the anisotropy the flow never had to model.

The samples land on the target distribution, width included.





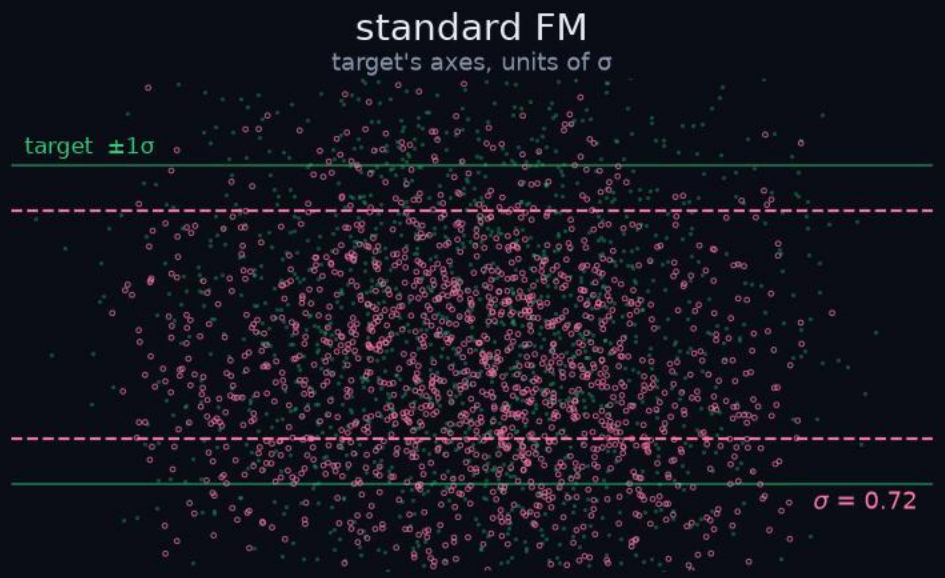


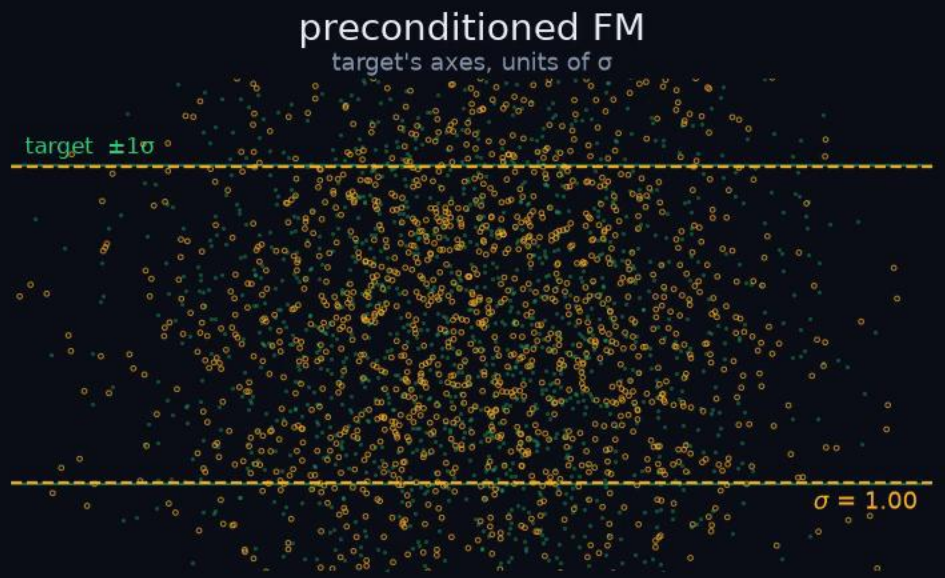
width across the narrow axis (σ)

standard FM	0.72× target
preconditioned	1.00× target
MMD	0.001 vs 0.009

Standard FM reaches the target's length but under-fills its width (0.72×).

Preconditioned FM reproduces both (1.00×) — the alignment standard FM never reaches.







width across the narrow axis (σ)

standard FM	0.72× target
preconditioned	1.00× target
MMD	0.001 vs 0.009

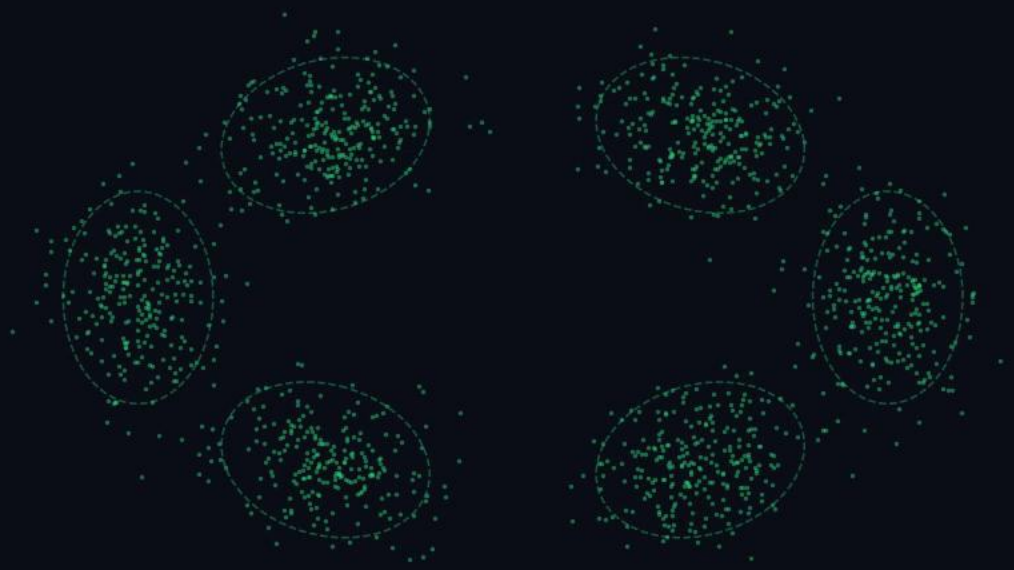
Standard FM reaches the target's length but under-fills its width (0.72×).

Preconditioned FM reproduces both (1.00×) — the alignment standard FM never reaches.

$$x_t = \cos(\pi t/2)\, x_0 + \sin(\pi t/2)\, x_1, \quad x_1 \sim \sum_k w_k\, \mathcal{N}(\mu_k, \Sigma_k)$$

$$v_t = (\pi/2)[\cos(\pi t/2)\, x_1 - \sin(\pi t/2)\, x_0]$$

six elongated modes, off the noise's scale



what the target demands

the flow has to haul mass a long way and land on six modes, each a few units wide.

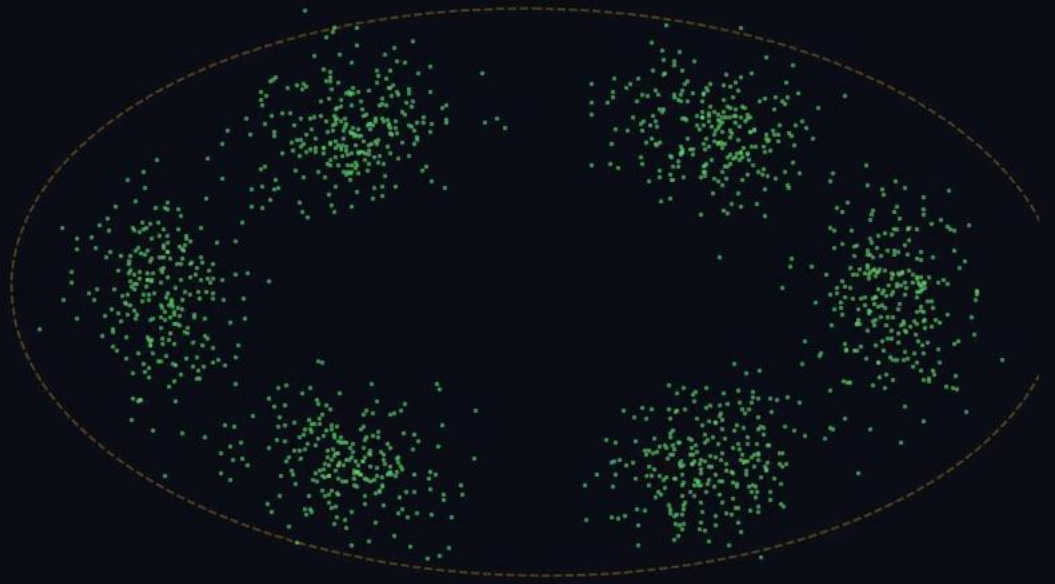
$\kappa(\Sigma) = 3.7 \quad \cdot \quad \kappa(H_k) \approx 2.0$

whitening fixes the mixture, not the shape of each mode.

