

000 001 002 003 004 005 006 007 008 009 010 011 012 013 014 015 016 017 018 019 020 021 022 023 024 025 026 027 028 029 030 031 032 033 034 035 036 037 038 039 040 041 042 043 044 045 046 047 048 049 050 051 052 053 LEARNING WHAT MATTERS: STEERING DIFFUSION VIA SPECTRALLY ANISOTROPIC FORWARD NOISE

Anonymous authors

Paper under double-blind review

ABSTRACT

Diffusion Probabilistic Models (DPMs) have achieved strong generative performance, yet their inductive biases remain largely implicit. In this work, we aim to build inductive biases into the training and sampling of diffusion models to better accommodate the target distribution of the data to model. We introduce an anisotropic noise operator that shapes these biases by replacing the isotropic forward covariance with a structured, frequency-diagonal covariance. **This operator unifies band-pass masks and power-law weightings, allowing us to emphasize or suppress designated frequency bands, while keeping the forward process Gaussian.** We refer to this as Spectrally Anisotropic Gaussian Diffusion (SAGD). In this work, we derive the score relation for anisotropic forward covariances and show that, under full support, the learned score converges to the true data score as $t \rightarrow 0$, while anisotropy reshapes the probability-flow path from noise to data. Empirically, we show the induced anisotropy outperforms standard diffusion across several vision datasets, and enables *selective omission*: learning while ignoring known corruptions confined to specific bands. Together, these results demonstrate that carefully designed anisotropic forward noise provides a simple, yet principled, handle to tailor inductive bias in DPMs.

1 INTRODUCTION

Diffusion Probabilistic Models (DPMs) have emerged as powerful tools for approximating complex data distributions, finding applications across a variety of domains, from image synthesis to probabilistic modeling (Yang et al., 2024; Ho et al., 2020; Sohl-Dickstein et al., 2015; Venkatraman et al., 2024; Sendera et al., 2024). These models operate by gradually transforming data into noise through a defined diffusion process and training a denoising model (Vincent et al., 2008; Alain & Bengio, 2014) to learn to reverse this process, enabling the generation of samples from the desired distribution via appropriate scheduling. Despite their success, the inductive biases inherent in diffusion models remain largely unexplored, particularly in how these biases influence model performance and the types of distributions that can be effectively modeled.

Inductive biases are known to play a crucial role in deep learning models, guiding the learning process by favoring certain types of data representations over others (Geirhos et al., 2019; Bietti & Mairal, 2019; Tishby & Zaslavsky, 2015). A well-studied example is the Frequency Principle (F-principle) or spectral bias, which suggests that neural networks tend to learn low-frequency components of data before high-frequency ones (Xu et al., 2019; Rahaman et al., 2019). Another related phenomenon is what is also known as the simplicity bias, or shortcut learning (Geirhos et al., 2020; Scimeca et al., 2021; 2023b), in which models are observed to preferentially pick up on simple, easy-to-learn, and often spuriously correlated features in the data for prediction. If left implicit, it is often unclear whether these biases will improve or hurt the models' performance on downstream tasks, potentially leading to undesired outcomes (Scimeca et al., 2023a). In this work, we aim to explicitly tailor the inductive biases of DPMs to better learn the target distribution of interest.

Recent studies have begun to explore the inductive biases inherent in diffusion models. For instance, Kadkhodaie et al. (2023) analyzes how the inductive biases of deep neural networks trained for image denoising contribute to the generalization capabilities of diffusion models. They demonstrate that these biases lead to geometry-adaptive harmonic representations, which play a crucial role in the models' ability to generalize beyond the training data. Similarly, Zhang et al. (2024) investigates

054 the role of inductive and primacy biases in diffusion models, particularly in the context of reward
 055 optimization. They propose methods to mitigate overoptimization by aligning the models' inductive
 056 biases with desired outcomes. Other methods, such as noise schedule adaptations (Sahoo et al., 2024)
 057 and the introduction of non-Gaussian noise (Bansal et al., 2022) have shown promise in improving
 058 the performance of diffusion models on various tasks. However, the exploration of frequency domain
 059 techniques within diffusion models is a relatively new area of interest. One of the pioneering studies
 060 in this domain investigates the application of diffusion models to time series data, where frequency
 061 domain methods have shown potential for capturing temporal dependencies more effectively (Crabbé
 062 et al., 2024). Similarly, the integration of spatial frequency components into the denoising process has
 063 been explored for enhancing image generation tasks (Qian et al., 2024; Yuan et al., 2023), showcasing
 064 the importance of considering frequency-based techniques as a means of refining the inductive biases
 065 of diffusion models.

066 In this work, we explore a new avenue, to build inductive biases in DPMs by frequency-based noise
 067 control. **Specifically, we replace the isotropic forward noise with an anisotropic forward Gaussian**
 068 **operator whose covariance is structured in the Fourier basis.** The main hypothesis in this paper is
 069 that the noising operator in a diffusion model has a direct influence on the model's representation
 070 of the data. Intuitively, the information erased by the noising process is the very information that
 071 the denoising model has pressure to learn, so that reconstruction is possible. Accordingly, by
 072 shaping the forward covariance we can steer which modes carry supervision signal during training
 073 and thus which aspects of the data distribution the model learns most effectively. We focus our
 074 attention on the generative learning of topologically structured data, and implement anisotropy via a
 075 frequency-parameterized schedule that emphasizes or de-emphasizes selected bands while keeping
 076 the forward process Gaussian. In what follows we refer to this setting as Spectrally Anisotropic
 077 Gaussian Diffusion (SAGD).

078 We report several key findings showing that SAGD provides a simple, principled handle to tailor
 079 inductive biases, including: improved learning of information lying in particular frequency bands;
 080 increased performance across several natural image datasets; and learning while ignoring (corrupted)
 081 information at predetermined frequency bands. Because SAGD modifies only the forward covariance,
 082 it integrates with existing diffusion implementations with a few lines of code and preserves the rest of
 083 the pipeline intact. We summarize our contributions as follows:

- 084 1. We introduce SAGD, an **anisotropic forward-noise operator with frequency-diagonal covari-**
085 ance (in the Fourier basis) that provides a simple handle on spectral inductive bias.
- 086 2. We provide a theoretical analysis showing that, under full spectral support, the learned
 087 score at $t \rightarrow 0$ recovers the true data score, while anisotropy deterministically reshapes the
 088 probability-flow path.
- 089 3. We show that SAGD can steer models to better approximate information concentrated in
 090 selected bands of the underlying data distribution.
- 091 4. We test and empirically show that models trained with SAGD anisotropic **forward covariance**
 092 **can match or outperform traditional (isotropic) diffusion across multiple datasets.**
- 093 5. We demonstrate *selective omission*: by zeroing chosen bands in the forward covariance,
 094 models learn to ignore known corruptions and recover the clean distribution.

096 2 METHODS

097 2.1 DENOISING PROBABILISTIC MODELS (DPMs)

100 Denoising Probabilistic Models are a class of generative models that learn to reconstruct complex
 101 data distributions by reversing a gradual noising process. DPMs are characterized by a *forward* and
 102 *backward* process. The *forward process* defines how data is corrupted, typically by Gaussian noise,
 103 over time. Given a data point \mathbf{x}_0 sampled from the data distribution $q(\mathbf{x}_0)$, the noisy versions of the
 104 data $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T$ are generated according to:

$$q(\mathbf{x}_t \mid \mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{\alpha_t} \mathbf{x}_{t-1}, (1 - \alpha_t) \mathbf{I}), \quad (1)$$

105 with variance schedule α_t . The *reverse process* models the denoising operation, attempting to recover
 106 \mathbf{x}_{t-1} from \mathbf{x}_t :

$$p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \mu_\theta(\mathbf{x}_t, t), \sigma_t^2 \mathbf{I}), \quad (2)$$

108 where $\mu_\theta(\mathbf{x}_t, t)$ is predicted by a neural network f_θ , and the variance σ_t^2 can be fixed, learned, or
 109 precomputed based on a schedule. We train the denoising model with the standard ϵ -parameterization
 110 by minimizing

$$112 \quad \mathcal{L} = \mathbb{E}_{t, \mathbf{x}_0, \epsilon \sim \mathcal{N}(0, \mathbf{I})} \left[\left\| \epsilon - \epsilon_\theta(\mathbf{x}_t, t) \right\|^2 \right], \quad \mathbf{x}_t = \sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sigma_t \epsilon, \quad (3)$$

114 where $\bar{\alpha}_t = \prod_{s=1}^t \alpha_s$ and $\sigma_t^2 = 1 - \bar{\alpha}_t$, ϵ is the Gaussian noise added to \mathbf{x}_0 , and ϵ_θ is the model's
 115 prediction of this noise. To generate new samples, we start from noise and apply the learned reverse
 116 process iteratively.

118 2.2 SPECTRALLY ANISOTROPIC GAUSSIAN DIFFUSION (SAGD)

120 The objective of this section is to convolve the data with a spectrally anisotropic gaussian noise
 121 during training to steer a model's tendency to learn particular aspects of the data distribution. To
 122 do so, we wish to generate spatial Gaussian noise whose frequency content can be systematically
 123 manipulated according to an arbitrary weighting function **in the Fourier basis, where stationary**
 124 **Gaussian covariances diagonalize and our power-law and band-pass families become low-dimensional**
 125 **parameterizations of the covariance eigenvalues.** The right-hand side of Equation 3 denotes how \mathbf{x}_t is
 126 generated by adding Gaussian noise $\epsilon \sim \mathcal{N}(0, \mathbf{I})$ to \mathbf{x}_0 .

127 Let us denote by $\mathbf{x} \in \mathbb{R}^{H \times W}$ an image (or noise field) in the spatial domain, and by \mathcal{F} the two-
 128 dimensional Fourier transform operator. Given white spatial Gaussian noise $\epsilon \sim \mathcal{N}(0, \mathbf{I})$, we form its
 129 Fourier transform $\mathbf{N}_{\text{freq}} = \mathcal{F}(\epsilon)$, where $\mathbf{N}_{\text{freq}} \in \mathbb{C}^{H \times W}$ is a complex-valued random field whose real
 130 and imaginary parts are i.i.d. Gaussian, i.e.:

$$131 \quad \mathbf{N}_{\text{freq}} = \mathbf{N}_{\text{real}} + i \mathbf{N}_{\text{imag}}, \quad \mathbf{N}_{\text{real}}, \mathbf{N}_{\text{imag}} \sim \mathcal{N}(0, \mathbf{I}), \quad (4)$$

132 We introduce a *weighting function* $w(f_x, f_y)$ that scales the amplitude of each frequency component.
 133 Let $\mathbf{f} = (f_x, f_y)$ denote coordinates in frequency space, where $f_x = \frac{k_x}{W}$, $f_y = \frac{k_y}{H}$, and k_x, k_y are
 134 integer indices (ranging over the width and height), while H and W are the image dimensions. We
 135 define the frequency-controlled noise $\mathbf{N}_{\text{freq}}^{(w)}(\mathbf{f})$ as:

$$137 \quad \mathbf{N}_{\text{freq}}^{(w)}(\mathbf{f}) = \mathbf{N}_{\text{freq}}(\mathbf{f}) \odot w(\mathbf{f}), \quad (5)$$

139 After applying $w(\mathbf{f})$ in the frequency domain, we invert back to the spatial domain to obtain the noise
 140 $\epsilon^{(w)}$:

$$141 \quad \epsilon^{(w)} = \Re(\mathcal{F}^{-1}(\mathbf{N}_{\text{freq}}^{(w)})), \quad (6)$$

143 where $\Re(\cdot)$ ensures that our final noise field is purely real.¹

144 In practice, any standard spatial noise can be converted to $\epsilon^{(w)}$ via this unified framework:

$$146 \quad \epsilon \xrightarrow{\mathcal{F}} \mathbf{N}_{\text{freq}} \xrightarrow{w(\mathbf{f})} \mathbf{N}_{\text{freq}}^{(w)} \xrightarrow{\mathcal{F}^{-1}} \epsilon^{(w)}.$$

148 Note that standard white Gaussian noise is a special case of this formulation, where $w(\mathbf{f}) = 1$ for all
 149 \mathbf{f} . In contrast, more sophisticated weightings allow one to emphasize, de-emphasize, or even remove
 150 specific bands of the frequency domain.

152 **Theoretical consistency** We keep the forward process Gaussian while reshaping its spectrum. Let
 153 \mathcal{F} be the unitary DFT and $w(\mathbf{f}) \geq 0$ a fixed spectral weight. The linear map $\mathbf{T}_w = \mathcal{F}^{-1} \text{Diag}(w) \mathcal{F}$
 154 sends white noise to frequency-based noise $\epsilon^{(w)} = \mathbf{T}_w \epsilon$ with covariance

$$155 \quad \Sigma_w = \mathbf{T}_w \mathbf{T}_w^* = \mathcal{F}^{-1} \text{Diag}(|w(\mathbf{f})|^2) \mathcal{F}. \quad (7)$$

157 Replacing ϵ by $\epsilon^{(w)}$ in the forward step yields the marginal

$$159 \quad q_w(\mathbf{x}_t \mid \mathbf{x}_0) = \mathcal{N}(\sqrt{\bar{\alpha}_t} \mathbf{x}_0, \sigma_t^2 \Sigma_w). \quad (8)$$

161 ¹Since the DFT of a real signal has Hermitian symmetry, multiplying by a real, pointwise weight w preserves
 162 Hermitian symmetry and yields a real-valued inverse transform.

