Manifolds, Random Matrices and Spectral Gaps: The geometric phases of generative diffusion

Anonymous authors
Paper under double-blind review

Abstract

In this paper, we investigate the latent geometry of generative diffusion models under the manifold hypothesis. To this purpose, we analyze the spectrum of eigenvalues (and singular values) of the Jacobian of the score function, whose discontinuities (gaps) reveal the presence and dimensionality of distinct sub-manifolds. Using a statistical physics approach, we derive the spectral distributions and formulas for the spectral gaps under several distributional assumptions and we compare these theoretical predictions with the spectra estimated from trained networks. Our analysis reveals the existence of three distinct qualitative phases during the generative process: a trivial phase; a manifold coverage phase where the diffusion process fits the distribution internal to the manifold; a consolidation phase where the score becomes orthogonal to the manifold and all particles are projected on the support of the data. This 'division of labor' between different timescales provides an elegant explanation on why generative diffusion models are not affected by the manifold overfitting phenomenon that plagues likelihood-based models, since the internal distribution and the manifold geometry are produced at different time points during generation.

1 Introduction

Generative diffusion models have revolutionized the fields of computer vision and generative modeling, achieving state-of-the-art performance on image generation (Ho et al., 2020; Song and Ermon, 2019; Song et al., 2021) and video generation (Ho et al., 2022; Singer et al., 2022; Blattmann et al., 2023; Brooks et al., 2024). Generative diffusion models synthesize images through a stochastic dynamical denoising process. Experimental and theoretical arguments suggest that different features such as frequency modes and class labels are generated at different times during the process. For example, it has been shown that separation between isolated classes, like in the case of mixture of Gaussian models, happens in critical phase transition points of spontaneous symmetry breaking (speciation events) (Biroli et al., 2024). It is also well known that subspaces corresponding to different frequency models emerge at different times of diffusion (Kingma and Gao, 2023). This idea has been recently refined by (Kadkhodaie et al., 2024), who showed that diffusion models give rise to a local decomposition of the image manifold into a basis of geometry-adaptive harmonic basis functions. These decomposition phenomena cannot be directly explained in terms of critical phase transitions as they are fundamentally linear processes. In this paper, we will provide a precise theoretical analysis of the separation of subspaces for data defined on low dimensional linear manifolds.

Our main contributions are: I) an in-depth theoretical random-matrix analysis of the distribution of Jacobian spectra in diffusion models on linear manifolds and II) a detailed experimental analysis of Jacobian spectra extracted from trained networks on linear manifolds and on image datasets. The analysis of this spectra is important as it provides a detailed picture of the latent geometry that guides the generative diffusion process. We show that the linear theory predicts several phenomena that we observed in trained networks. Based on our result, we divide the generative process in three qualitatively different phases: **trivial phase**, **manifold coverage phase** and **manifold consolidation phase**. Using these concepts, we

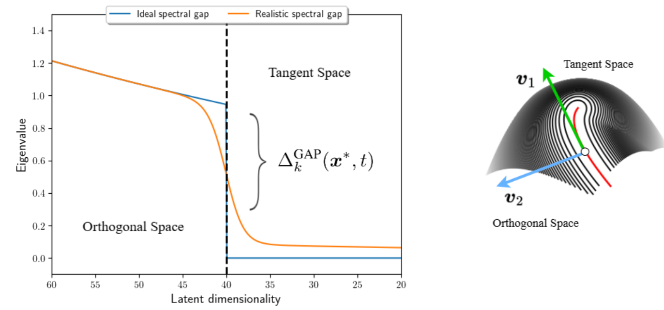


Figure 1: Visualization of the gaps in the spectrum of the (negative) Jacobian of the score for data supported on a latent manifold. Blue line: idealize spectrum of distribution with uniform internal density; Orange line: spectrum of a more realistic distribution.

provide an elegant explanation of why diffusion models can avoid the manifold overfitting pathology that characterizes likelihood-based models (Loaiza-Ganem et al., 2022).