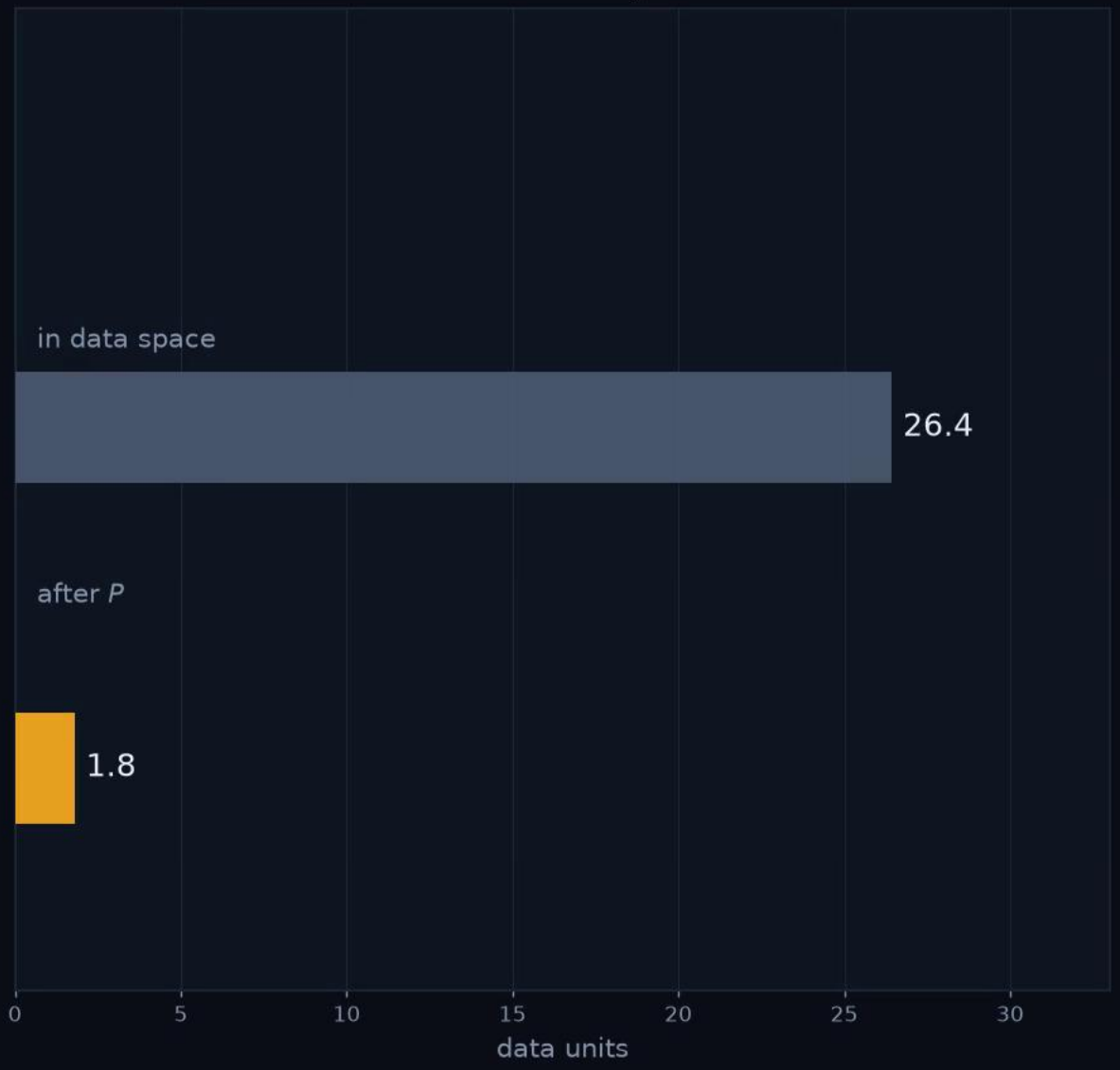
A target the flow cannot trivialise: six modes on an ellipse, far from the noise.

P applied to the target

unit covariance, same six modes



distance the flow must move each sample



P maps the target to unit covariance and unit scale.

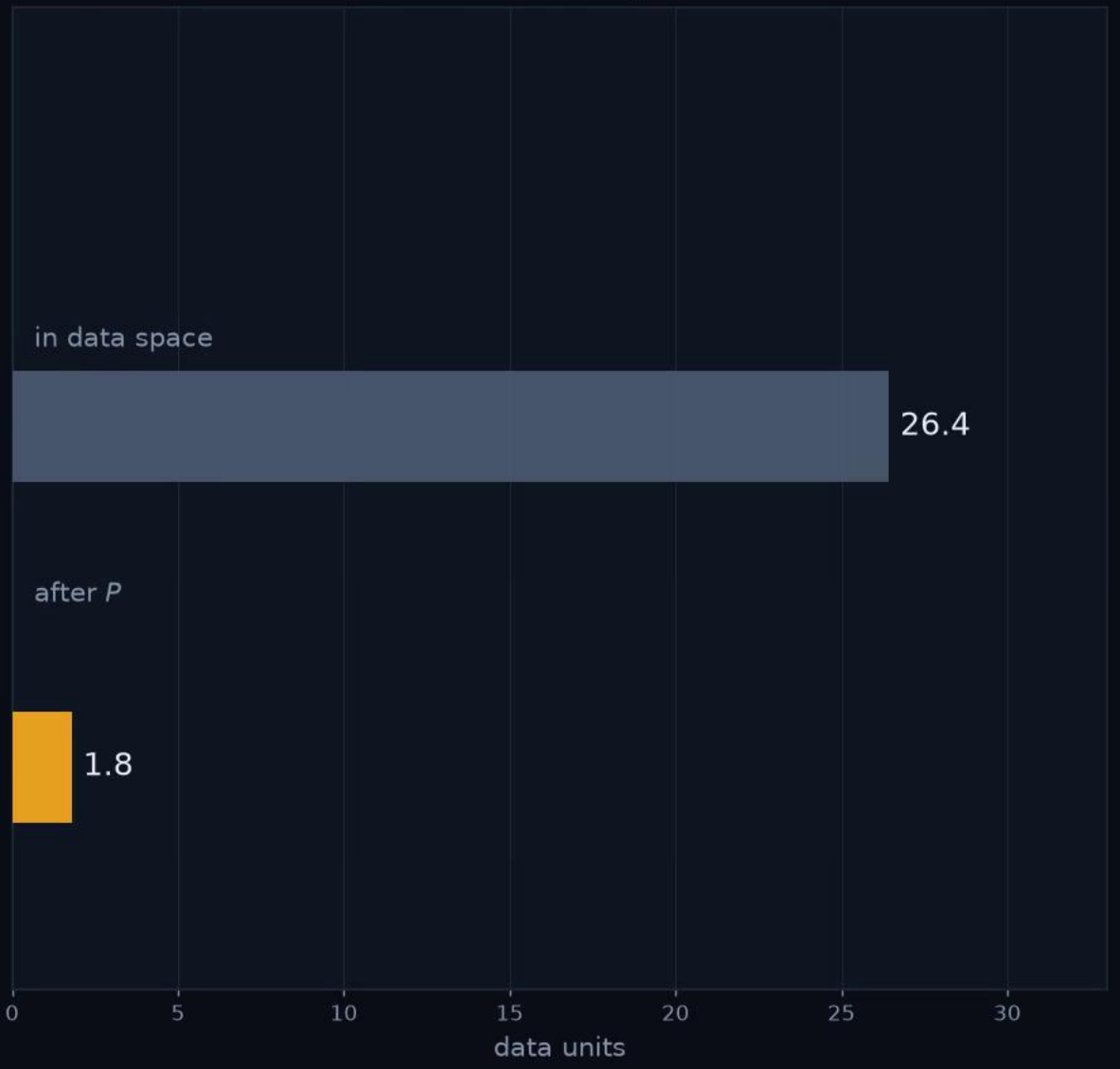
The flow reshapes the noise locally instead of hauling mass 26 units onto the modes.