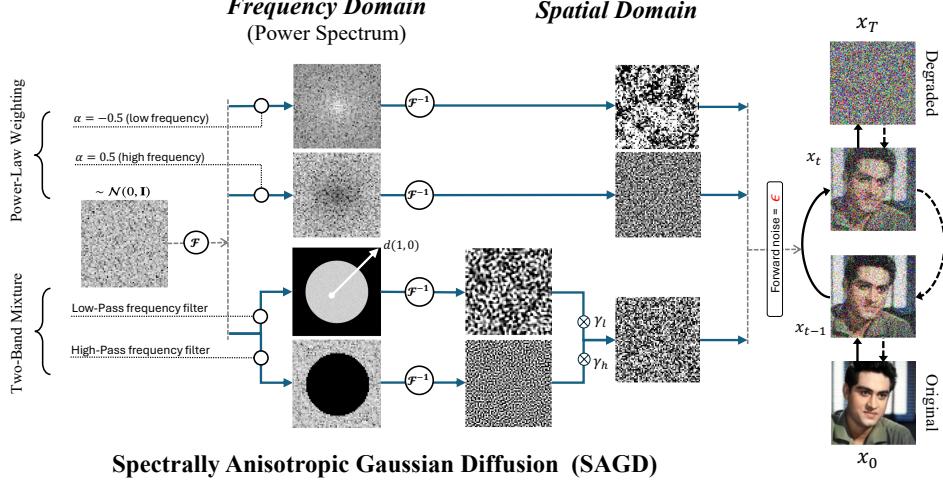


Figure 1: Spectrally Anisotropic Gaussian Diffusion under a generalized framework.

Training with the standard ℓ_2 objective on the added noise remains optimal: $\epsilon_\theta^\star(\mathbf{x}_t, t) = \mathbb{E}[\epsilon^{(w)} | \mathbf{x}_t]$, and the corresponding score satisfies

$$\nabla_{\mathbf{x}_t} \log q_{w,t}(\mathbf{x}_t) = -\frac{1}{\sigma_t} \Sigma_w^{-1} \epsilon_\theta^\star(\mathbf{x}_t, t), \quad (9)$$

so converting ϵ -predictions to scores simply multiplies by Σ_w^{-1} . As $t \rightarrow 0$, if $\Sigma_w \succ 0$ and q has a locally positive density with $\nabla \log q \in L^1_{\text{loc}}$, the anisotropic Gaussian smoothing collapses to a Dirac and $\nabla \log q_{w,t} \rightarrow \nabla \log q$ almost everywhere. Thus, shaping the forward spectrum preserves the endpoint score while altering the path to the score at the data distribution (see Appendix subsection D.1 for proofs and extensions).

2.3 FREQUENCY NOISE OPERATORS

In this work, the design of $w(\mathbf{f})$ is especially important. We propose two particular choices which provide a flexible design bench: power-law weighting and a two-band mixture.

POWER-LAW WEIGHTING (*plw*-SAGD)

We implement a radial power-law anisotropic noise operator that imposes a linear slope in the log–log power spectrum. Let $\mathbf{f} = (f_x, f_y)$ denote normalized frequency coordinates on $[-\frac{1}{2}, \frac{1}{2}]^2$, and define the radial frequency

$$r(\mathbf{f}) = \sqrt{f_x^2 + f_y^2}.$$

Given white spatial Gaussian noise $\epsilon \sim \mathcal{N}(0, \mathbf{I})$, we form its Fourier transform $\mathbf{N}_{\text{freq}} = \mathcal{F}(\epsilon)$ and scale each frequency bin by

$$w_\alpha(\mathbf{f}) = (r(\mathbf{f}) + \varepsilon)^\alpha, \quad \varepsilon = 10^{-10}, \quad (10)$$

where $\alpha \in \mathbb{R}$ controls the slope and ε is a small weight to prevent a DC singularity. The shaped spectrum and spatial noise are

$$\mathbf{N}_{\text{freq}}^{(\alpha)}(\mathbf{f}) = \mathbf{N}_{\text{freq}}(\mathbf{f}) \cdot w_\alpha(\mathbf{f}), \quad \epsilon^{(\alpha)} = \mathcal{R}(\mathcal{F}^{-1}[\mathbf{N}_{\text{freq}}^{(\alpha)}]), \quad (11)$$

which we use in the forward step $\mathbf{x}_t = \sqrt{\alpha_t} \mathbf{x}_{t-1} + \sqrt{1 - \alpha_t} \epsilon^{(\alpha)}$. A minimal code implementation of *plw*-SAGD can be found in Appendix E.

Effect on spectrum and learning signal. Let $w_\alpha(r) = (r + \varepsilon)^\alpha$ be the radial spectral weight. Because amplitudes are scaled by w_α while power scales with $|w_\alpha|^2$, the radially averaged power spectral density (RAPSD) follows: $\log \text{PSD}(r) \approx (2\alpha) \log r + \text{const}$, so $\alpha > 0$ tilts energy toward

216 high frequencies (sharper textures), $\alpha < 0$ toward low frequencies (coarser structure), and $\alpha = 0$
 217 recovers white noise. Note also that the global scalar rescaling of Σ_w can be absorbed into the scalar
 218 variance σ_t^2 in the forward process (or the noise schedule). What we aim to do, instead, is the relative
 219 weighting of the modes (eigenvalues) of Σ_w in the Fourier basis, (shape of $w(\mathbf{f})$).
 220

221 BAND-PASS MASKING AND TWO-BAND MIXTURE (*bpm*-SAGD))

222 A band-pass mask can be viewed as a special case of a more general weighting function:
 223

$$224 \quad w(\mathbf{f}) \in \{0, 1\}. \quad (12)$$

225 In this case, the frequency domain is split into a set of permitted and excluded regions, or radial
 226 thresholds. With this, we can construct several types of filters, including a low-pass filter retaining
 227 only frequencies below a cutoff (e.g., $\|\mathbf{f}\| \leq \omega_c$), a high-pass filter keeping only frequencies above a
 228 cutoff, or more generally a filter restricting $\|\mathbf{f}\|$ to lie between two thresholds $[a, b]$. We thus define a
 229 simple band-pass filter as:
 230

$$231 \quad w_{a,b}(\mathbf{f}) = \mathbf{M}_{[a,b]}(f_x, f_y) = \begin{cases} 1, & \text{if } a \leq d(f_x, f_y) \leq b, \\ 0, & \text{otherwise,} \end{cases} \quad (13)$$

233 In this special case, $w(\mathbf{f})$ is simply a binary mask, selecting only those frequencies within $[a, b]$.
 234

235 For the experiments in this paper we formulate a simple two-band mixture, where we limit ourselves
 236 to constructing noise as a linear combination of two band-pass filtered components. Specifically, we
 237 generate frequency-filtered noise ϵ_f via:
 238

$$239 \quad \epsilon^{(w)} = \gamma_l \epsilon_{[a_l, b_l]} + \gamma_h \epsilon_{[a_h, b_h]}, \quad (14)$$

240 where $\gamma_l, \gamma_h \geq 0$ denote the relative contributions of a low and a high-frequency component ($\gamma_l + \gamma_h = 1$),
 241 each filtering noise respectively in the ranges $[a_l, b_l]$ (low-frequency range) and $[a_h, b_h]$ (high-
 242 frequency range). We uniquely refer to $\epsilon_{[a,b]}$ as the noise filtered in the $[a, b]$ frequency range
 243 following Equation 5 and Equation 6. Standard Gaussian noise emerges as a particular instance of
 244 this formulation, for $\gamma_l = 0.5$, $\gamma_h = 0.5$, $a_l = 0$, $b_l = a_h = 0.5$, and $b_h = 1$. We provide a minimal
 245 code implementation of *bpm*-SAGD in Appendix E.
 246

247 **Selective omission:** If w vanishes on a band, then Σ_w is rank-deficient and the model learns
 248 the score *projected* onto $\text{range}(\Sigma_w)$. In the two-band mixture operator, we can achieve this for
 249 $b_l < a_h$, leaving the $[b_l, a_h]$ frequency band unsupported by the anisotropic covariance. Note that
 250 the $t \rightarrow 0$ score-consistency result in our analysis requires full spectral support ($\Sigma_w \succ 0$); when Σ_w is
 251 singular, the smoothed marginals are not absolutely continuous and the estimator converges only to
 252 the *projected* score, i.e., $\Pi \nabla_{\mathbf{x}} \log q(\mathbf{x})$ with Π the orthogonal projector onto $\text{range}(\Sigma_w)$ (equivalently,
 253 replace Σ_w^{-1} by the Moore–Penrose pseudoinverse in the score– ϵ mapping). As later shown, we
 254 exploit this to avoid learning components in the omitted band, while learning and recovering the
 255 information in the bands of interest.
 256

257 3 RESULTS

258 3.1 EXPERIMENTAL DETAILS

259 All experiments involve separately training and testing DPMs with various SAGD schedules, along-
 260 side standard isotropic Gaussian baselines. We consider six image datasets—MNIST, CIFAR-10,
 261 DomainNet-Quickdraw, WikiArt, FFHQ, and ImageNet-1k—spanning widely different visual distri-
 262 butions, scales, and statistics. We study both pixel-space diffusion with U-Net denoisers and latent
 263 diffusion with DiT backbones in a DINOv2 feature space; the latter uses the public RAE implemen-
 264 tation of state-of-the-art DiT models on ImageNet-1k at 256×256 resolution (Zheng et al., 2025). We
 265 use DDIM sampling (Song et al., 2021) in all experiments, so no step noise is injected at test time. As
 266 quality metrics, we report FID and KID between generated samples and held-out data, computed from
 267 Inception-v3 features: for all datasets except ImageNet-1k we use a 768-dimensional intermediate
 268 block, while for ImageNet-1k we follow standard practice and use the 2048-dimensional penulti-
 269 mate block. Unless otherwise stated, we report averages over multiple random seeds. Additional
 270 experimental details are provided in Appendix C.
 271

270 3.2 IMPROVED SAMPLING VIA SAGD
271

272 In the first set of experiments, we wish to test our main hypothesis, i.e. that appropriate spectral
273 manipulations of the forward noise can better support the learning of particular aspects of the sampling
274 distribution. In the following experiments, we use the *plw*-SAGD formulation in Equation 10 and
275 Equation 11 to train and compare diffusion models with different anisotropic power-law weighted
276 operators, while varying the value of α to emphasize or de-emphasize the learning of higher or lower
277 frequency components of the target distribution, where $\alpha \in [-0.1, 0.1]$.

278 3.2.1 QUALITATIVE OVERVIEW
279

280 First, we show a qualitative example of a
281 standard forward linear noising schedule in
282 DPMs, as compared to two particular settings of *plw*-SAGD, emphasizing high and low-
283 frequency noising in Figure 2. With standard
284 noise, information is uniformly removed from
285 the image, with sample quality degrading evenly
286 over time. In the high-frequency noising sched-
287 ule ($\alpha > 0$), sharpness and texture are affected
288 more significantly than general contours and
289 shapes, which remain intact over longer trajec-
290 tories; in the low-frequency noising schedule
291 ($\alpha < 0$), instead, general shapes and homo-
292 geneous pixel clusters are quickly affected, yield-
293 ing qualitatively different information destruc-
294 tion operations over the sampling time steps. As
295 discussed previously, we hypothesize that this
296 will, in turn, purposely affect the information
297 learned by the denoiser, effectively focusing the
298 diffusion sampling process on different parts of
299 the distribution to learn.

300 3.2.2 LEARNING TARGET DISTRIBUTIONS
301 FROM FREQUENCY-BOUNDED INFORMATION
302

303 We conduct a controlled experiment to test
304 whether SAGD yields better samplers in the case
305 where the information content in the data lies, by
306 construction, in the high frequencies. We use the
307 CIFAR-10 dataset, and corrupt the original data
308 with low-frequency noise $\epsilon_{[0,3]}$, thus erasing the
309 low-frequency content while predominantly pre-
310 serving the high-frequency details in the range
311 $\epsilon_{[3,1]}$. We separately train DPMs with twelve
312 different α values, as well as a standard DPM at
313 $\alpha = 0$ (baseline). We repeat the experiment over
314 three seeds and report the average FID and error
315 in Figure 3. In the figure, we observe an almost
316 monotonically decreasing relationship between
317 the mean FID and increasing values of α , with a
318 significant 0.3 decrease in average FID score for
319 $\alpha = 0.08$ from the standard DPM baseline. The
320 observation is in line with our original intuition,
321 whereby improved learning can be achieved by
322 aligning the forward noising operation with the
323 data's dominant spectral content (the frequency bands
324 carrying most of the information).