2 The manifold hypothesis

The manifold hypothesis states that the distribution on natural data such as images and sound recordings is (approximately) supported on a $m-$dimensional manifolds $\mathcal{M}$ embedded in a larger euclidean ambient space $\mathbb{R}^d$ (Peyré, 2009; Fefferman et al., 2016). The probability of data supported on a $m < d$ manifold $\mathcal{M}$ cannot be expressed using a a density function, and it can instead be written as

$$p_0(\boldsymbol{x}) = \delta(1 - \mathcal{I}_{\mathcal{M}}(\boldsymbol{x}))\, \rho_{\text{int}}(\boldsymbol{x}) \ , \tag{1}$$

where $\mathcal{I}_{\mathcal{M}}(\boldsymbol{x})$ is the index function of the manifold and the *internal density* $\rho_{\text{int}}(\boldsymbol{x})$ is non-zero for $\boldsymbol{x} \in \mathcal{M}$. Loosely speaking, this expression is divergent on the manifold and zero everywhere else in order to ensure that only events containing the manifold have non-zero probability.

2.1 Manifold overfitting in likelihood-based models

Likelihood-based generative models are defined by a highly parameterized likelihood function $f(\boldsymbol{x}; \boldsymbol{\theta})$, whose parameters are trained by minimizing the loss

$$\mathcal{L}(\boldsymbol{\theta}) = -\mathbb{E}_{\boldsymbol{x} \sim p_0(\boldsymbol{x})}[f(\boldsymbol{x}; \boldsymbol{\theta})] \ , \tag{2}$$

which maximizes the probability of the data given the model. This maximum likelihood loss is minimized if $f(\boldsymbol{x}; \boldsymbol{\theta}) = p_0(\boldsymbol{x})$. Unfortunately, the divergence of $p_0(\boldsymbol{x})$ on the manifold implies that the model density $f(\boldsymbol{x}; \boldsymbol{\theta})$ on the manifold becomes larger and larger during training. More problematically, the optimization problem becomes almost insensitive to the internal density $\rho_{\text{int}}(\boldsymbol{x})$. This phenomenon is called *manifold overfitting* (Loaiza-Ganem et al., 2022), since the trained model fits the manifold while ignoring its internal density, resulting in poor generation. While generative diffusion models have often been characterized in terms of a likelihood function, experimental evidence suggests that they are not affected by this pathology. In the rest of the paper, we will provide an elegant explanation of their behavior by analyzing their latent geometry.

3 Background on generative diffusion models

Here, we will consider a simple variance-exploding forward process where the data $\boldsymbol{x}_0 \sim p_0(\boldsymbol{x})$ evolves according to the equation

$$\mathrm{d}\boldsymbol{x}_t = \mathrm{d}\boldsymbol{Z}_t \tag{3}$$

where $\mathrm{d}\boldsymbol{Z}_t$ is a standard Brownian motion. The formal solution of Eq. 3 can be given in term of the heat kernel: $p(\boldsymbol{x}_t, t) = \mathbb{E}_{\boldsymbol{x}_0 \sim p_0}\left[\frac{1}{\sqrt{2\pi t}} e^{-\frac{\|\boldsymbol{x}_t - \boldsymbol{x}_0\|_2^2}{2t}}\right]$. The *target distribution* $p_0(\boldsymbol{x})$ is then recovered by reversing the diffusion process (Anderson, 1982). We initialize this reverse process from $\boldsymbol{x}_{t_f} \sim \mathcal{N}(0, I_d)$, which evolves backward according to

$$\mathrm{d}\boldsymbol{x}_t = -\nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x}_t)\mathrm{d}t + d\boldsymbol{Z}_t \tag{4}$$

The function $s(\boldsymbol{x}, t) = \nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x})$ is is the so-called score function. From a set of training points $\{\boldsymbol{y}^1, \ldots, \boldsymbol{y}^N\} \overset{\text{iid}}{\sim} p_0$, we can train a neural approximation of $s(\boldsymbol{x}, t)$ by learning a denoising autoencoder $\hat{\boldsymbol{\epsilon}}_{\boldsymbol{\theta}}(\mathbf{x}, t)$, which is trained to recover the standardized noise ϵ_t from the noisy state $x_t = x_0 + t\epsilon_t$ (Hyvärinen and Dayan, 2005; Vincent, 2011; Ho et al., 2020). The learned score is then obtained using $\hat{s}_{\boldsymbol{\theta}}(\boldsymbol{x}, t) = -\frac{\hat{\boldsymbol{\epsilon}}_{\boldsymbol{\theta}}(\boldsymbol{x}, t)}{\sqrt{t}}$.