P applied to the target

unit covariance, same six modes



distance the flow must move each sample



P maps the target to unit covariance and unit scale.

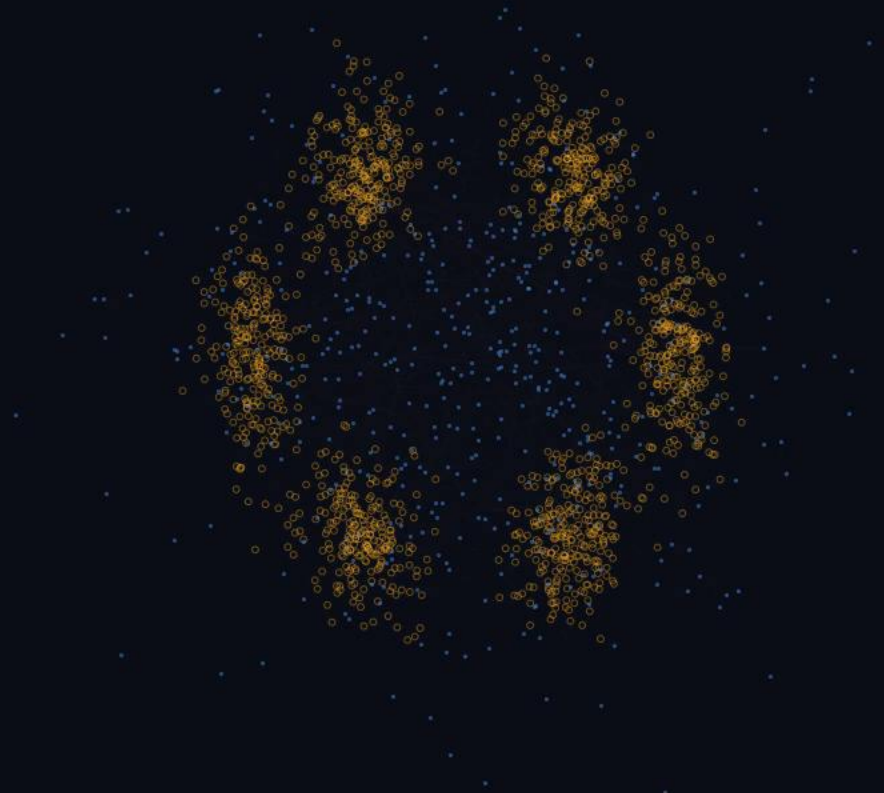
The flow reshapes the noise locally instead of hauling mass 26 units onto the modes.

$$x_t = \cos(\pi t/2)\, x_0 + \sin(\pi t/2)\, x_1, \quad x_1 \sim \sum_k w_k\, \mathcal{N}(\mu_k, \Sigma_k)$$

$$v_t = (\pi/2)[\cos(\pi t/2)\, x_1 - \sin(\pi t/2)\, x_0]$$

each pair gives a target velocity

$$v^\star = \tilde{x}_1 - x_0$$



240 of 500 pairs shown

Noise samples are paired with target samples; the velocity between them is the label.

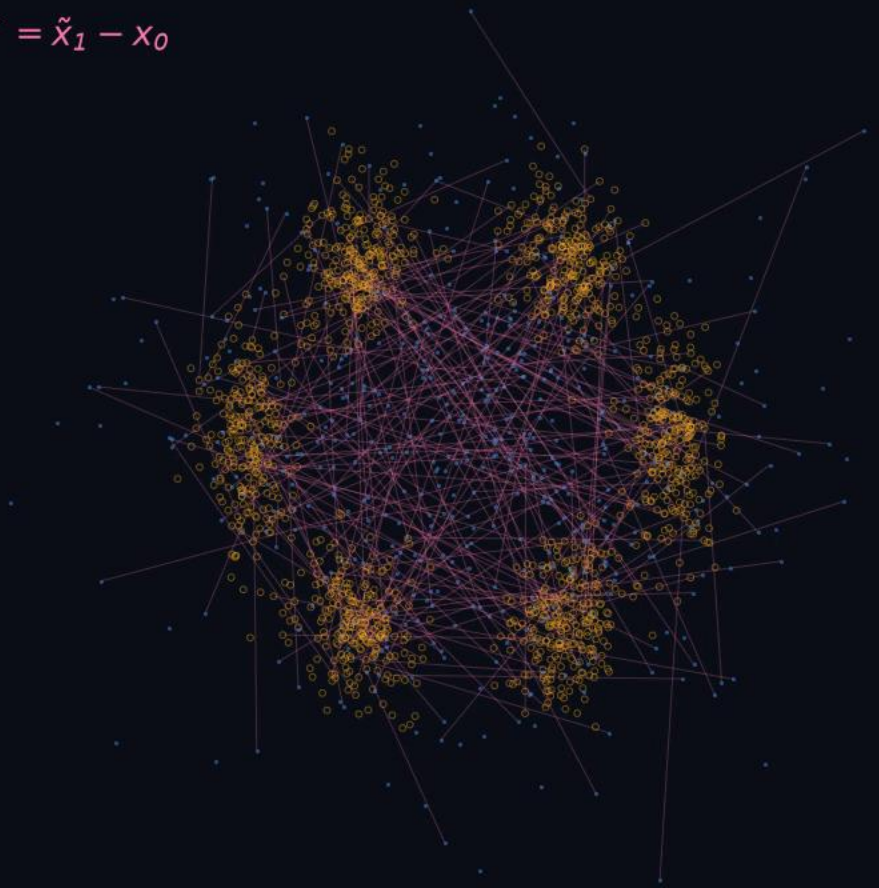
The pairs cross — the network has to learn their conditional mean, a smooth field.