325 3.2.3 SAGD IN NATURAL DATASETS

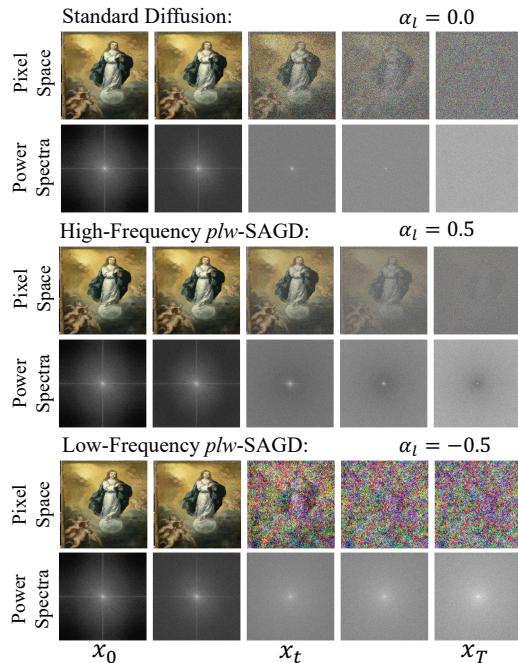


Figure 2: Power spectra and image visuals of the forward Process in standard diffusion, as compared to high ($\alpha = 0.5$) and low-frequency ($\alpha = -0.5$) noise settings of a power-law weighted SAGD.

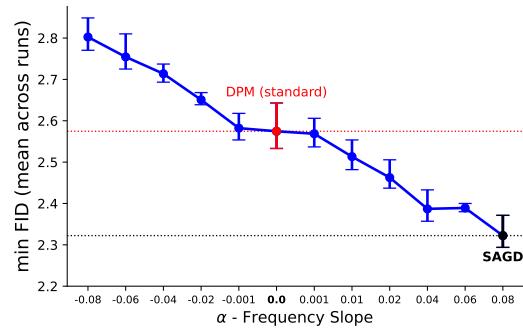


Figure 3: Mean FID across seeds of *plw*-SAGD diffusion samplers trained on CIFAR-10, pre-
325 processed to only retain high frequency information ($\alpha = 0$ yields standard diffusion).

Table 1: **FID** across selected α (frequency slope) settings for all datasets (mean \pm standard error across seeds). FID is computed from block 768 for all datasets except ImageNet1k (block 2048).

Dataset →	MNIST	CIFAR-10	Domainnet-Quickdraw	Wiki-Art	FFHQ	ImageNet1k
baseline	$0.42 \pm 8.52e - 03$	0.75 ± 0.01	0.60 ± 0.05	1.06 ± 0.03	1.11 ± 0.01	8.6819 ± 0.0739
$\alpha = -0.060$	0.28 ± 0.02	0.94 ± 0.02	0.52 ± 0.05	1.35 ± 0.08	1.74 ± 0.10	8.1098 ± 0.0229
$\alpha = -0.040$	$0.31 \pm 7.76e - 03$	0.86 ± 0.02	0.49 ± 0.02	1.25 ± 0.07	1.76 ± 0.08	7.5534 ± 0.0556
$\alpha = -0.020$	$0.37 \pm 6.36e - 03$	0.76 ± 0.01	0.52 ± 0.03	1.14 ± 0.05	1.68 ± 0.19	7.6419 ± 0.0581
$\alpha = -0.010$	0.37 ± 0.02	0.75 ± 0.01	0.54 ± 0.04	1.09 ± 0.04	1.48 ± 0.12	8.0400 ± 0.0236
$\alpha = -0.001$	0.40 ± 0.02	0.76 ± 0.01	0.56 ± 0.04	$1.02 \pm 5.66e - 03$	$1.04 \pm 5.17e - 03$	8.5288 ± 0.0112
$\alpha = 0.010$	0.43 ± 0.02	0.80 ± 0.02	0.66 ± 0.02	1.20 ± 0.07	2.06 ± 0.06	9.3867 ± 0.0348

We further test our hypothesis by training twelve SAGD noising models with $\alpha \in [-0.08, 0.08]$ against a standard DPM baseline for each of the datasets considered in our experiments, varying greatly in size, resolution, visual distribution, and complexity, namely: MNIST, CIFAR-10, Domainnet-Quickdraw, Wiki-Art, FFHQ and ImageNet-1k. We report additional information on the datasets and our pre-processing pipeline in Appendix B and Appendix C. We repeat the experiments over 3 seeds and present a focused report of the mean FID and KID metrics for all ablations in Table 1 (see additional results in Appendix F). In Table 1, we observe significant SAGD improvements over standard DPM training (first row) in almost all datasets, where in 5/6 we achieve substantially improved baseline FID scores.

On ImageNet-1k at 256×256 resolution, using the RAE DiT backbone in DINOv2 latent space, the isotropic baseline ($\alpha=0$) attains an FID of 8.68 ± 0.07 , whereas SAGD with a low-frequency tilt ($\alpha=-0.04$) reaches 7.55 ± 0.06 , i.e., an absolute improvement of ≈ 1.1 FID (about 13% relative). As summarized in Table 1 and visualized in Figure 4, FID decreases almost monotonically as α moves from 0 toward moderately negative values (down to around -0.04) and then worsens again for more negative (-0.06) and positive values of α , indicating a non-trivial optimum away from the standard Gaussian setting. This demonstrates that SAGD yields measurable gains even on large-scale, high-resolution, natural-image generation in a state-of-the-art latent DiT setup, and that its benefits are not limited to small or low-resolution datasets.

3.3 SELECTIVE LEARNING: NOISE CONTROL TO OMIT TARGETED INFORMATION

Following our original intuition, learning pressure in DPMs is aligned with the information deletion induced by forward noising. Conversely, when the noising operator is crafted to leave parts of the original distribution intact, no such pressure exists, and the sampler can effectively discard the left-out information during generation.

In this section, we perform experiments whereby the original data is corrupted with noise at different frequency ranges. The objective is to manipulate the inductive biases of diffusion samplers to avoid learning the corruption noise, while correctly approximating the relevant information in the data. We formulate our corruption process as $\mathbf{x}' = A_c(\mathbf{x})$, where $A_c(\mathbf{x}) = \mathbf{x} + \gamma_c \epsilon_{f[a_c, b_c]}$ and $\epsilon_{f[a_c, b_c]}$ denotes noise in the $[a_c, b_c]$ frequency range. We perform the corruption on the fly, and use the original MNIST dataset for training while testing on 10K images. We default $\gamma_c = 1$. and show samples of the original and corrupted distributions in Figure 5. For any standard DPM training procedure, as expected, the sampler learns the corrupted distribution presented at training time. As such, the recovery of the original, noiseless, distribution would normally be impossible. Assuming knowledge of the corruption process, we use our two-band mixture formulation in Equation 14 and frame the frequency diffusion learning procedures as a noiseless distribution recovery process, where $a_l = 0$, $b_h = 1$, $b_l = a_c$, and $a_h = b_c$. This formulation effectively allows for the forward frequency-noising operator to omit the range of frequencies in which the corruption lies. In line with our previous

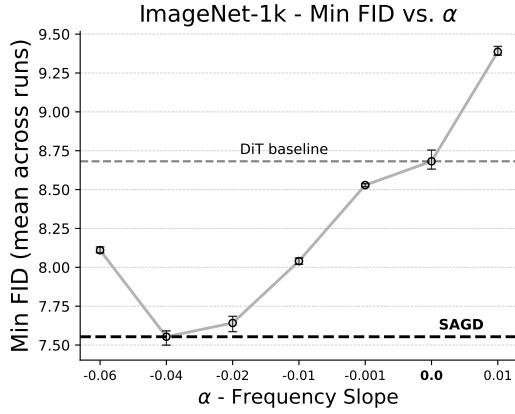


Figure 4: Mean FID across seeds of *plw*-SAGD diffusion samplers trained on ImageNet1k ($\alpha = 0$ yields standard diffusion).

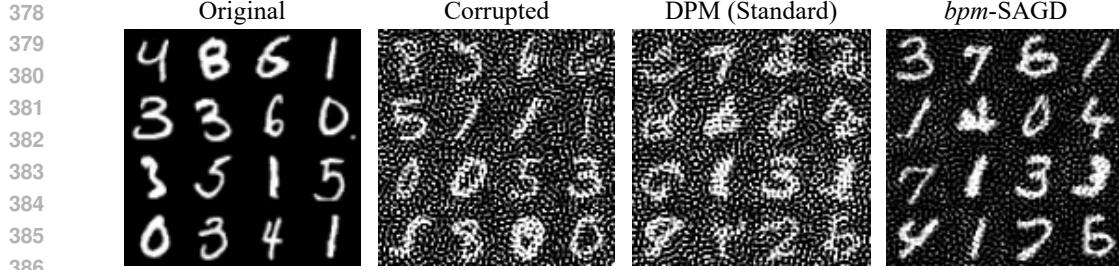


Figure 5: Samples from the original data distribution, the degraded data distribution, a standard diffusion sampler trained on the degraded data distribution, and a *frequency diffusion* sampler trained on the degraded data distribution. We generate noise for data corruption in the frequency range [$a_c = 0.4$, $b_c = 0.5$]).

rationale, this would effectively put no pressure on the denoiser to learn the corrupted part of the target distribution, and focus instead on the frequency ranges holding the true signal.

We compare original and corrupted samples from MNIST, as well as samples from standard and *bpm*-SAGD-trained models in Figure 5. In line with our hypothesis, we observe frequency diffusion DPMs trained with an appropriate two-band SAGD to be able to discard the corrupting information and recover the original distribution after severe corruption. We further measure the FID and KID of the samples generated by the baseline and frequency DPMs against the original (uncorrupted) data samples in Table 2. We perform 8 ablation studies, considering noises at 0.1 non-overlapping intervals in the [0.1, .9] frequency range. We observe appropriately designed *bpm*-SAGD samplers to outperform standard diffusion training across all tested ranges. Interestingly, we observe better performance (lower FID) for data corruption in the high-frequency ranges, and reduced performance for data corruption in low-frequency ranges, confirming a marginally higher information content in the low frequencies for the MNIST dataset.

Table 2: Resulting FID and KID between standard diffusion and *bpm*-SAGD samplers trained on noise-corrupted data, with respect to samples from the true uncorrupted distribution (mean \pm standard error across 3 seeds). We report eight ablation experiments across different non-overlapping corruption noise schemes.

Dataset \rightarrow	Baseline		<i>bpm</i> -SAGD	
	Corruption \downarrow	FID (\downarrow)	KID (\downarrow)	FID (\downarrow)
$\epsilon_{[0.1, 0.2]}$	3.2273 \pm 8.50e -03	0.0114 \pm 3.13e -05	2.7572 \pm 3.56e -02	0.0095 \pm 1.47e -04
$\epsilon_{[0.2, 0.3]}$	3.6601 \pm 4.43e -03	0.0132 \pm 1.67e -05	3.0416 \pm 4.47e -02	0.0107 \pm 1.79e -04
$\epsilon_{[0.3, 0.4]}$	3.4771 \pm 4.79e -03	0.0125 \pm 1.89e -05	2.9952 \pm 3.35e -02	0.0106 \pm 1.23e -04
$\epsilon_{[0.4, 0.5]}$	3.4281 \pm 5.46e -03	0.0123 \pm 1.98e -05	2.9218 \pm 2.54e -02	0.0105 \pm 8.79e -05
$\epsilon_{[0.5, 0.6]}$	3.3638 \pm 6.31e -03	0.0121 \pm 2.32e -05	2.8267 \pm 2.81e -02	0.0102 \pm 9.32e -05
$\epsilon_{[0.6, 0.7]}$	3.2444 \pm 7.10e -03	0.0116 \pm 2.55e -05	2.7026 \pm 3.90e -02	0.0097 \pm 1.28e -04
$\epsilon_{[0.7, 0.8]}$	3.0442 \pm 6.32e -03	0.0109 \pm 2.29e -05	2.5469 \pm 6.39e -02	0.0091 \pm 2.00e -04
$\epsilon_{[0.8, 0.9]}$	3.4660 \pm 7.90e -03	0.0124 \pm 2.96e -05	2.5138 \pm 9.63e -02	0.0090 \pm 3.07e -04

4 RELATED WORK

Beyond schedule tuning and architectural changes, several works have explicitly altered the forward corruption to shape inductive biases. Cold Diffusion replaces Gaussian noise with deterministic degradations (e.g., blur, masking), learning to invert arbitrary transforms without relying on stochastic noise (Bansal et al., 2023). Others introduce colored/ correlated Gaussian noise: for example, Huang et al. (2024) construct spatially correlated (blue/red) noise to emphasize selected frequency bands and show improved fidelity in image synthesis. **More recently, frequency-domain guidance has been used to shape what diffusion models learn during sampling: FDG-Diff (Zhang et al., 2025) modulates feature spectra via a frequency-domain guidance module, and Frequency-Guided Diffusion (Gao et al., 2025) adjusts high-frequency components during text-driven image translation. Both operate in the frequency domain but retain an isotropic Gaussian forward process.** Diffusion-like models based on reversing the heat equation similarly bias learning toward coarse-to-fine structure (Rissanen et al., 2023).