It is also convenient to define the support score $\tilde{s}_{\mathcal{M}}(\boldsymbol{x}, t)$, defined as the score function obtained from the uniform data distribution $p_0(\boldsymbol{x}) = \delta(1 - \mathcal{I}_{\mathcal{M}}(\boldsymbol{x}))$.

4 Dynamic latent manifolds and spectral gaps

Consider a generative diffusion model with $p_0(\boldsymbol{x}_0)$ defined on a d-dimensional manifold $\mathcal{M}_0$ according to Eq. 2. In the course of the diffusion process, we can define a time-dependent family of stable latent manifolds

$$\mathcal{M}_t = \{\boldsymbol{x}^* \mid \tilde{s}_{\mathcal{M}}(\boldsymbol{x}^*, t) = 0, \text{ with } J_{\mathcal{M}}(\boldsymbol{x}^*, t) \;\; \text{n.s.d.}\} \;, \tag{5}$$

where the negative semi-definiteness (n.s.d.) is a stability condition on the Jacobi matrix $J_{\mathcal{M}}(\boldsymbol{x}, t)$ of the support score $\tilde{s}_{\mathcal{M}}(\boldsymbol{x}^*, t)$. Due to the noise, the diffusing particles are likely to be found in shells of radius $\sqrt{t}$ around each point of the latent manifold.

For a small perturbation $\boldsymbol{p}$ around a point $\boldsymbol{x}^*$ on the latent manifold at time t, the score function is well approximated by its linearization:

$$s(\boldsymbol{p}, t) \approx J(\boldsymbol{x}^*, t)\; \boldsymbol{p} = -\sum_j (\boldsymbol{v}_j \cdot \boldsymbol{p})\, \lambda_j(\boldsymbol{x}^*, t) \boldsymbol{v}_j \quad , \tag{6}$$

where $J(\boldsymbol{x}^*, t)$ is the Jacobian of the score and the $\boldsymbol{v}_j$ and $\lambda_j(\boldsymbol{x}^*, t)$ are respectively the j-th eigenvector and the associated eigenvalue of $-J(\boldsymbol{x}^*, t)$. The spectrum of eigenvectors provides detailed information concerning the local geometry of the manifold around. Perturbations aligned to the tangent space of $\mathcal{M}_t$ correspond to small eigenvalues, while orthogonal perturbations correspond to high eigenvalues, as the score tends to push the stochastic dynamics towards its fixed-points. Therefore, we can estimate the dimensionality of the manifold from the location of a gap (i.e. a sharp change) in the sorted spectrum of eigenvalues (Stanczuk et al., 2022). This is visualized in Fig. 1.

4.1 Subspaces and intermediate gaps

Consider the situation where the internal density ρ_{int} is not locally flat around a point $\boldsymbol{x}^* \in \mathcal{M}_0$. In this case, at a finite time t the actual score function does not vanish on the latent manifold $\mathcal{M}_t$ as there is a gradient of log-density along the tangent directions. This implies that the spectrum of tangent eigenvalues can have a series of sub-gaps with separate different tangent subspaces with different 'local variance'. In image generation tasks, these subspaces are often associated with different frequency modes, as noted in (Kingma and Gao, 2023).

Consequently, we can quantify the sensitivity to the internal density at time t by studying the statistics and temporal evolution of intermediate gaps

$$\Delta_k^{\text{GAP}}(\boldsymbol{x}^*, t) = \lambda_{k+1}(\boldsymbol{x}^*, t) - \lambda_k(\boldsymbol{x}^*, t) \;, \tag{7}$$

where the indices k depend on the dimensionality of the subspaces. Note however that under realistic data distributions it is unlikely to find sharp intermediate discontinuities since each subspace will have a different eigenvalue, resulting in a smooth gradient.