$$x_t = \cos(\pi t/2)\, x_0 + \sin(\pi t/2)\, x_1, \quad x_1 \sim \sum_k w_k\, \mathcal{N}(\mu_k, \Sigma_k)$$

$$v_t = (\pi/2)[\cos(\pi t/2)\, x_1 - \sin(\pi t/2)\, x_0]$$

each pair gives a target velocity

$$v^\star = \tilde{x}_1 - x_0$$



240 of 500 pairs shown

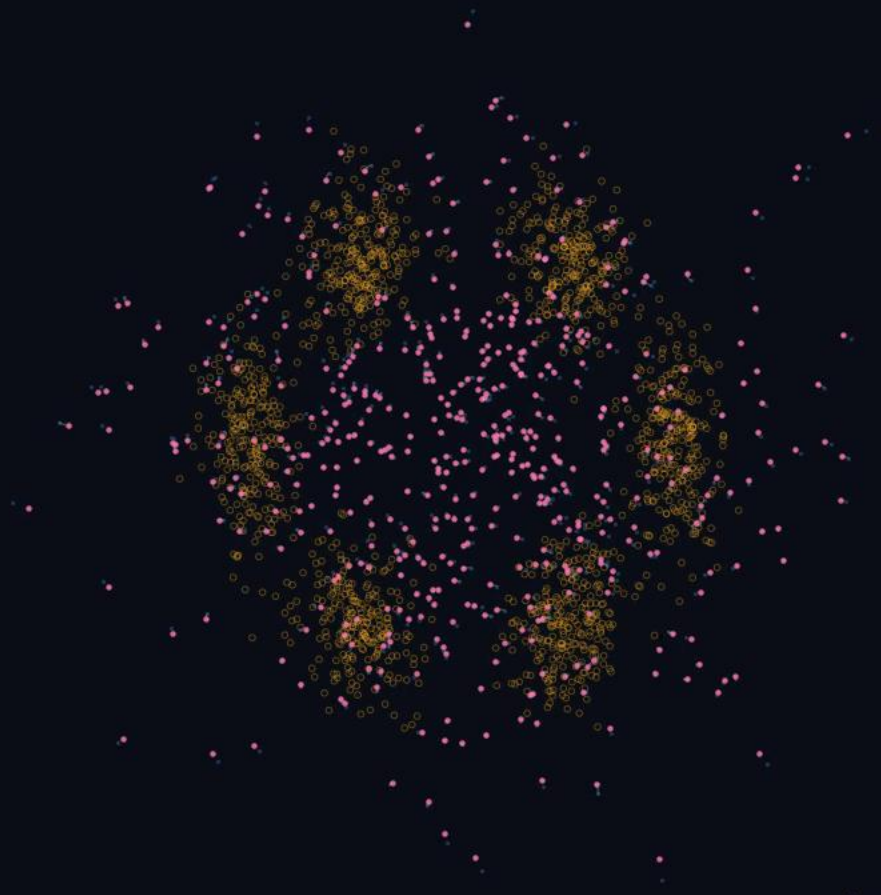
Noise samples are paired with target samples; the velocity between them is the label.

The pairs cross — the network has to learn their conditional mean, a smooth field.

$\tilde{x}_t = \cos(\pi t/2)\, x_0 + \sin(\pi t/2)\, \tilde{x}_1, \quad \tilde{x}_1 = P(x_1)$

$v_t = (\pi/2)[\cos(\pi t/2)\, \tilde{x}_1 - \sin(\pi t/2)\, x_0]$

integrate $dx/dt = v_\theta(x, t)$



t = 0.21

the same MLP, retrained here

one field, integrated from noise to the preconditioned target: short paths, all at unit scale, none of them long.

the modes are where the flow does its real work.

1.8 units per sample, not 26

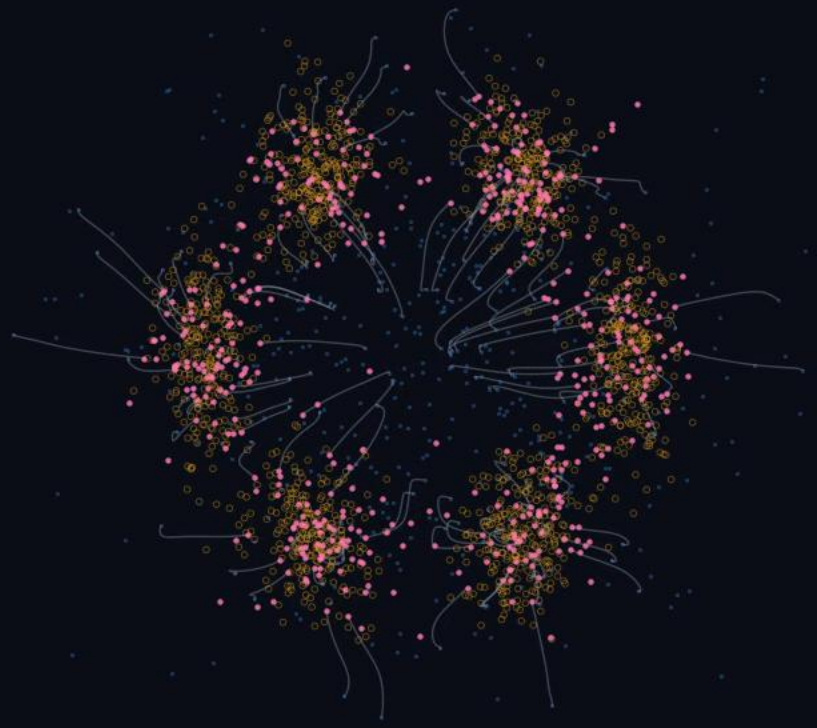
Integrating the learned field carries noise onto the preconditioned target.

Short, well-conditioned paths: this is what the better geometry buys.

$\tilde{x}_t = \cos(\pi t/2)\, x_0 + \sin(\pi t/2)\, \tilde{x}_1, \quad \tilde{x}_1 = P(x_1)$

$v_t = (\pi/2)[\cos(\pi t/2)\, \tilde{x}_1 - \sin(\pi t/2)\, x_0]$

integrate $dx/dt = v_\theta(x, t)$



t = 1.00

the same MLP, retrained here

one field, integrated from noise to the preconditioned target: short paths, all at unit scale, none of them long.

the modes are where the flow does its real work.

1.8 units per sample, not 26

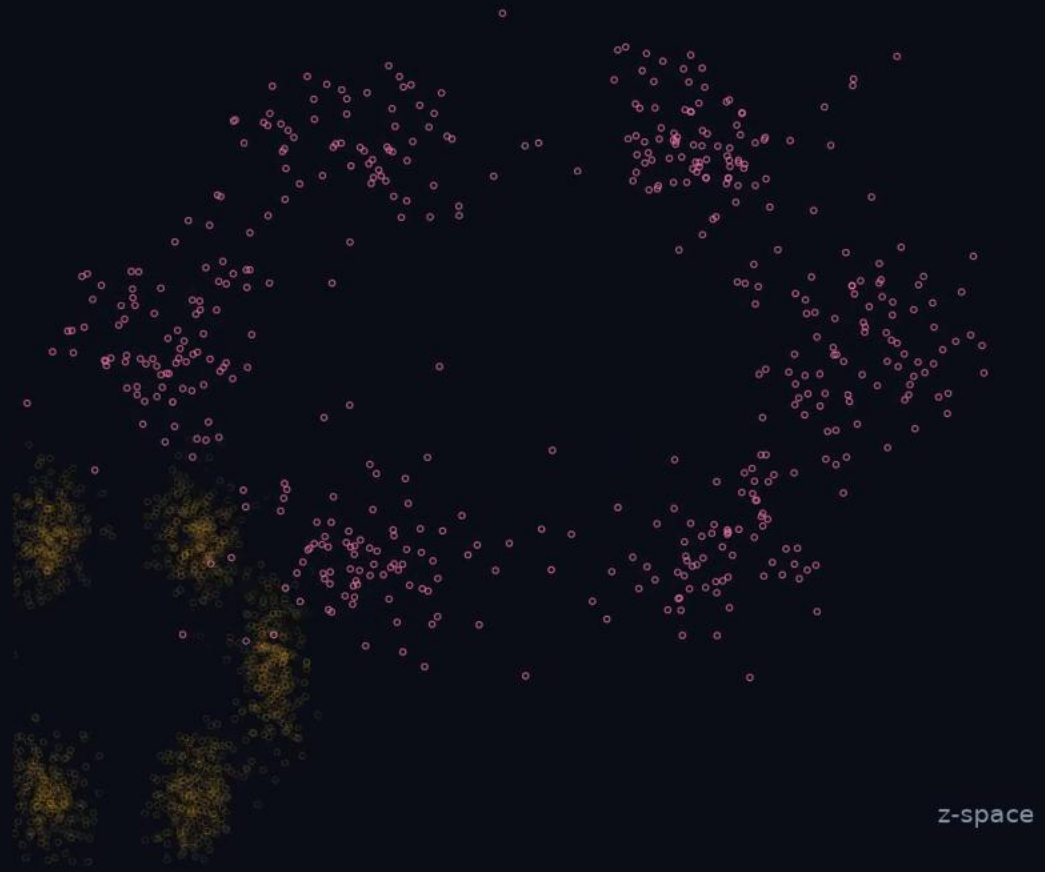
Integrating the learned field carries noise onto the preconditioned target.