In contrast to these methods, SAGD retains a Gaussian forward model with frequency-diagonal covariance, providing a probabilistically consistent framework that is compatible with standard samplers (e.g., DDIM) and supported by a proof that, under full spectral support, the learned score converges to the true data score as $t \rightarrow 0$ (Sec. D). Closest to our setting, Voleti et al. (2022) formulate

432 diffusion with a non-isotropic Gaussian forward covariance Σ , and derive the corresponding denoising
 433 relations; their experiments, however, do not instantiate a spectrally structured operator and only show
 434 preliminary results on CIFAR-10. In contrast, we construct two noising operators while constraining
 435 Σ to be diagonal in the Fourier basis (i.e., $\Sigma = \mathcal{F}^{-1}\text{Diag}(|w(\mathbf{f})|^2)\mathcal{F}$), which (i) yields a simple,
 436 drop-in implementation via *per-frequency weighting* \rightarrow IFFT, (ii) leaves the standard ℓ_2 ϵ -objective
 437 unchanged while enabling a closed-form score- ϵ conversion by multiplying with Σ^{-1} (trivial in the
 438 spectral domain), and (iii) admits principled selective omission by zeroing prescribed bands. To our
 439 knowledge, this is the first forward-covariance manipulation method implemented with a frequency-
 440 diagonal Gaussian and validated through both spectral ablations and multi-dataset studies—while
 441 recovering standard isotropic Gaussian diffusion as a special case.

443 5 DISCUSSION AND CONCLUSION

444
 445 **Summary.** In this work, we studied the potential to build inductive biases in the training and
 446 sampling of Diffusion Probabilistic Models (DPMs) by purposeful manipulation of the forward
 447 covariance in the noising process. We introduced spectrally anisotropic Gaussian diffusion (SAGD),
 448 an approach that guides DPMs via an anisotropic Gaussian forward operator with frequency-diagonal
 449 covariance. We compare SAGD to DPMs trained with standard Gaussian noise on generative
 450 visual tasks spanning datasets with significantly varying structures and scales. We show several
 451 key findings. First, we show that under full spectral support, the learned score converges to the
 452 true data score as $t \rightarrow 0$, while anisotropy deterministically reshapes the probability-flow path from
 453 noise to data. Second, we show how shaping the forward covariance serves as a strong inductive
 454 bias that steers diffusion samplers to better learn information at particular frequencies. **Third, this**
 455 **property can be leveraged on both standard natural-image benchmarks (e.g., FFHQ, ImageNet-1k)**
 456 **and less conventional datasets (e.g., MNIST, DomainNet-Quickdraw), often yielding comparable or**
 457 **superior sampling quality to standard diffusion schedules, while remaining a minimal drop-in change**
 458 **to existing diffusion pipelines.** Finally, SAGD enables selective omission, whereby by zeroing chosen
 459 bands the model ignores unwanted content and recovers clean signals in desired ranges.

460 **Future Work.** In our approach, we crafted two particular choices of SAGD forward-noise operators:
 461 a power-law weighting and a two-band mixture. In future work, several other alternatives may be
 462 considered, which can serve as more flexible tools to inject useful inductive biases for similar tasks.
 463 Moreover, the approach can be extended beyond constant schedules. For instance, time-varying
 464 spectral strategies (Σ_{w_t}) could shift focus from low-frequency (general shapes) to high-frequency
 465 (edges/textures) components over the sampling trajectory. Such methods could more closely align
 466 with human visual processing, which progressively sharpens details over time, offering a more natural
 467 sampling process. Additionally, other domains of noise manipulation, outside of spectral operations,
 468 may also present new opportunities for further steerability and improvements.

469 **Limitations and Considerations.** A current limitation lies in the complexity of relating spatial
 470 percepts to their frequency-domain representations. The perception of information in the frequency
 471 domain does not always translate straightforwardly to visual content, impeding the process of
 472 designing optimal ad hoc operators. In practice, empirical validation is still required to identify
 473 the best inductive biases for a given dataset. **We believe it worthwhile for future work to develop**
 474 **analytical tools to guide operator design using data-specific considerations (e.g., spectral diagnostics**
 475 **and band-consistent metrics).**

476 Finally, while our experiments focus on images, the approach applies to any domain with an intrinsic
 477 spectral basis in which the forward covariance can be specified and (approximately) diagonalized
 478 (e.g., 1D time series, 2D/3D grids and videos via Fourier/DCT, geospatial fields, and graph/mesh
 479 signals via Laplacian eigenbases). In contrast, domains lacking a coherent topology or spectral
 480 geometry (e.g., unordered sets or purely categorical tabular data) offer no natural basis for anisotropic
 481 forward covariances, making our construction less applicable.

482 **Conclusion** Overall, this work opens the door for more targeted and flexible diffusion generative
 483 modeling by building inductive biases through the manipulation of the forward nosing process.
 484 The ability to design noise schedules that align with specific data characteristics holds promise for
 485 advancing the state of the art in generative modeling.

486 **Reproducibility Statement.** We take several steps to facilitate the reproducibility of our results.
 487 The SAGD formulation and training objective are specified in Secs. 2.1–2.2 (forward step in Eq. equation
 488 3, spectral operator in Eqs. equation 5–equation 6), with the score– ϵ relation, posterior, and
 489 probability–flow ODE derived in Appendix D. We provide minimal code to generate noise according
 490 to our two proposed SAGD operators (power-law and band-pass) in Appendix E. Datasets and prepro-
 491 cessing details appear in Appendices B and C; controlled corruptions and synthetic data procedures
 492 (power-law random fields) are documented alongside the code. All ablation settings (e.g., α grids,
 493 two-band ranges and weights, and seed counts) are enumerated in the Results tables/figures (e.g., ??,
 494 Figure F.1). Finally, all evaluations follow standard FID/KID protocols using Inception-v3 (block
 495 768); sample counts and metrics are reported per experiment.

496 **REFERENCES**

497 Guillaume Alain and Yoshua Bengio. What regularized auto-encoders learn from the data-generating
 498 distribution. *The Journal of Machine Learning Research*, 15(1):3563–3593, 2014.

500 Arpit Bansal, Eitan Borgnia, Hong-Min Chu, Jie S. Li, Hamid Kazemi, Furong Huang, Micah
 501 Goldblum, Jonas Geiping, and Tom Goldstein. Cold Diffusion: Inverting Arbitrary Image
 502 Transforms Without Noise, August 2022. URL <http://arxiv.org/abs/2208.09392>.
 503 arXiv:2208.09392 [cs].

504 Arpit Bansal, Xuhui Lin, Sergey Tulyakov, Hao Liu, Yijun Wang, Hsin-Ying Wang, Yong Jae
 505 Lee, and Abhinav Shrivastava. Cold diffusion: Inverting arbitrary image transforms without
 506 supervision. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023. URL
 507 <https://arxiv.org/abs/2208.09392>.

508 Alberto Bietti and Julien Mairal. On the inductive bias of neural tangent kernels. In *Advances in
 509 Neural Information Processing Systems*, 2019.

510 Jonathan Crabbé, Nicolas Huynh, Jan Stanczuk, and Mihaela van der Schaar. Time series diffusion in
 511 the frequency domain, 2024. URL <https://arxiv.org/abs/2402.05933>.

512 Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale
 513 hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*,
 514 pp. 248–255. Ieee, 2009.

515 Zheng Gao, Jifei Song, Zhensong Zhang, Jiankang Deng, and Ioannis Patras. Frequency-guided diffu-
 516 sion for training-free text-driven image translation. In *Proceedings of the IEEE/CVF International
 517 Conference on Computer Vision*, pp. 19195–19205, 2025.

518 Robert Geirhos, Patricia Rubisch, Claudio Michaelis, Matthias Bethge, Felix A Wichmann, and
 519 Wieland Brendel. Imagenet-trained cnns are biased towards texture; increasing shape bias improves
 520 accuracy and robustness. In *International Conference on Learning Representations*, 2019.

521 Robert Geirhos, Jörn-Henrik Jacobsen, Claudio Michaelis, Richard Zemel, Wieland Brendel, Matthias
 522 Bethge, and Felix A. Wichmann. Shortcut learning in deep neural networks. *Nature Machine
 523 Intelligence*, 2(11):665–673, 2020. doi: 10.1038/s42256-020-00257-z. URL <https://doi.org/10.1038/s42256-020-00257-z>.

524 Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising Diffusion Probabilistic Models, June 2020.
 525 URL <https://arxiv.org/abs/2006.11239v2>.

526 Xingchang Huang, Yuhao Zhou, Yifan Wu, Zhengyang Chen, Chongyang Han, Xiaokang Chen,
 527 Leonidas Guibas, Gordon Wetzstein, and Jiajun Zhang. Blue noise for diffusion models. *ACM
 528 Transactions on Graphics (Proc. SIGGRAPH)*, 2024. URL <https://dl.acm.org/doi/10.1145/3641519.3657435>.

529 Zahra Kadkhodaie, Florentin Guth, Eero P Simoncelli, and Stéphane Mallat. Generalization
 530 in diffusion models arises from geometry-adaptive harmonic representation. *arXiv preprint
 531 arXiv:2310.02557*, 2023.

540 Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative
 541 adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern*
 542 *recognition*, pp. 4401–4410, 2019.

543

544 Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009.

545 Loic Matthey, Irina Higgins, Demis Hassabis, and Alexander Lerchner. dsprites: Disentanglement
 546 testing sprites dataset. <https://github.com/deepmind/dsprites-dataset/>, 2017.

547

548 Xingchao Peng, Qinxun Bai, Xide Xia, Zijun Huang, Kate Saenko, and Bo Wang. Moment matching
 549 for multi-source domain adaptation. In *Proceedings of the IEEE/CVF international conference on*
 550 *computer vision*, pp. 1406–1415, 2019.

551 Yurui Qian, Qi Cai, Yingwei Pan, Yehao Li, Ting Yao, Qibin Sun, and Tao Mei. Boosting diffusion
 552 models with moving average sampling in frequency domain. In *Proceedings of the IEEE/CVF*
 553 *Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8911–8920, June 2024.

554 Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Draxler, Min Lin, Fred Hamprecht, Yoshua
 555 Bengio, and Aaron Courville. On the Spectral Bias of Neural Networks. In *Proceedings of the*
 556 *36th International Conference on Machine Learning*, pp. 5301–5310. PMLR, May 2019. URL
 557 <https://proceedings.mlr.press/v97/rahaman19a.html>. ISSN: 2640-3498.

558

559 Santeri Rissanen, Konstantin Grigoryev, and Antti Honkela. Generative modelling with inverse
 560 heat dissipation. In *International Conference on Learning Representations (ICLR)*, 2023. URL
 561 <https://arxiv.org/abs/2206.13397>. arXiv:2206.13397.

562 Subham Sekhar Sahoo, Aaron Gokaslan, Chris De Sa, and Volodymyr Kuleshov. Diffusion Models
 563 With Learned Adaptive Noise, June 2024. URL <http://arxiv.org/abs/2312.13236>.
 564 arXiv:2312.13236 [cs].

565

566 Babak Saleh and Ahmed Elgammal. Large-scale classification of fine-art paintings: Learning the
 567 right metric on the right feature. *arXiv preprint arXiv:1505.00855*, 2015.

568

569 Luca Scimeca, Seong Joon Oh, Sanghyuk Chun, Michael Poli, and Sangdoo Yun. Which shortcut
 570 cues will dnns choose? a study from the parameter-space perspective. In *International Conference*
 571 *on Learning Representations*, 2021.

572

573 Luca Scimeca, Alexander Rubinstein, Armand Nicolicioiu, Damien Teney, and Yoshua Bengio.
 574 Leveraging diffusion disentangled representations to mitigate shortcuts in underspecified visual
 575 tasks. In *NeurIPS 2023 Workshop on Diffusion Models*, 2023a. URL <https://openreview.net/forum?id=AvUAVYRA70>.

576

577 Luca Scimeca, Alexander Rubinstein, Damien Teney, Seong Joon Oh, Armand Mihai Nicolicioiu,
 578 and Yoshua Bengio. Shortcut bias mitigation via ensemble diversity using diffusion probabilistic
 579 models. *arXiv preprint arXiv:2311.16176*, 2023b.

580

581 Marcin Sendera, Minsu Kim, Sarthak Mittal, Pablo Lemos, Luca Scimeca, Jarrid Rector-Brooks,
 582 Alexandre Adam, Yoshua Bengio, and Nikolay Malkin. On diffusion models for amortized
 583 inference: Benchmarking and improving stochastic control and sampling. *arXiv preprint*
 584 *arXiv:2402.05098*, 2024.

585

586 Jascha Sohl-Dickstein, Eric A. Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep Unsupervised
 587 Learning using Nonequilibrium Thermodynamics, November 2015. URL <http://arxiv.org/abs/1503.03585>. arXiv:1503.03585 [cond-mat, q-bio, stat].

588

589 Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *International*
 590 *Conference on Learning Representations (ICLR)*, 2021.

591

592 Naftali Tishby and Noga Zaslavsky. Deep learning and the information bottleneck principle. In *2015*
 593 *IEEE Information Theory Workshop (ITW)*, pp. 1–5. IEEE, 2015.

594

595 Siddarth Venkatraman, Moksh Jain, Luca Scimeca, Minsu Kim, Marcin Sendera, Mohsin Hasan, Luke
 596 Rowe, Sarthak Mittal, Pablo Lemos, Emmanuel Bengio, et al. Amortizing intractable inference in
 597 diffusion models for vision, language, and control. *arXiv preprint arXiv:2405.20971*, 2024.

594 Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. Extracting and
 595 composing robust features with denoising autoencoders. In *Proceedings of the 25th international*
 596 *conference on Machine learning*, pp. 1096–1103, 2008.

