

Why Adversarial Diffusion Trains More Stably Than GANs: A Local Jacobian View

Florian Ochs, University of Oxford



The question

Diffusion models train far more reliably than GANs. Does that reliability come from the **scalar ELBO objective**, or from the **step-wise denoising structure** itself? We answer locally, by comparing the Jacobian of the training dynamics near equilibrium for one-step GAN, adversarial diffusion (DDGAN), and ELBO diffusion (DDPM). The short answer: min-max training rotates around the equilibrium, scalar objectives do not, and step-wise denoising averages the rotation down by $O(\log T/T)$.

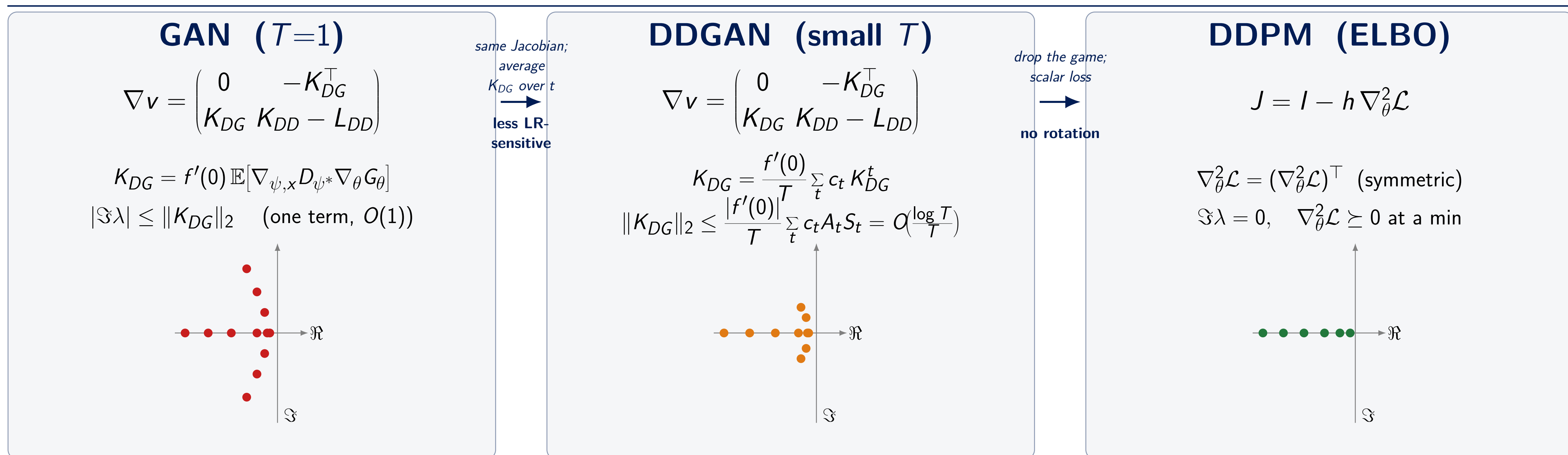
Setup: training as a dynamical system

All three objectives are trained by gradient updates. The adversarial games (GAN, DDGAN) use the simultaneous-gradient operator, resp. the scalar ELBO (DDPM) uses plain gradient descent:

$$F_h(\theta, \psi) = (\theta, \psi) + h \mathbf{v}(\theta, \psi), \quad \mathbf{v} = (-\nabla_\theta L, \nabla_\psi L) \quad \text{resp.} \quad F_h(\theta) = \theta - h \nabla_\theta \mathcal{L}.$$

Local convergence near an equilibrium is governed by the Jacobian of the update: $\nabla \mathbf{v}$ for the game, resp. $J = I - h \nabla_\theta^2 \mathcal{L}$ for DDPM. An eigenvalue of $\nabla \mathbf{v}$ is $a + ib$ (real part $\Re \lambda := a$, imaginary part $\Im \lambda := b$); with $\Re \lambda < 0$ the update eigenvalue $\lambda = 1 + h(a + ib)$ is contractive iff $0 < h < -2a/(a^2 + b^2)$. The imaginary part $\Im \lambda$ is *rotation* around the equilibrium: a larger $|\Im \lambda|$ shrinks the stable learning-rate window, so the same rotation that makes min-max training oscillate is what makes it learning-rate sensitive. The whole question reduces to *how much imaginary part each objective's Jacobian carries*, compared in the three panels below.

The mechanism: three objectives, one Jacobian



GAN and DDGAN share the *same* Jacobian; only the coupling block K_{DG} differs. Timestep averaging weights each per-step block $\|K_{DG}^t\|_2 \leq A_t S_t$ (discriminator/generator gradient norms) by $c_t \leq 1$, while the dissipative block stays $O(1)$, so the rotation-to-dissipation ratio shrinks, not undoable by rescaling h . DDPM replaces the game with a symmetric Jacobian.