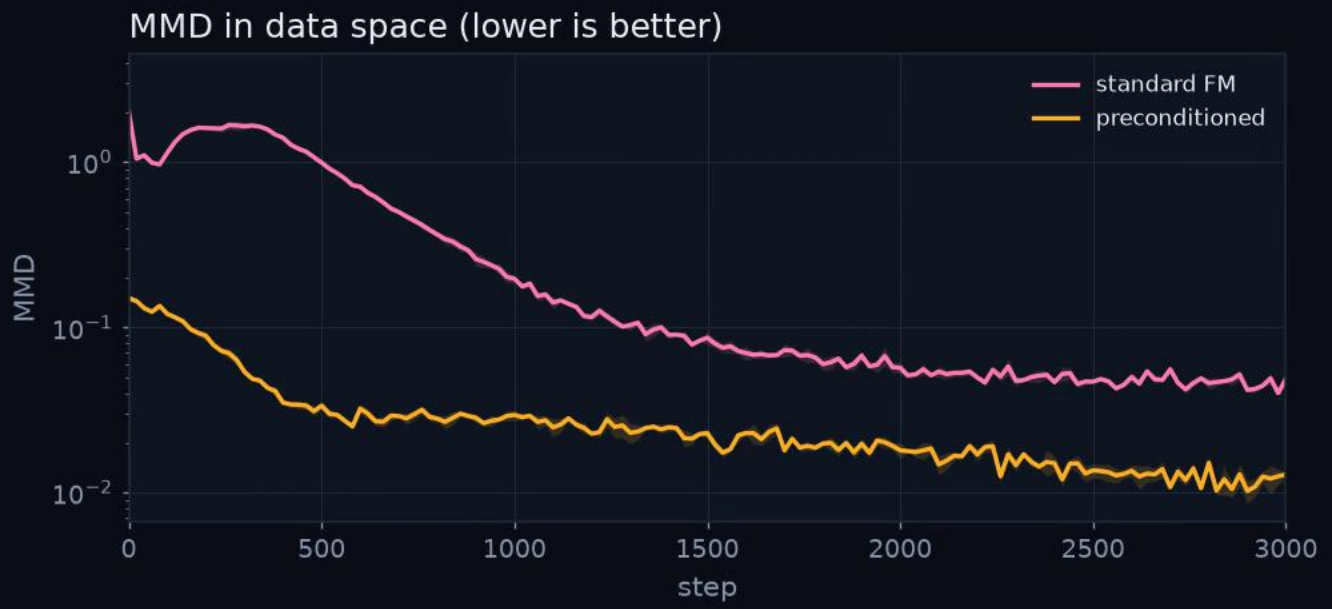
Short, well-conditioned paths: this is what the better geometry buys.

$\tilde{x}_t = \cos(\pi t/2)\, x_0 + \sin(\pi t/2)\, \tilde{x}_1, \quad \tilde{x}_1 = P(x_1)$

$v_t = (\pi/2)[\cos(\pi t/2)\, \tilde{x}_1 - \sin(\pi t/2)\, x_0]$

$$\hat{x}_1 = P^{-1}(\hat{\tilde{x}}_1)$$





the samples stretch back onto the original target, in its own units — the whitened cloud is left behind.

MMD 0.013 vs 0.049

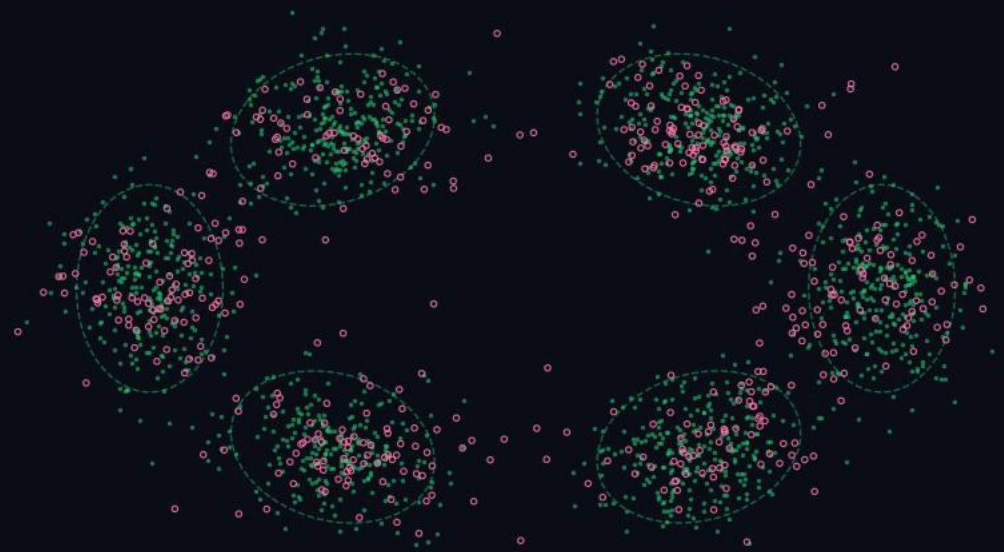
P^{-1} is applied to the transported points: they lift out of z-space onto the target.

The distribution they land on is the original one, not the whitened one.

$\tilde{x}_t = \cos(\pi t/2)\, x_0 + \sin(\pi t/2)\, \tilde{x}_1, \quad \tilde{x}_1 = P(x_1)$

$v_t = (\pi/2)[\cos(\pi t/2)\, \tilde{x}_1 - \sin(\pi t/2)\, x_0]$

$$\hat{x}_1 = P^{-1}(\hat{\tilde{x}}_1)$$



data space



the samples stretch back onto the original target, in its own units — the whitened cloud is left behind.

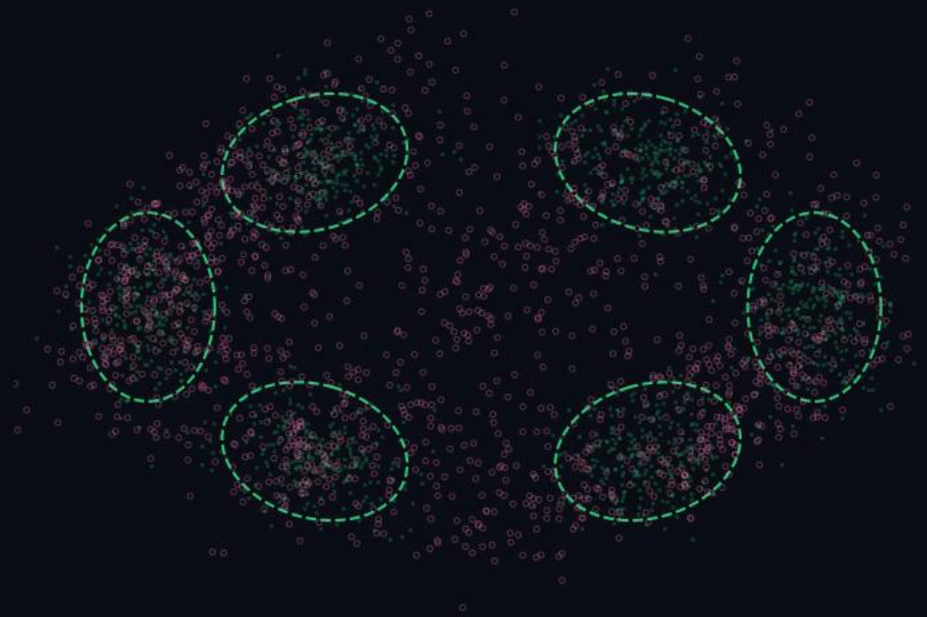
MMD 0.013 vs 0.049

P^{-1} is applied to the transported points: they lift out of z-space onto the target.