597

598 Vikram Voleti, Christopher Pal, and Adam Oberman. Score-based denoising diffusion with non-
 599 isotropic gaussian noise models. In *NeurIPS Workshop on Score-Based Methods*, 2022. URL
 600 <https://arxiv.org/abs/2210.12254>.

601 Zhi-Qin John Xu, Yaoyu Zhang, and Yanyang Xiao. Training behavior of deep neural net-
 602 work in frequency domain, October 2019. URL <http://arxiv.org/abs/1807.01251>.
 603 arXiv:1807.01251 [cs, math, stat].

604

605 Ling Yang, Zhilong Zhang, Yang Song, Shenda Hong, Runsheng Xu, Yue Zhao, Wentao Zhang,
 606 Bin Cui, and Ming-Hsuan Yang. Diffusion Models: A Comprehensive Survey of Methods and
 607 Applications, June 2024. URL <http://arxiv.org/abs/2209.00796>. arXiv:2209.00796
 608 [cs].

609 Xin Yuan, Linjie Li, Jianfeng Wang, Zhengyuan Yang, Kevin Lin, Zicheng Liu, and Lijuan Wang.
 610 Spatial-frequency u-net for denoising diffusion probabilistic models, 2023. URL <https://arxiv.org/abs/2307.14648>.

611

612 Ruicheng Zhang, Kanghui Tian, Zeyu Zhang, Qixiang Liu, and Zhi Jin. Fdg-diff: Frequency-
 613 domain-guided diffusion framework for compressed hazy image restoration. *arXiv preprint*
 614 *arXiv:2501.12832*, 2025.

615

616 Ziyi Zhang, Sen Zhang, Yibing Zhan, Yong Luo, Yonggang Wen, and Dacheng Tao. Confronting
 617 reward overoptimization for diffusion models: A perspective of inductive and primacy biases.
 618 *arXiv preprint arXiv:2402.08552*, 2024.

619

620 Boyang Zheng, Nanye Ma, Shengbang Tong, and Saining Xie. Diffusion transformers with represen-
 621 tation autoencoders. *arXiv preprint arXiv:2510.11690*, 2025.

622

623

624

625

626

627

628

629

630

631

632

633

634

635

636

637

638

639

640

641

642

643

644

645

646

647

648 A ANIMATED PARTICLE TRAJECTORIES (SUPPLEMENTARY VIDEOS)
649650 We include three short animations of particle flows under the probability-flow ODE (deterministic
651 DDIM; cf. Eq. equation 35). Each video shows N particles transported from an isotropic Gaussian
652 prior toward a fixed mixture-of-Gaussians target in \mathbb{R}^2 .
653654 **Files:**
655656 • `particles_alpha-0p5git`: SAGD with power-law weighting $w_\alpha(r) = (r + \varepsilon)^\alpha$
657 (Eq. equation 10) and $\alpha = -0.5$ (low-frequency tilt).
658 • `particles_iso.gif`: isotropic baseline, $\alpha = 0$ (white-noise forward covariance).
659 • `particles_alpha0p5.gif`: SAGD with $\alpha = +0.5$ (high-frequency tilt).
660661 All three runs share the same prior, target, and $\beta(t)$; only the forward covariance differs via $\Sigma_w =$
662 $\mathcal{F}^{-1}\text{Diag}(|w_\alpha|^2)\mathcal{F}$. As time decreases $t : 1 \rightarrow 0$, there is a perceived spatial bias in the particle
663 trajectory to the target. In all cases with full spectral support ($\Sigma_w \succ 0$), the endpoint score at $t \rightarrow 0$ is
664 consistent with the true data score. In practice, the anisotropic path deviations favor the learning of
665 particular aspects of the sampling distribution.
666667 B DATASETS
668669 For the experiments, we consider five datasets, namely: MNIST, CIFAR-10, Domainnet-Quickdraw,
670 Wiki-Art and FFHQ; providing examples of widely different visual distributions, scales, and domain-
671 specific statistics.672 **MNIST:** MNIST consists of 70,000 grayscale images of handwritten digits (0-9) (Matthey et al.,
673 2017). MNIST provides a simple test-bed to for the hypothesis in this work, as a well understood
674 dataset with well structured, and visually coherent samples.
675676 **CIFAR-10:** CIFAR-10 contains 60,000 color images distributed across 10 object categories
677 (Krizhevsky et al., 2009). The dataset is highly diverse in terms of object appearance, backgrounds,
678 and colors, with the wide-ranging visual variations across classes like animals, vehicles, and other
679 common objects.
680681 **DomainNet-Quickdraw:** DomainNet-Quickdraw features 120,750 sketch-style images, covering
682 345 object categories (Peng et al., 2019). These images, drawn in a minimalistic, abstract style,
683 present a distribution that is drastically different from natural images, with sparse details and heavy
684 visual simplifications.
685686 **WikiArt:** WikiArt consists of over 81,000 images of artwork spanning a wide array of artistic styles,
687 genres, and historical periods (Saleh & Elgammal, 2015). The dataset encompasses a rich and varied
688 distribution of textures, color palettes, and compositions, making it a challenging benchmark for
689 generative models, which must capture both the global structure and fine-grained stylistic variations
690 that exist across different forms of visual art.
691692 **FFHQ:** The Flickr-Faces-HQ (FFHQ) dataset comprises 70,000 high-resolution, aligned face
693 images (Karras et al., 2019), curated to increase diversity in age, ethnicity, pose, lighting, accessories
694 (e.g., eyeglasses, hats), and backgrounds. The dataset offers a rich distribution of facial features, and it
695 is a strong benchmark for testing generative models’ ability to capture fine-grained identity-preserving
696 details and global facial structure.
697698 **ImageNet-1k:** ImageNet-1k is a large-scale benchmark of over 1.2 million training images and
699 50,000 validation images, spanning 1,000 object categories drawn from everyday visual concepts such
700 as animals, vehicles, tools, and scenes (Deng et al., 2009). The dataset consists of high-resolution
701 high-variability natural images with complex backgrounds, diverse viewpoints, and significant intra-
class variation in appearance, scale, and context, making it a challenging benchmark for generative
models.
702

702 C IMPLEMENTATION DETAILS
703704 DATA PREPROCESSING
705706 We form a minimal preprocessing pipeline across datasets and vary only the target spatial resolution.
707 We standardize pixel intensities to $[-1, 1]$ and resample each image to at three resolutions: 32×32
708 (MNIST, CIFAR-10, and DomainNet-Quickdraw); 64×64 (WikiArt), 256×256 (FFHQ, and
709 Imagenet1k). MNIST is treated as single-channel (`channels` = 1), while all other datasets are
710 RGB (`channels` = 3).
711712 CODE BASE
713714 For our small- and medium-scale diffusion experiments on MNIST, CIFAR-10, DomainNet-
715 Quickdraw, WikiArt, and FFHQ, we build on the `diffusers` library from Hugging Face, using its
716 standard DDPM-style training loop and replacing only the forward noise with our SAGD operators.
717 All other components (optimizer, scheduler, and sampling code) follow the default configurations
718 described in the main text. For large-scale ImageNet-1k experiments, we adapt the public codebase
719 of Zheng et al. (2025), which implements latent DiT models in a DINOv2 feature space and has been
720 shown to reach state-of-the-art performance in diffusion modeling at scale. In this setting, we again
721 change only the forward noise generation to SAGD, leaving the architecture, optimizer, and training
722 schedule unchanged.
723724 DENOISER ARCHITECTURE
725726 For all 2D image experiments on MNIST, CIFAR-10, DomainNet-Quickdraw, WikiArt, and FFHQ,
727 we use a standard U-Net denoiser with four resolution levels and channel widths (32, 64, 128, 256)
728 in the encoder (mirrored in the decoder), resulting in approximately 15.9M trainable parameters.
729 This architecture is kept fixed across all SAGD and isotropic baselines. For ImageNet-1k, we use
730 the DiT-based latent diffusion model from Zheng et al. (2025), operating in the DINOv2 latent
731 space at 256×256 resolution with 196M parameters. We do not modify the DiT architecture or
732 hyperparameters; SAGD is introduced solely as a drop-in replacement for the isotropic forward noise,
733 demonstrating that our method is compatible with—and beneficial for—state-of-the-art large-scale
734 diffusion setups.
735736 POWER-LAW WEIGHTING IMPLEMENTATION
737738 **Discretization and batching.** In code, we construct the grid with $f_x(k) = \frac{k}{W} - \frac{1}{2}$ and $f_y(\ell) = \frac{\ell}{H} - \frac{1}{2}$
739 for $k \in \{0, \dots, W-1\}$, $\ell \in \{0, \dots, H-1\}$ (equivalently, `np.linspace(-0.5, 0.5, W)` and
740 `H`). This follows the standard `fft freq` convention, where frequencies are expressed in cycles per
741 pixel on $[-\frac{1}{2}, \frac{1}{2}]$ and $\pm \frac{1}{2}$ correspond to Nyquist. If one instead parameterizes frequencies on $[-1, 1]$
742 via $f' = 2f$ (and hence $r'(f') = 2r(f)$), the same power-law shape can be recovered by defining
743 $w'_\alpha(f') = (r'(f')/2 + \varepsilon)^\alpha$; any resulting global scale factor in w (and thus in Σ_w) is absorbed into
744 σ_t^2 or removed by the per-sample variance normalization used in our implementation, so the inductive
745 bias depends only on the relative weighting across frequencies, not on the chosen numeric range. The
746 weight w_α is broadcast across batch (and channels, if present). For convenience, one may multiply
747 in the `fftshift`-centered domain and undo the shift before the inverse FFT; this is equivalent to
748 multiplying in the unshifted domain since w_α is radial.
749750 **Optional variance calibration.** To keep $\mathbb{E}\|\epsilon^{(\alpha)}\|_2^2$ roughly constant across α , an energy-preserving
751 scalar

752
$$C_\alpha = \left(\frac{1}{HW} \sum_{u,v} |w_\alpha(f_{uv})|^2 \right)^{-\frac{1}{2}} \quad (15)$$

753

754 can be applied in equation 11, i.e., $\mathbf{N}_{\text{freq}}^{(\alpha)} \leftarrow C_\alpha \mathbf{N}_{\text{freq}} \cdot w_\alpha$. (Our experiments omit this by default,
755 matching the implementation in the main text.)

756 PRACTICAL CONSIDERATIONS.
757

758 We instantiate SAGD by sampling $\epsilon^{(w)}$ via FFTs, multiplying Fourier coefficients by a real, per-
759 frequency weight $w(\mathbf{f})$, and inverting to the spatial domain (Eqs. equation 5–equation 6). The
760 ℓ_2 training objective equation 3 remains unchanged, and converting ϵ_θ to a score only requires
761 multiplying by Σ_w^{-1} (see above). The computational overhead is non-existent at inference time, and
762 negligible during training, as FFTs dominate and weights are broadcastable; for RGB, we apply w
763 per channel.

764 D NOISE PARAMETERIZATION, SCORES, AND FREQUENCY-BASED DYNAMICS
765

766 In this section, we wish to formalize the role of frequency diffusion in correctly learning the gradient
767 of the log probability density of the data distribution at various noise levels (the score function). We
768 model frequency-based corruption as an anisotropic Gaussian forward process, derive the score– ϵ
769 relation for this general case, and prove that as $t \rightarrow 0$ the learned score converges to the true data score
770 whenever all frequencies are represented (full-rank covariance). We also derive the reverse/posterior
771 formulas and discuss how shaping the forward covariance changes the path to the score, shifting the
772 information burden across frequencies. Finally, we formalize the selective-omission case when some
773 bands are removed.

774 **Setup and notation.** Let $\alpha_t \in (0, 1)$ be the per-step scaling, $\bar{\alpha}_t = \prod_{s=1}^t \alpha_s$, and $\sigma_t^2 = 1 - \bar{\alpha}_t$. In
775 standard DDPM, the forward marginal is

$$776 q(\mathbf{x}_t | \mathbf{x}_0) = \mathcal{N}(\sqrt{\bar{\alpha}_t} \mathbf{x}_0, \sigma_t^2 \mathbf{I}), \quad (16)$$

777 and one trains an ϵ -predictor $\epsilon_\theta(\mathbf{x}_t, t)$ by minimizing

$$778 \mathcal{L} = \mathbb{E}_{t, \mathbf{x}_0, \epsilon \sim \mathcal{N}(0, \mathbf{I})} [\|\epsilon - \epsilon_\theta(\mathbf{x}_t, t)\|_2^2], \quad \mathbf{x}_t = \sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sigma_t \epsilon. \quad (17)$$

779 The optimal predictor is $\epsilon_\theta^*(\mathbf{x}_t, t) = \mathbb{E}[\epsilon | \mathbf{x}_t]$ and the true score relates to it via

$$780 \nabla_{\mathbf{x}_t} \log q_t(\mathbf{x}_t) = -\frac{1}{\sigma_t} \epsilon^*(\mathbf{x}_t, t), \quad \text{where } \epsilon^*(\mathbf{x}_t, t) = \mathbb{E}[\epsilon | \mathbf{x}_t]. \quad (18)$$

781 D.1 FREQUENCY-BASED FORWARD PROCESS AS ANISOTROPIC GAUSSIAN

782 Let $w(\mathbf{f}) > 0$ be a (time-independent) radial spectral weight and let \mathcal{F} denote the discrete Fourier
783 transform (unitary). The linear operator

$$784 \mathbf{T}_w := \mathcal{F}^{-1} \circ \text{Diag}(w(\mathbf{f})) \circ \mathcal{F} \quad (19)$$

785 maps spatial white noise to frequency-based noise. Writing $\xi \sim \mathcal{N}(0, \mathbf{I})$ and $\epsilon^{(w)} = \mathbf{T}_w \xi$, we have
786 $\epsilon^{(w)} \sim \mathcal{N}(0, \Sigma_w)$ with

$$787 \Sigma_w = \mathbf{T}_w \mathbf{T}_w^\top = \mathcal{F}^{-1} \text{Diag}(|w(\mathbf{f})|^2) \mathcal{F}, \quad (20)$$

788 i.e., Σ_w is circulant and diagonalized by the Fourier basis, with eigenvalues given by the power
789 spectrum $|w|^2$.²

790 Our forward process uses this shaped noise at each step:

$$791 \mathbf{x}_t = \sqrt{\alpha_t} \mathbf{x}_{t-1} + \sqrt{1 - \alpha_t} \epsilon_t^{(w)}, \quad \epsilon_t^{(w)} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Sigma_w). \quad (21)$$

792 A simple induction gives the marginal

$$793 q_w(\mathbf{x}_t | \mathbf{x}_0) = \mathcal{N}(\sqrt{\bar{\alpha}_t} \mathbf{x}_0, \sigma_t^2 \Sigma_w). \quad (22)$$

794 Hence, relative to equation 16, we have replaced the isotropic covariance by Σ_w , while $\bar{\alpha}_t$ and σ_t^2
795 remain unchanged.