Not a free lunch

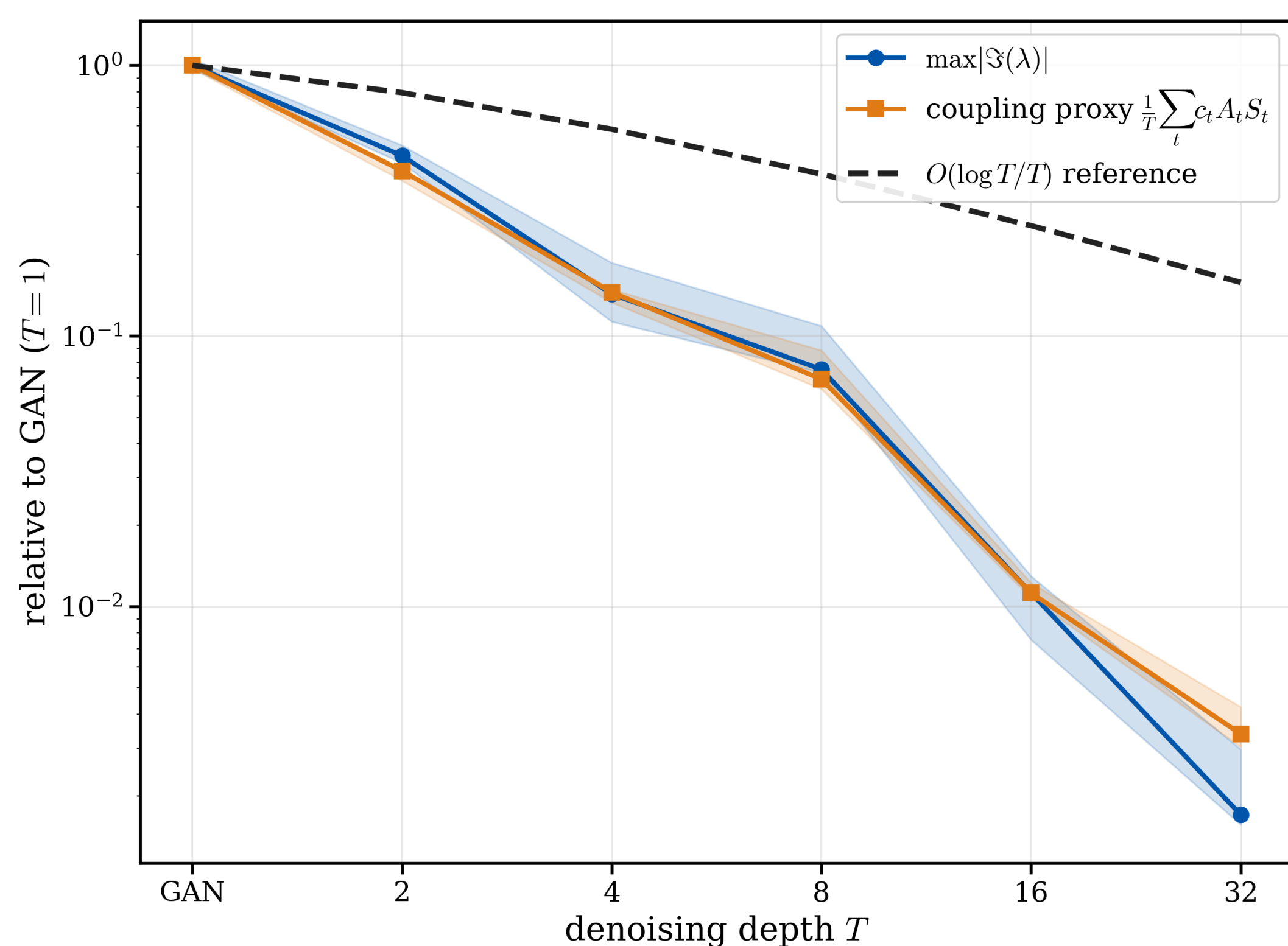
The top-left block of $\nabla \mathbf{v}$ is zero, so the generator converges *only* through the same coupling K_{DG} . Shrinking K_{DG} suppresses rotation but starves the generator, since $\|K_{DG}\|_2 \sim O(\log T/T)$ as T grows. So we do not expect a monotone win in T at fixed compute: adversarial diffusion targets small-to-moderate T (e.g. $T=8$), and as the coupling vanishes the ELBO parametrization becomes the natural and most stable choice.

Why DDGAN lands between GAN and DDPM

- GAN: full coupling, so large $|\Im \lambda|$ and a tight LR window.
- DDGAN: coupling $\times O(\log T/T)$, so smaller $|\Im \lambda|$ and a wider window than GAN.
- DDPM: symmetric Jacobian, so $\Im \lambda = 0$ unconditionally and the widest window.