The distribution they land on is the original one, not the whitened one.

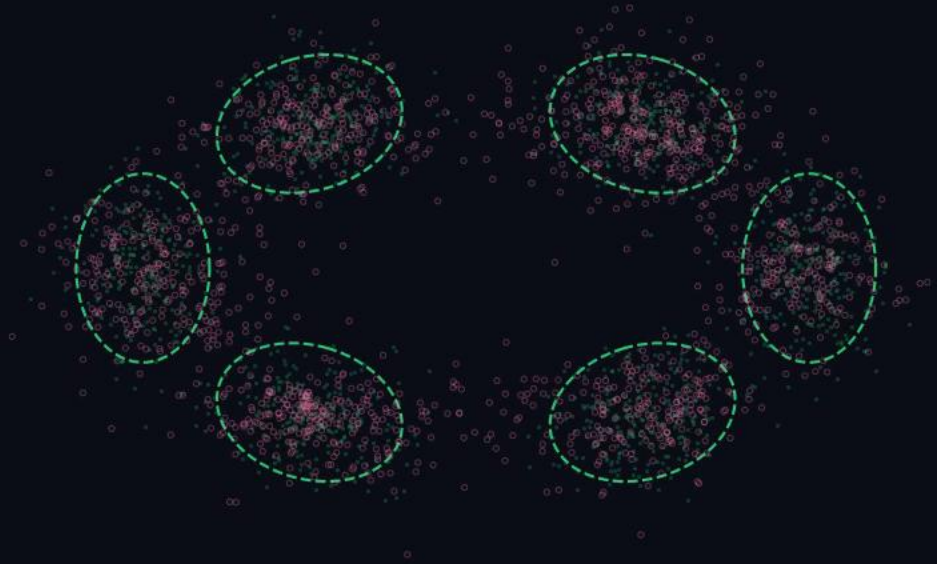
standard FM

MMD 0.049



preconditioned FM

MMD 0.013

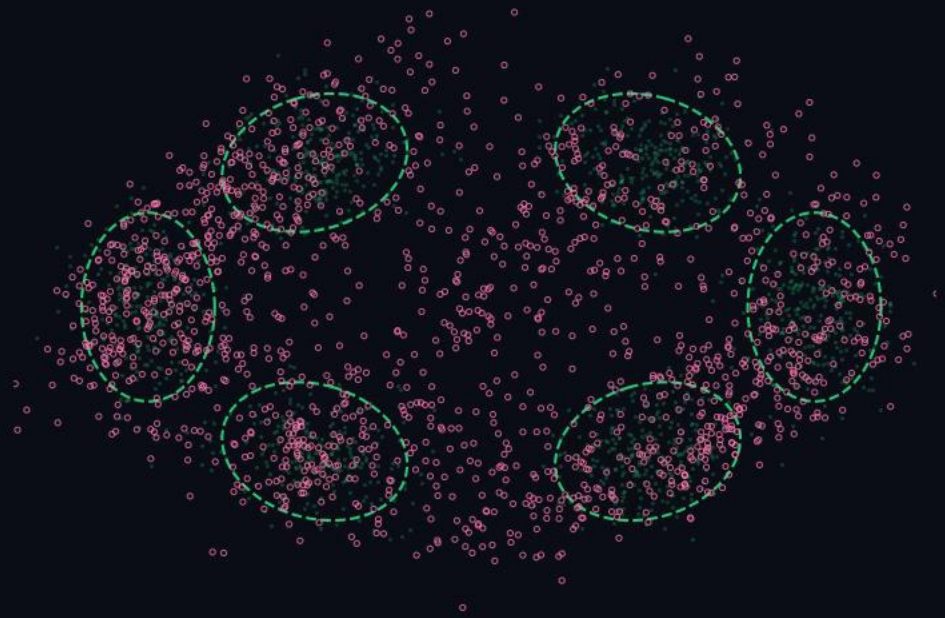


Same architecture, same 3k-step budget. Standard FM smears the modes.

Preconditioned FM lands on all six — 3.8× lower MMD.