805 **Support condition.** If $w(\mathbf{f}) > 0$ for all \mathbf{f} , then $\Sigma_w \succ 0$ (full rank) and the forward kernels have full
806 support in \mathbb{R}^{HW} . If w vanishes on a band, Σ_w is singular and the forward kernels are supported on a
807 strict subspace (Section D.5). In practice, adding a small DC floor (e.g., $r(\mathbf{f}) \mapsto r(\mathbf{f}) + \varepsilon$ with $\varepsilon > 0$)
808 ensures $w(0) > 0$ and thus $\Sigma_w \succ 0$.

809 ²With the usual Hermitian pairing in the discrete Fourier basis, $\epsilon^{(w)}$ is real-valued.

810 D.2 SCORE- ϵ RELATION UNDER ANISOTROPIC COVARIANCE
811

812 From equation 22,

813
$$\nabla_{\mathbf{x}_t} \log q_w(\mathbf{x}_t | \mathbf{x}_0) = -\frac{1}{\sigma_t^2} \Sigma_w^{-1} (\mathbf{x}_t - \sqrt{\bar{\alpha}_t} \mathbf{x}_0). \quad (23)$$

814
815

816 Taking the posterior expectation over $q_w(\mathbf{x}_0 | \mathbf{x}_t)$ and using $\mathbf{x}_t - \sqrt{\bar{\alpha}_t} \mathbf{x}_0 = \sigma_t \epsilon^{(w)}$, we obtain the
817 marginal score

818
$$\nabla_{\mathbf{x}_t} \log q_{w,t}(\mathbf{x}_t) = -\frac{1}{\sigma_t} \Sigma_w^{-1} \underbrace{\mathbb{E}[\epsilon^{(w)} | \mathbf{x}_t]}_{:= \epsilon_{\star}^{(w)}(\mathbf{x}_t, t)}. \quad (24)$$

819
820
821

822 Training with the natural generalization of equation 17,
823

824
$$\mathcal{L}^{(w)} = \mathbb{E}_{t, \mathbf{x}_0, \epsilon^{(w)} \sim \mathcal{N}(0, \Sigma_w)} [\|\epsilon^{(w)} - \epsilon_{\theta}(\mathbf{x}_t, t)\|_2^2], \quad \mathbf{x}_t = \sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sigma_t \epsilon^{(w)}, \quad (25)$$

825

826 the optimal predictor is $\epsilon_{\theta}^{\star}(\mathbf{x}_t, t) = \mathbb{E}[\epsilon^{(w)} | \mathbf{x}_t]$. Therefore, a consistent score estimator is
827

828
$$s_{\theta}(\mathbf{x}_t, t) := \nabla_{\mathbf{x}_t} \log q_{w,t}(\mathbf{x}_t) \approx -\frac{1}{\sigma_t} \Sigma_w^{-1} \epsilon_{\theta}(\mathbf{x}_t, t). \quad (26)$$

829
830

831 Equation equation 26 reduces to equation 18 when $\Sigma_w = \mathbf{I}$. Since the corruption covariance Σ_w is
832 fixed, the ℓ_2 objective needs no reweighting—the optimal ϵ_{θ} remains the conditional mean; Σ_w^{-1}
833 appears only when converting ϵ_{θ} to the score via Eq. equation 26.834 D.3 TWEEDIE’S IDENTITY AND THE LIMIT $t \rightarrow 0$
835836 Write the marginal as a (scaled) Gaussian smoothing of the data:
837

838
$$q_{w,t}(\mathbf{x}) = \int q(\mathbf{x}_0) \mathcal{N}(\mathbf{x} ; \sqrt{\bar{\alpha}_t} \mathbf{x}_0, \sigma_t^2 \Sigma_w) d\mathbf{x}_0. \quad (27)$$

839
840

841 Let $\mathbf{z}_t := \mathbf{x}_t / \sqrt{\bar{\alpha}_t}$; then $\mathbf{z}_t = \mathbf{x}_0 + \tilde{\sigma}_t \epsilon^{(w)}$ with $\tilde{\sigma}_t^2 = \sigma_t^2 / \bar{\alpha}_t$. The anisotropic Tweedie identity gives
842

843
$$\mathbb{E}[\mathbf{x}_0 | \mathbf{z}_t] = \mathbf{z}_t + \tilde{\sigma}_t^2 \Sigma_w \nabla_{\mathbf{z}_t} \log p_t(\mathbf{z}_t), \quad p_t = \text{law}(\mathbf{z}_t). \quad (28)$$

844

845 Equivalently, in the original variable,

846
$$\nabla_{\mathbf{x}_t} \log q_{w,t}(\mathbf{x}_t) = \frac{\sqrt{\bar{\alpha}_t}}{\sigma_t^2} \Sigma_w^{-1} \left(\mathbb{E}[\mathbf{x}_0 | \mathbf{x}_t] - \frac{\mathbf{x}_t}{\sqrt{\bar{\alpha}_t}} \right). \quad (29)$$

847
848

849 As $t \rightarrow 0$, $\bar{\alpha}_t \rightarrow 1$, $\sigma_t \rightarrow 0$, and $q_{w,t} \Rightarrow q$. If $\Sigma_w \succ 0$ and q admits a locally positive C^1 density with
850 $\nabla \log q \in L^1_{\text{loc}}$, the anisotropic Gaussian mollifier is an approximate identity and
851

852
$$\lim_{t \rightarrow 0} \nabla_{\mathbf{x}_t} \log q_{w,t}(\mathbf{x}_t) = \nabla_{\mathbf{x}} \log q(\mathbf{x}) \quad \text{for a.e. } \mathbf{x}. \quad (30)$$

853
854

855 Intuitively, the anisotropic Gaussian kernel in equation 27 shrinks to a Dirac as $\sigma_t \rightarrow 0$ regardless
856 of its orientation, so the smoothed score converges to the true data score. Combining equation 24–
857 equation 30, the ϵ -parameterization with frequency-based noise yields a correct score at $t = 0$,
858 provided $\Sigma_w \succ 0$.859 To visualize how frequency noising alters trajectories and score geometry through time, Fig. D.1 shows
860 particle flows under the probability-flow ODE for isotropic noise (top), and plw -SAGD high ($\alpha=0.1$,
861 middle), and low-frequency ($\alpha=-0.1$, bottom) tilt, while Fig. D.2 shows the corresponding score
862 fields $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$ at five equally spaced times. Frequency noising changes the path deterministically
863 by reweighting modes through Σ_w , while preserving the $t \rightarrow 0$ endpoint score under full support (see
Sec. D).

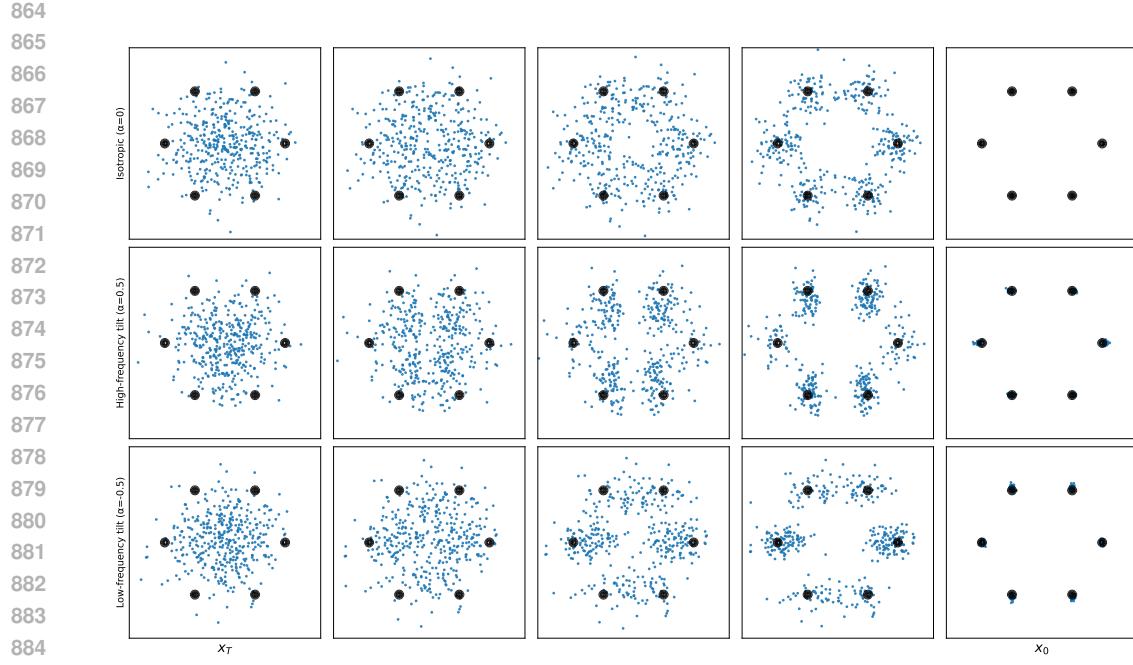


Figure D.1: Particle trajectories under the probability-flow ODE from a Gaussian prior to a mixture-of-Gaussians target (black contours), visualized at five equally spaced times (left to right). Rows: (top) isotropic noise ($\alpha=0$), (middle) high-frequency tilt ($\alpha=0.1$), (bottom) low-frequency tilt ($\alpha=-0.1$). *plw*-SAGD alters the path by reweighting modes via Σ_w while keeping the endpoint consistent under full support (cf. Sec. D).

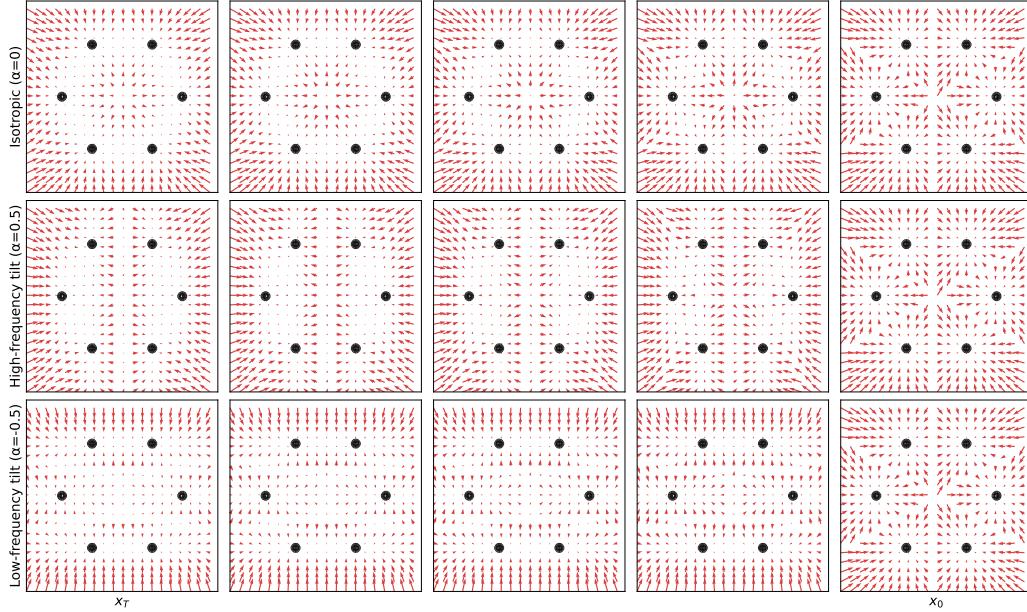


Figure D.2: Evolving score fields $\nabla_x \log p_t(x)$ for the same three settings as Fig. D.1. Arrows indicate the instantaneous score on a grid; black contours show the target density. *plw*-SAGD stretches/compresses the field along principal modes, biasing the trajectory toward frequencies emphasized by Σ_w .