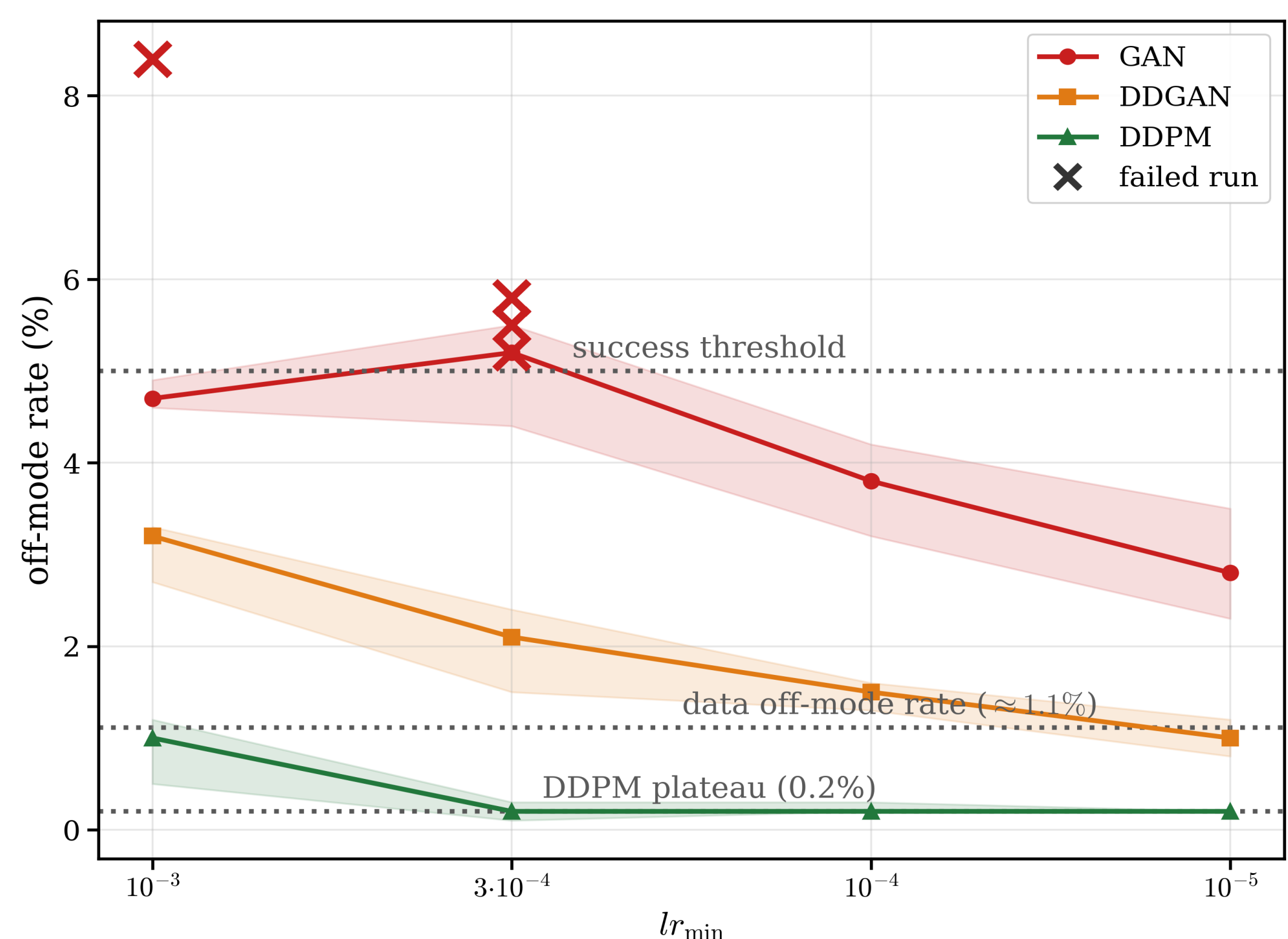
The stability ordering DDPM > DDGAN > GAN follows directly from how much imaginary part each objective's Jacobian carries.

Depth kills rotation



8-Gaussians: rotation and coupling proxy decay with T , staying below the $O(\log T/T)$ bound.

Learning-rate robustness



8-Gaussians LR sweep: DDPM robust throughout, GAN most sensitive, DDGAN between.

Takeaways

Rotation drives instability. Min-max Jacobians carry imaginary eigenvalues. This rotation near equilibria is what makes GAN training learning-rate sensitive.

Denoising suppresses it. Step-wise structure weights the coupling by c_t and averages over t , shrinking $|\Im \lambda|$ by $O(\log T/T)$, at the cost of a more weakly driven generator.

Scalar objectives are most stable. A scalar (ELBO) objective has a symmetric Jacobian and hence a real spectrum, removing rotational dynamics near optima.