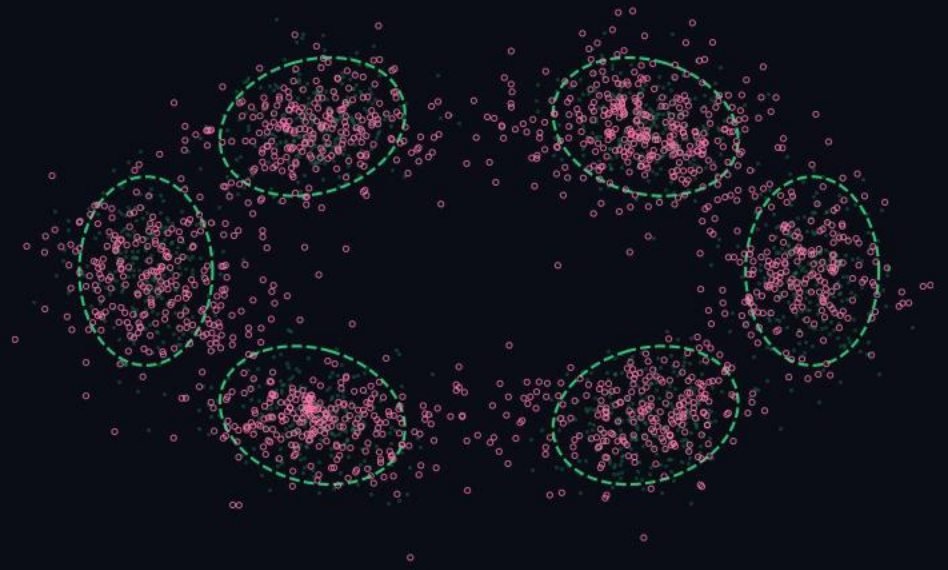
standard FM

MMD 0.049



preconditioned FM

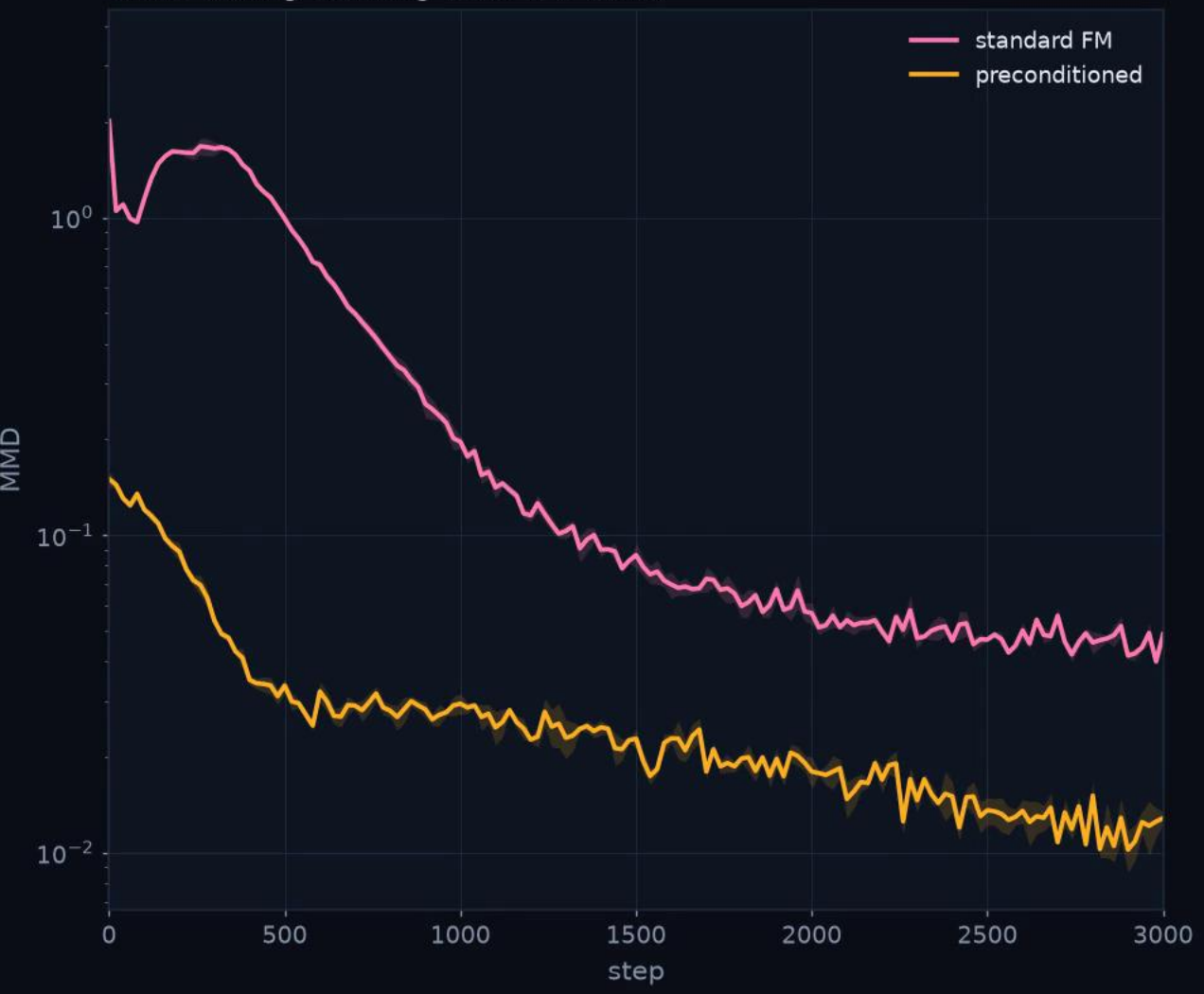
MMD 0.013



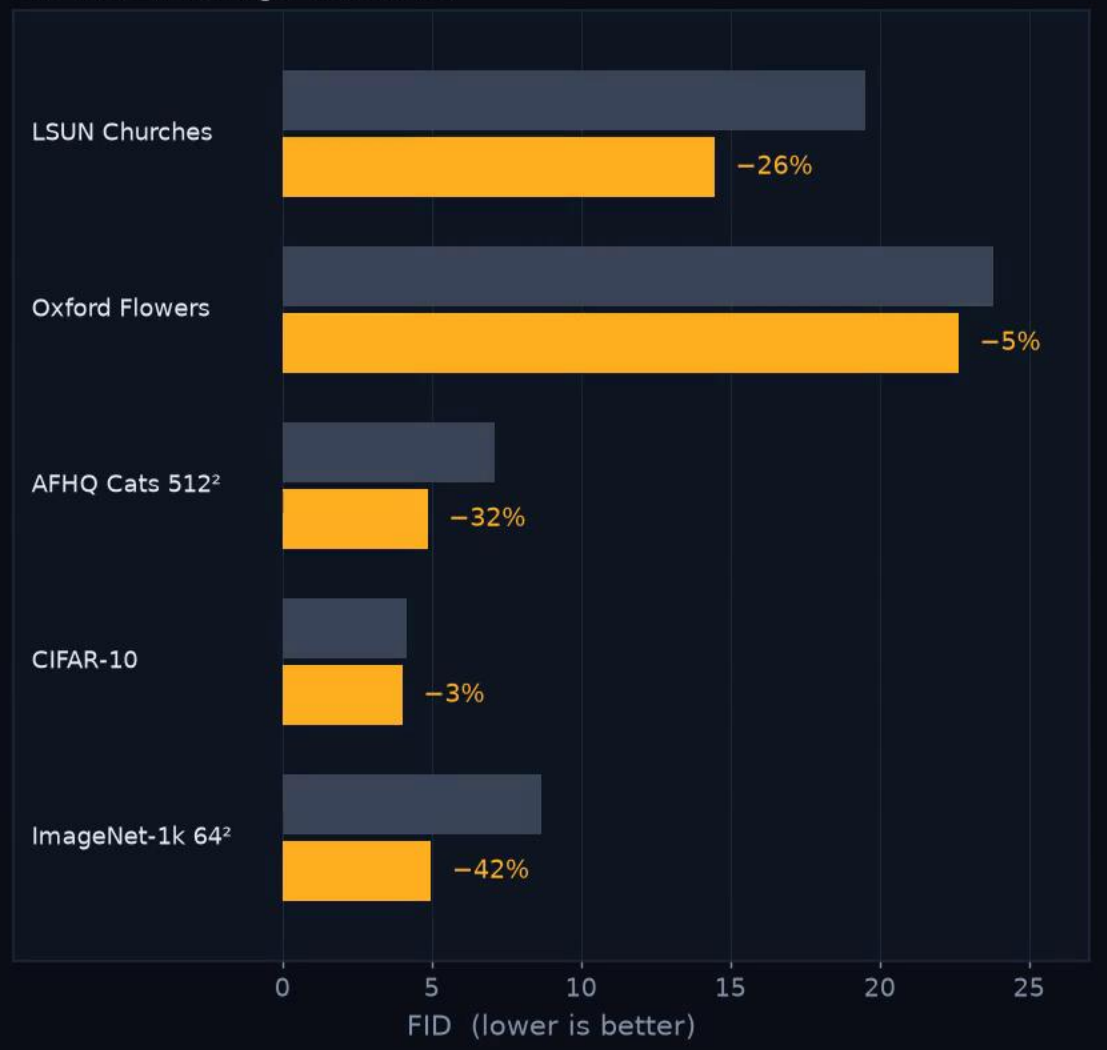
Same architecture, same 3k-step budget. Standard FM smears the modes.

Preconditioned FM lands on all six — 3.8× lower MMD.

MMD during training (GMM, 3 runs)



FID across image datasets



Better conditioning, better samples: consistent FID gains from 32×32 to 512×512.

Compute-matched baselines improve only marginally — the gain is geometry, not capacity.

The control: latent MNIST

One backbone, one sampler, one budget — only the preconditioner changes.

PRECONDITIONING METHOD	COMPUTE MATCHED	PARAMS	FID ↓	MMD ↓	PRECISION ↑	RECALL ↑
No preconditioning	No	2.32M	5.97	77.10	0.5774	0.2856
No preconditioning	Yes	3.37M	5.42	75.46	0.5869	0.2618
ZCA whitening (Ours)	No	2.32M	4.20	6.25	0.6022	0.3107
ZCA whitening (Ours)	Yes	3.37M	4.02	7.02	0.6117	0.2885
NF preconditioner (Ours)	**No**	**3.37M**	**2.74**	**2.95**	**0.7271**	**0.3561**
FM preconditioner (Ours)	**No**	**3.37M**	**3.98**	**3.02**	**0.8440**	**0.5486**

Adding capacity to the baseline moves FID 5.97 → 5.42. The preconditioner reaches **2.74** — and the cheapest one gives the best precision 0.844 and recall 0.549.