918 D.4 REVERSE/POSTERIOR WITH FREQUENCY-BASED NOISE
919

920 Since all covariances are proportional to the same Σ_w , linear-Gaussian posteriors retain the standard
921 scalar coefficients while the covariances inherit Σ_w as a factor. In particular,

$$922 \quad q_w(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{x}_0) = \mathcal{N}\left(\tilde{\mu}_t(\mathbf{x}_t, \mathbf{x}_0), \tilde{\beta}_t \Sigma_w\right), \quad (31)$$

923 with
925

$$926 \quad \tilde{\beta}_t = \frac{1 - \bar{\alpha}_{t-1}}{1 - \bar{\alpha}_t} (1 - \alpha_t), \quad (32)$$

$$928 \quad \tilde{\mu}_t(\mathbf{x}_t, \mathbf{x}_0) = \frac{1}{\sqrt{\bar{\alpha}_t}} \left(\mathbf{x}_t - \frac{1 - \alpha_t}{1 - \bar{\alpha}_t} (\mathbf{x}_t - \sqrt{\bar{\alpha}_t} \mathbf{x}_0) \right). \quad (33)$$

930 Replacing \mathbf{x}_0 by $\hat{\mathbf{x}}_0$ in equation 33 yields the usual mean update. For the ϵ -parameterization we
931 recover an estimate of \mathbf{x}_0 via
932

$$933 \quad \hat{\mathbf{x}}_0(\mathbf{x}_t, t) = \frac{1}{\sqrt{\bar{\alpha}_t}} (\mathbf{x}_t - \sigma_t \epsilon_\theta(\mathbf{x}_t, t)). \quad (34)$$

936 **Stochastic sampling:** if one samples stochastically (e.g., DDPM), the injected noise should be drawn
937 as $\eta_t^{(w)} \sim \mathcal{N}(0, \Sigma_w)$ (not $\mathcal{N}(0, \mathbf{I})$) for consistency with the forward process.

938 **Probability-flow ODE:** if one uses the deterministic sampler (e.g., DDIM), no step noise is injected.
939 In continuous time, the associated probability-flow ODE with frequency-based forward noise reads
940

$$941 \quad \frac{d\mathbf{x}}{dt} = -\frac{1}{2} \beta(t) \mathbf{x} - \frac{1}{2} \beta(t) \Sigma_w \nabla_{\mathbf{x}} \log p_t(\mathbf{x}), \quad (35)$$

943 which reduces to the standard probability-flow ODE when $\Sigma_w = \mathbf{I}$. In practice with the ϵ -
944 parameterization, one uses $\hat{\mathbf{x}}_0$ from equation 34 in the standard DDIM deterministic update; no
945 extra noise term appears.
946

947 D.5 SELECTIVE OMISSION AND RANK-DEFICIENT Σ_w
948

949 If w vanishes on a measurable band, then $\Sigma_w \succeq 0$ is singular. The forward kernels in equation 22 are
950 supported on an affine subspace determined by $\text{range}(\Sigma_w)$, and the smoothed marginals $q_{w,t}$ are not
951 strictly positive in \mathbb{R}^{HW} . The score $\nabla \log q_{w,t}$ exists only on that subspace and is undefined along
952 the null space. Training with equation 25 then recovers the projected score, i.e., the model learns to
953 ignore the omitted bands by construction (this is the mechanism exploited in our corruption-recovery
954 experiments).

955 D.6 PATH TO THE SCORE EFFECTS OF SAGD
956

957 Even though the $t \rightarrow 0$ limit recovers the true data score under $\Sigma_w \succ 0$, the evolution of the score with
958 t changes substantially:
959

- 960 **Geometry of the score.** From equation 24, the conversion from ϵ -prediction to score
961 multiplies by Σ_w^{-1} . In the Fourier basis (where Σ_w is diagonal), modes with larger variance
962 (large $|w|^2$) are downweighted in the score, while low-variance modes are amplified. Thus,
963 shaping the forward spectrum changes the relative gradient magnitudes across frequencies
964 during training and sampling.
- 965 **Signal-to-noise during supervision.** The target $\epsilon^{(w)}$ has covariance Σ_w , so its per-mode
966 variance follows $|w|^2$. The ℓ_2 loss in equation 25 therefore exposes the model to larger target
967 amplitudes (and larger gradients) in bands where $|w|$ is large, shifting the inductive bias
968 toward fitting those modes sooner/more accurately.
- 969 **Reverse dynamics.** The reverse posterior covariance in equation 31 is $\tilde{\beta}_t \Sigma_w$, so the
970 stochasticity injected at each reverse step is anisotropic. This changes the trajectory taken
971 from t down to 0, biasing the generation process to consolidate structure along directions
972 favored by Σ_w . Under DDIM, no step noise is instead injected, so the anisotropic stochastic

972 effect disappears. However, the drift in the probability-flow ODE equation 35 still carries
 973 Σ_w through the term $-\beta(t) \Sigma_w \nabla_{\mathbf{x}} \log p_t(\mathbf{x})$, inducing trajectories to be frequency-biased
 974 deterministically. In this context, modes emphasized by Σ_w contribute more strongly to the
 975 drift, reshaping the path from $t=T$ to $t=0$ even without randomness.
 976

977 Collectively, these effects explain why different datasets benefit from different w : the endpoint
 978 score is consistent (under full rank), but the path—and thus the optimization landscape and sample
 979 trajectories—is reshaped by frequency weighting.

980 **Time-varying weights.** If one uses a schedule $w_t(\mathbf{f})$, the t -step marginal covariance becomes a
 981 scalar-weighted sum of commuting matrices:
 982

$$983 \quad \text{Cov}(\mathbf{x}_t \mid \mathbf{x}_0) = \sum_{s=1}^t \left(\beta_s \prod_{k=s+1}^t \alpha_k \right) \Sigma_{w_s}, \quad \beta_s := 1 - \alpha_s. \quad (36)$$

984 When all Σ_{w_s} are diagonal in the Fourier basis (true for any per-frequency diagonal weight, not neces-
 985 sarily radial), the analysis carries through modewise with eigenvalues replaced by the corresponding
 986 positive weighted sums $\sum_s w_s |w_s(\mathbf{f})|^2$ (which form a convex combination after normalization by
 987 $\sigma_t^2 = \sum_s w_s$).
 988

989 E REFERENCE CODE FOR SAGD NOISE GENERATORS

990 Below we provide minimal, implementation-oriented pseudo-code for the two SAGD operators used
 991 in this paper: (i) a *power-law* (radial) weighting and (ii) a *band-pass* mask (building block for the
 992 two-band mixture). Both follow the same template:
 993

$$994 \quad \epsilon \sim \mathcal{N}(0, \mathbf{I}) \xrightarrow{\mathcal{F}} \mathbf{N}_{\text{freq}} \xrightarrow{w(\mathbf{f})} \mathbf{N}_{\text{freq}}^{(w)} \xrightarrow{\mathcal{F}^{-1}} \epsilon^{(w)}.$$

995 We normalize the spatial noise to unit variance so that the overall scale still comes from the schedule
 996 via σ_t . Multiplying by a *real*, per-frequency weight preserves Hermitian symmetry and thus yields a
 997 real inverse transform.
 998

1000 **Power-law weighting.** This implements a *plw*-SAGD with $w_\alpha(\mathbf{f}) = (r(\mathbf{f}) + \varepsilon)^\alpha$, $r(\mathbf{f}) = \sqrt{f_x^2 + f_y^2}$
 1001 and a small DC floor $\varepsilon > 0$.
 1002

```
1003
1004 import numpy as np
1005
1006
1007
1008 def sagd_powerlaw_noise(shape, alpha, eps=1e-10):
1009     # shape: (B, C, H, W)
1010     B, C, H, W = shape
1011     # 1) white noise
1012     eps_white = np.random.randn(B, C, H, W)
1013     # 2) FFT over spatial axes
1014     F = np.fft.fftn(eps_white, axes=(-2, -1))
1015     # 3) radial grid in normalized frequency (cycles/pixel)
1016     fy = np.fft.freq(H)[:, None] # shape Hx1
1017     fx = np.fft.freq(W)[None, :] # shape 1xW
1018     r = np.sqrt(fx**2 + fy**2) # shape HxW
1019     # 4) power-law weight
1020     w = (r + eps)**alpha # shape HxW (real)
1021     # 5) apply weight and invert
1022     Fw = F * w[None, None, ...]
1023     eps_w = np.fft.ifftn(Fw, axes=(-2, -1)).real
1024     # 6) unit-variance normalization (per-sample)
1025     std = eps_w.std(axis=(-2, -1), keepdims=True) + 1e-8
1026     return eps_w / std
```

```

1026
1027 Band-pass mask (single band  $[a, b]$ ). This implements a bpm-SAGD with  $w_{a,b}(\mathbf{f}) = \mathbf{1}\{a \leq r(\mathbf{f}) \leq b\}$ . Cutoffs  $a, b$  are given in frequency units of cycles/pixel (e.g.,  $a=0.1, b=0.3$ ); we can also
1028 express them as fractions of the Nyquist radius.
1029
1030 import numpy as np
1031
1032 def sagd_bandpass_noise(shape, a, b):
1033     # shape: (B, C, H, W); choose 0 <= a <= b <= ~sqrt(2)/2 (cycles/pixel)
1034     B, C, H, W = shape
1035     # 1) white noise
1036     eps_white = np.random.randn(B, C, H, W)
1037     # 2) FFT over spatial axes
1038     F = np.fft.fftn(eps_white, axes=(-2, -1))
1039     # 3) radial grid in normalized frequency (cycles/pixel)
1040     fy = np.fft.fftfreq(H)[ :, None] # shape Hx1
1041     fx = np.fft.fftfreq(W)[None, :] # shape 1xW
1042     r = np.sqrt(fx**2 + fy**2) # shape HxW
1043     # 4) band-pass mask (keep a <= r <= b)
1044     M = ((r >= a) & (r <= b)).astype(float) # shape HxW
1045     # 5) apply mask and invert
1046     Fw = F * M[None, None, ...]
1047     eps_w = np.fft.ifftn(Fw, axes=(-2, -1)).real
1048     # 6) unit-variance normalization (per-sample)
1049     std = eps_w.std(axis=(-2, -1), keepdims=True) + 1e-8
1050     return eps_w / std
1051
1052
1053 Two-band mixture. We combine two band-pass noises with nonnegative coefficients  $\gamma_l, \gamma_h$  (typically  $\gamma_l+\gamma_h=1$ ):
1054
1055     
$$\epsilon^{(w)} = \gamma_l \text{sagd\_bandpass\_noise}(\cdot, a_l, b_l) + \gamma_h \text{sagd\_bandpass\_noise}(\cdot, a_h, b_h).$$

1056
1057 SAGD Forward step. Replace the isotropic noise in the DDPM forward step with either generator
1058 above:
1059
1060     
$$\mathbf{x}_t = \sqrt{\alpha_t} \mathbf{x}_{t-1} + \sqrt{1 - \alpha_t} \epsilon^{(w)}.$$

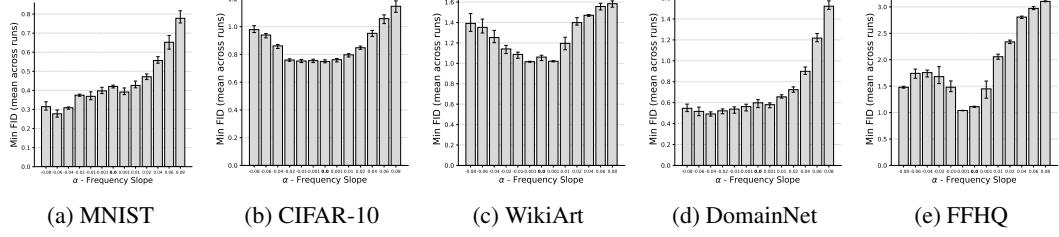
1061
1062 The training loss and DDIM update remain unchanged; when converting  $\epsilon_\theta$  to a score, multiply by
1063  $\Sigma_w^{-1}$  (diagonal in the Fourier basis).
1064
1065
1066 F ADDITIONAL RESULTS
1067
1068
1069 F.1 SAGD IN NATURAL DATASETS - FULL RESULTS
1070
1071
1072 We present in Table F.1 and Table F.2 the full set of results from running pwd-SAGD on all datasets
1073 and values of  $\alpha$  considered. In Figure F.1, we can better discern the learning performance over
1074 different  $\alpha$  settings, where we observe interesting monotonically decreasing performance (increasing
1075 FID) for MNIST and DomainNet, showcasing improved learning for negative values of  $\alpha$ , and
1076 suggesting that semantically informative content (strokes/contours) for these datasets may lie in
1077 the lower frequency ranges. By contrast, CIFAR-10, Wiki-Art, and FFHQ exhibit a much flatter
1078 dependence with shallow optima near  $\alpha \approx 0$  (within  $\pm 0.02$ ), consistent with their broader, mixed
1079 spectra. In Figure F.1, we can better discern the learning performance over different  $\alpha$  settings, where
1080 we observe interesting monotonically decreasing performance (increasing FID) for MNIST and
1081 DomainNet, showcasing improved learning for negative values of  $\alpha$ , and suggesting that semantically
1082 informative content (strokes/contours) for these datasets may lie in the lower frequency ranges. By
1083 contrast, CIFAR-10, Wiki-Art, and FFHQ exhibit a much flatter dependence with shallow optima
1084 near  $\alpha \approx 0$  (within  $\pm 0.02$ ), consistent with their broader, mixed spectra.
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
20100
20101
20102
20103
20104
20105
20106
20107
20108
20109
20110
20111
20112
20113
20114
20115
20116
20117
20118
20119
20120
20121
20122
20123
20124
20125
20126
20127
20128
20129
20130
20131
20132
20133
20134
20135
20136
20137
20138
20139
20140
20141
20142
20143
20144
20145
20146
20147
20148
20149
20150
20151
20152
20153
20154
20155
20156
20157
20158
20159
20160
20161
20162
20163
20164
20165
20166
20167
20168
20169
20170
20171
20172
20173
20174
20175
20176
20177
20178
20179
20180
20181
20182
20183
20184
20185
20186
20187
20188
20189
20190
20191
20192
20193
20194
20195
20196
20197
20198
20199
20200
20201
20202
20203
20204
20205
20206
20207
20208
20209
20210
20211
20212
20213
20214
20215
20216
20217
20218
20219
20220
20221
20222
20223
20224
20225
20226
20227
20228
20229
20230
20231
20232
20233
20234
20235
20236
20237
20238
20239
20240
20241
20242
20243
20244
20245
20246
20247
20248
20249
20250
20251
20252
20253
20254
20255
20256
20257
20258
20259
20260
20261
20262
20263
20264
20265
20266
20267
20268
20269
20270
20271
20272
20273
20274
20275
20276
20277
20278
20279
20280
20281
20282
20283
20284
20285
20286
20287
20288
20289
20290
20291
20292
20293
20294
20295
20296
20297
20298
20299
20300
20301
20302
20303
20304
20305
20306
20307
20308
20309
20310
20311
20312
20313
20314
20315
20316
20317
20318
20319
20320
20321
20322
20323
20324
20325
20326
20327
20328
20329
20330
20331
20332
20333
20334
20335
20336
20337
20338
20339
20340
20341
20342
20343
20344
20345
20346
20347
20348
20349
20350
20351
20352
20353
20354
20355
20356
20357
20358
20359
20360
20361
20362
20363
20364
20365
20366
20367
20368
20369
20370
20371
20372
20373
20374
20375
20376
20377
20378
20379
20380
20381
20382
20383
20384
20385
20386
20387
20388
20389
20390
20391
20392
20393
20394
20395
20396
20397
20398
20399
20400
20401
20402
20403
20404
20405
20406
20407
20408
20409
20410
20411
20412
20413
20414
20415
20416
20417
20418
20419
20420
20421
20422
20423
20424
20425
20426
20427
20428
20429
20430
20431
20432
20433
20434
20435
20436
20437
20438
20439
20440
20441
20442
20443
20444
20445
20446
20447
20448
20449
20450
20451
20452
20453
20454
20455
20456
20457
20458
20459
20460
20461
20462
20463
20464
20465
20466
20467
20468
20469
20470
20471
20472
20473
20474
20475
20476
20477
20478
20479
20480
20481
20482
20483
20484
20485
20486
20487
20488
20489
20490
20491
20492
20493
20494
20495
20496
20497
20498
20499
20500
20501
20502
20503
20504
20505
20506
20507
20508
20509
20510
20511
20512
20513
20514
20515
20516
20517
20518
20519
20520
20521
20522
20523
20524
20525
20526
20527
20528
20529
20530
20531
20532
20533
20534
20535
20536
20537
20538
20539
20540
20541
20542
20543
20544
20545
20546
20547
20548
20549
20550
20551
20552
20553
20554
20555
20556
20557
20558
20559
20560
20561
20562
20563
20564
20565
20566
20567
20568
20569
20570
20571
20572
20573
20574
20575
20576
20577
20578
20579
20580
20581
20582
20583
20584
20585
20586
20587
20588
20589
20590
20591
20592
20593
20594
20595
20596
20597
20598
20599
20600
20601
20602
20603
20604
20605
20606
20607
20608
20609
20610
20611
20612
20613
20614
20615
20616
20617
20618
20619
20620
20621
20622
20623
20624
20625
20626
20627
20628
20629
20630
20631
20632
20633
20634
20635
20636
20637
20638
20639
20640
20641
20642
20643
20644
20645
20646
20647
20648
20649
20650
20651
20652
20653
20654
20655
20656
20657
20658
20659
20660
20661
20662
20663
20664
20665
20666
20667
20668
20669
20670
20671
20672
20673
20674
20675
20676
20677
20678
20679
20680
20681
20682
20683
20684
20685
20686
20687
20688
20689
20690
20691
20692
20693
20694
20695
20696
20697
20698
20699
20700
20701
20702
20703
20704
20705
20706
20707
20708
20709
20710
20711
20712
20713
20714
20715
20716
20717
20718
20719
20720
20721
20722
20723
20724
20725
20726
20727
20728
20729
20730
20731
20732
20733
20734
20735
20736
20737
20738
20739
20740
20741
20742
20743
20744
20745
20746
20747
20748
20749
20750
20751
20752
20753
20754
20755
20756
20757
20758
20759
20760
20761
20762
20763
20764
20765
20766
20767
20768
20769
20770
20771
20772
20773
20774
20775
20776
20777
20778
20779
20780
20781
20782
20783
20784
20785
20786
20787
20788
20789
20790
20791
20792
20793
20794
20795
20796
20797
20798
20799
20800
20801
20802
20803
20804
20805
20806
20807
20808
20809
20810
20811
20812
20813
20814
20815
20816
20817
20818
20819
20820
20821
20822
20823
20824
20825
20826
20827
20828
20829
20830
20831
20832
20833
20834
20835
20836
20837
20838
20839
20840
20841
20842
20843
20844
20845
20846
20847
20848
20849
20850
20851
20852
20853
20854
20855
20856
20857
20858
20859
20860
20861
20862
20863
20864
20865
20866
20867
20868
20869
20870
20871
20872
20873
20874
20875
20876
20877
20878
20879
20880
20881
20882
20883
20884
20885
20886
20887
20888
20889
20890
20891
20892
20893
20894
20895
20896
20897
20898
20899
20900
20901
20902
20903
20904
20905
20906
20907
20908
```

1080
1081
1082 Table F.1: Results for FID across different α (frequency slope) settings (mean \pm standard error across
1083 seeds). The setting for $\alpha = 0$ corresponds to standard DPM training (baseline).
1084

Dataset →	MNIST	CIFAR-10	Domainnet-Quickdraw	Wiki-Art	FFHQ	ImageNet1k
baseline	$0.42 \pm 8.52e - 03$	0.75 ± 0.01	0.60 ± 0.05	1.06 ± 0.03	1.11 ± 0.01	8.6819 ± 0.0739
$\alpha = -0.080$	0.31 ± 0.03	0.98 ± 0.03	0.55 ± 0.04	1.39 ± 0.10	1.48 ± 0.02	-
$\alpha = -0.060$	0.28 ± 0.02	0.94 ± 0.02	0.52 ± 0.05	1.35 ± 0.08	1.74 ± 0.10	8.1098 ± 0.0229
$\alpha = -0.040$	$0.31 \pm 7.76e - 03$	0.86 ± 0.02	0.49 ± 0.02	1.25 ± 0.07	1.76 ± 0.08	7.5534 ± 0.0556
$\alpha = -0.020$	$0.37 \pm 6.36e - 03$	0.76 ± 0.01	0.52 ± 0.03	1.14 ± 0.05	1.68 ± 0.19	7.6419 ± 0.0581
$\alpha = -0.010$	0.37 ± 0.02	0.75 ± 0.01	0.54 ± 0.04	1.09 ± 0.04	1.48 ± 0.12	8.0400 ± 0.0236
$\alpha = -0.001$	0.40 ± 0.02	0.76 ± 0.01	0.56 ± 0.04	$1.02 \pm 5.66e - 03$	$1.04 \pm 5.17e - 03$	8.5288 ± 0.0112
$\alpha = 0.001$	0.39 ± 0.02	0.76 ± 0.02	0.58 ± 0.03	$1.02 \pm 6.63e - 03$	1.45 ± 0.19	-
$\alpha = 0.010$	0.43 ± 0.02	0.80 ± 0.02	0.66 ± 0.02	1.20 ± 0.07	2.06 ± 0.06	9.3867 ± 0.0348
$\alpha = 0.020$	0.47 ± 0.02	0.85 ± 0.01	0.72 ± 0.03	1.40 ± 0.05	2.34 ± 0.04	-
$\alpha = 0.040$	0.56 ± 0.02	0.95 ± 0.02	0.90 ± 0.04	1.47 ± 0.01	2.81 ± 0.03	-
$\alpha = 0.060$	0.65 ± 0.04	1.06 ± 0.04	1.22 ± 0.05	1.56 ± 0.04	2.97 ± 0.03	-
$\alpha = 0.080$	0.78 ± 0.04	1.15 ± 0.05	1.52 ± 0.05	1.58 ± 0.04	$3.10 \pm 9.95e - 03$	-

1097
1098
1099
1100 Table F.2: Results for KID across different α (frequency slope) settings (mean \pm standard error across
1101 seeds). The setting for $\alpha = 0$ corresponds to standard DPM training (baseline).
1102

Dataset →	MNIST	CIFAR-10	Domainnet-Quickdraw	Wiki-Art	FFHQ
baseline	$9.16e - 04 \pm 6.44e - 05$	$2.11e - 04 \pm 1.15e - 05$	$7.15e - 04 \pm 1.06e - 04$	$8.08e - 04 \pm 1.28e - 04$	$1.42e - 03 \pm 8.12e - 05$
$\alpha = -0.080$	$5.63e - 04 \pm 7.19e - 05$	$8.64e - 04 \pm 6.72e - 05$	$6.54e - 04 \pm 5.45e - 05$	$2.03e - 03 \pm 3.32e - 04$	$2.79e - 03 \pm 1.01e - 04$
$\alpha = -0.060$	$4.53e - 04 \pm 8.19e - 05$	$7.42e - 04 \pm 5.04e - 05$	$5.28e - 04 \pm 8.75e - 05$	$1.97e - 03 \pm 3.32e - 04$	$3.81e - 03 \pm 3.43e - 04$
$\alpha = -0.040$	$5.29e - 04 \pm 5.44e - 05$	$4.71e - 04 \pm 5.95e - 05$	$4.73e - 04 \pm 1.54e - 05$	$1.69e - 03 \pm 2.94e - 04$	$3.75e - 03 \pm 3.14e - 04$
$\alpha = -0.020$	$7.87e - 04 \pm 2.19e - 05$	$2.25e - 04 \pm 1.42e - 05$	$5.47e - 04 \pm 4.14e - 05$	$1.32e - 03 \pm 1.50e - 04$	$3.44e - 03 \pm 6.55e - 04$
$\alpha = -0.010$	$7.62e - 04 \pm 8.03e - 05$	$1.90e - 04 \pm 2.01e - 05$	$5.87e - 04 \pm 7.17e - 05$	$9.64e - 04 \pm 1.16e - 04$	$2.82e - 03 \pm 3.94e - 04$
$\alpha = -0.001$	$8.32e - 04 \pm 6.08e - 05$	$2.29e - 04 \pm 3.56e - 05$	$6.59e - 04 \pm 8.39e - 05$	$6.87e - 04 \pm 7.94e - 05$	$1.23e - 03 \pm 5.66e - 05$
$\alpha = 0.001$	$8.00e - 04 \pm 6.49e - 05$	$2.49e - 04 \pm 3.57e - 05$	$7.02e - 04 \pm 6.25e - 05$	$7.21e - 04 \pm 7.45e - 05$	$2.48e - 03 \pm 6.59e - 04$
$\alpha = 0.010$	$9.49e - 04 \pm 8.17e - 05$	$3.28e - 04 \pm 2.90e - 05$	$8.99e - 04 \pm 5.53e - 05$	$1.36e - 03 \pm 2.58e - 04$	$4.74e - 03 \pm 1.00e - 04$
$\alpha = 0.020$	$1.09e - 03 \pm 6.57e - 05$	$5.05e - 04 \pm 2.52e - 05$	$1.06e - 03 \pm 7.36e - 05$	$2.18e - 03 \pm 2.05e - 04$	$5.80e - 03 \pm 1.26e - 04$
$\alpha = 0.040$	$1.39e - 03 \pm 7.64e - 05$	$8.03e - 04 \pm 5.90e - 05$	$1.51e - 03 \pm 1.30e - 04$	$2.45e - 03 \pm 1.26e - 04$	$7.65e - 03 \pm 8.05e - 05$
$\alpha = 0.060$	$1.72e - 03 \pm 1.53e - 04$	$1.12e - 03 \pm 8.61e - 05$	$2.37e - 03 \pm 1.47e - 04$	$2.83e - 03 \pm 2.40e - 04$	$8.37e - 03 \pm 1.59e - 04$
$\alpha = 0.080$	$2.18e - 03 \pm 1.59e - 04$	$1.37e - 03 \pm 1.09e - 04$	$3.30e - 03 \pm 1.68e - 04$	$2.94e - 03 \pm 1.84e - 04$	$8.87e - 03 \pm 7.44e - 05$



1120
1121
1122 Figure F.1: Minimum FID vs. frequency slope (α). Bars show mean FID with inter-quartile error
1123 bars across runs across three seeds.
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133