Every resolution, same budget

Latent space for all sets except CIFAR-10, which is modelled in pixels.

DATASET	FID ↓	PRECISION ↑	RECALL ↑
	no preconditioning → with preconditioning		
LSUN Churches (256×256)	19.53 → **14.47**	0.454 → **0.531**	0.324 → **0.326**
Oxford Flowers-102 (256×256)	23.82 → **22.64**	0.521 → **0.543**	0.276 → **0.285**
AFHQ Cats (512×512)	7.11 → **4.86**	0.776 → **0.832**	0.540 → **0.530**
CIFAR-10 (32×32)	4.15 → **4.02**	0.810 → **0.786**	0.588 → **0.589**
ImageNet-1k (64×64)	8.66 → **4.98**	0.780 → **0.817**	0.634 → **0.643**

FID falls on all five datasets — **–26%** churches, **–32%** cats, **–42%** ImageNet-1k. Flowers and CIFAR gain least: their paths were already well conditioned.

Same sampler, same step count

Top row in each block: standard flow matching. Bottom row: preconditioned.

LSUN CHURCHES · 256×256



AFHQ CATS · 512×512

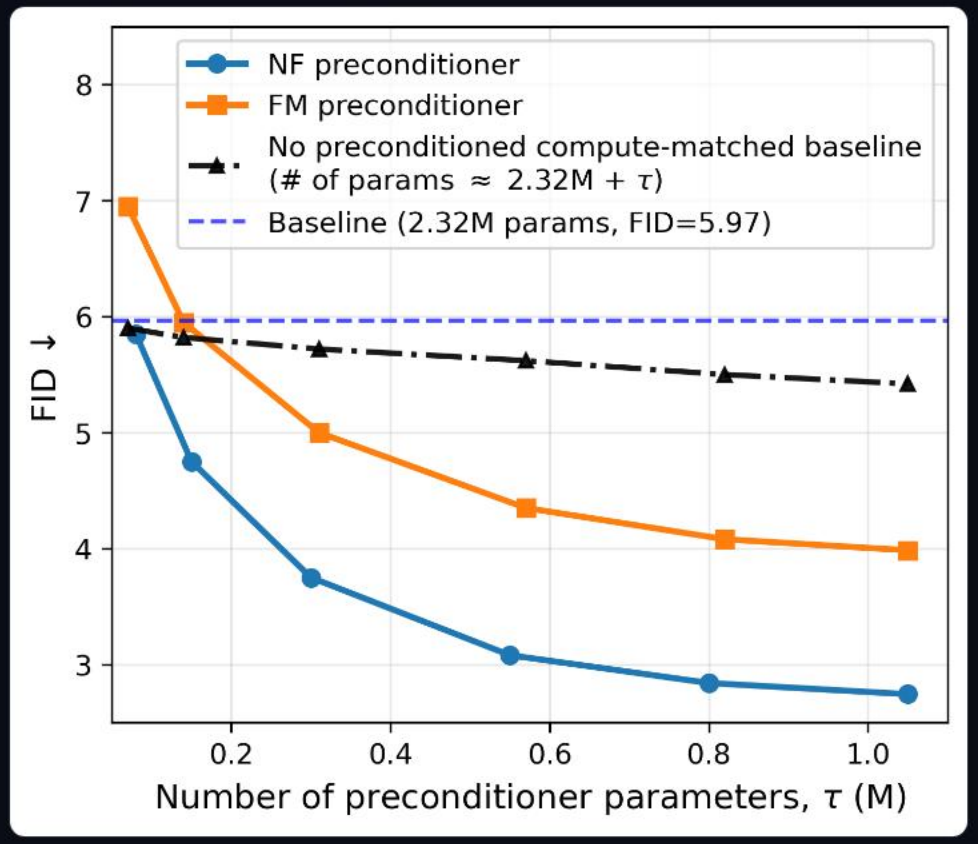


Window and roof structure survives in the churches; ear and whisker edges hold in the cats — the blur of an under-trained narrow direction is gone.

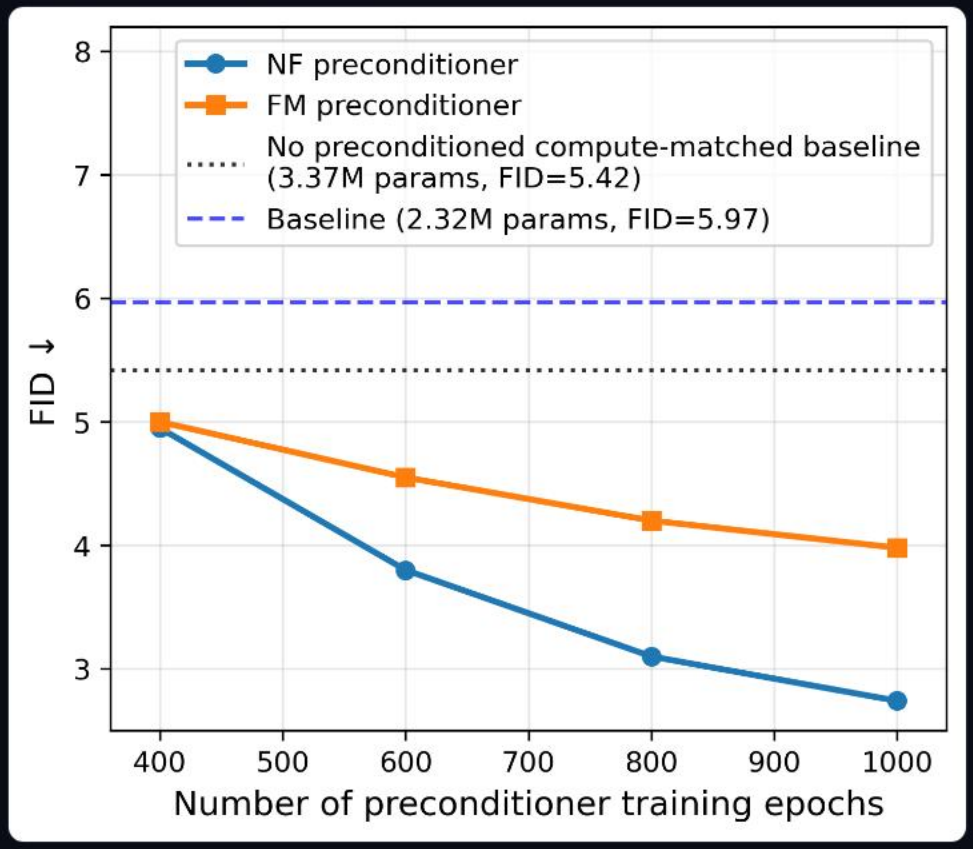
The gain is not extra capacity

Capacity-matched baseline in black; preconditioned runs in colour.

FID VS PRECONDITIONER SIZE



FID VS TRAINING EPOCHS



FID is flat past a few thousand preconditioner parameters, and the compute-matched baseline stays flat with training — so the gain is **geometry, not parameters**.

Preconditioned Flow Matching

Same architecture. Same budget. Better-conditioned geometry.

PRECONDITION	MATCH	MAP BACK
$x_1 \to \tilde{x}_1$	$\mathcal{N}(0, I) \to \tilde{x}_1$	$\hat{x}_1 = P^{-1}(\tilde{x}_1)$

PAPER arxiv.org/abs/2603.02337

CODE github.com/ahxmeds/preconditioned-flow-matching

PAGE ahxmeds.github.io/preconditioned-flow-matching

Shadab Ahamed · Eshed Gal · Md Shahriar Rahim Siddiqui · Simon Ghyselincks · Moshe Eliasof · Eldad Haber