5 Phenomenology of generative diffusion on manifolds

This section contains an intuitive picture that follows from our theoretical results on linear models, which we will fully outline in the next section. The theory considers the case where linear manifolds are sampled according to fixed-distributions, which account for the fact that the exact geometry of the manifold is not known in advance when training a diffusion model. While only linear models are theoretically tractable, we conjecture that their phenomenology captures the main features of subspace separation in the tangent space of curved manifolds. We validated the theory using networks trained on both linear data and highly non-linear data such as natural images (see section 8). Based on the dynamics of the spectral gaps, we found that the generative dynamics of $\boldsymbol{x}_t$ according to Eq. 4 can be separated into three distinct phases. These phases do not correspond to singularities and there are therefore cross-over events, not genuine phase transitions. During all our analysis we will exclusively work with *eigenvalues* since, in the linear manifold model, the Jacobian of the score is symmetric, hence invertible. The same phenomenology is nevertheless fully appreaciable when using the *singular values* in our experimental tests.

5.1 Phase I: The trivial phase

In the **trivial phase**, the diffusing particle moves according to the noise distribution without strong biases towards the manifold directions. In this dynamic regime, the latent manifold $\mathcal{M}_t$ is a single point surrounded by an isotropic quadratic well of potential. The spectral gaps are not visible and all eigenvalues have approximately the same value due to the isotropy of the noise distribution. This trivial phase is analogous to the initial phases described in (Raya and Ambrogioni, 2023) and (Biroli et al., 2024).

5.2 Phase II: Manifold coverage

The **manifold coverage phase** begins with the opening of the first of a series of spectral gaps corresponding to local subspaces. In this phase, different subspaces with different variances can therefore be identified by intermediate gaps in the spectra, as sketched in Fig. 2. When the intermediate gaps are opened, the diffusing particles spread across the relevant manifold directions according to their relative variances. In other words, during this regime of generative diffusion the process fits the distribution of the data internal to the manifold.

In term of random matrix theory, the gap-forming phenomenology provides for two distinct processes: the emergence of intermediate gaps (i.e. steps in the dimensionality plot) between separated bulks of the spectrum, the opening of a final gap that allows to infer the dimensionality of the full manifold. Our analysis has allowed us to define the time scale at which such intermediate gaps are maximally opened, i.e.

$$t_{\max}^{(k)} = \sqrt{\gamma_+(\sigma_k)\gamma_-(\sigma_{k+1})}, \tag{8}$$

where $\gamma_-(\sigma_{k+1})$ and $\gamma_+(\sigma_k)$ are specific eigenvalues in the spectrum of the projection matrix $F^\top F$ (see Fig. 9), associated to two hierarchically consecutive variances (see Appendix A.2 for an exhaustive analysis). In most of the cases, when $\sigma_{k+1}^2 \ll \sigma_k^2$, the dependence on the two variances is $\mathcal{O}(\sigma_k \cdot \sigma_{k+1})$. This is the timescale where the score is maximally sensitive to the relative variance of the two subspaces, which guides the particles toward the correct internal distribution.

5.3 Phase III: Manifold consolidation

Finally, the **manifold consolidation phase** is characterized by the asymptotic closure of the intermediate gaps and the sharpening of the total manifold gap, indicating the full dimensionality of $\mathcal{M}$. In this final regime, the score assumes the form

$$\nabla_{\mathbf{x}} \log p_t(\mathbf{x}) \simeq \frac{1}{t}\Big[\Pi - I_d\Big]\mathbf{x}. \tag{9}$$

where $\Pi = F(F^\top F)^{-1}F^\top$ is the projection matrix over the manifold. Without any posterior normalization, the component of the score orthogonal to the manifold diverges proportionally to t^{-1}, while the tangent components converge to a constant and become therefore negligible in this regime. This results in the consolidation of the gap corresponding to the manifold dimensionality m and to the (relative) closure of the intermediate gaps. Therefore, in this final phase the dynamics of the model simply projects the particles into the manifold $\mathcal{M}_t \to \mathcal{M}_0$. In the generative modeling literature, this phenomenon is also known as *manifold overfitting* as the terms corresponding the the internal distribution is negligible (Loaiza-Ganem et al., 2022). Fortunately, this internal coverage occurs in the model throughout intermediate phases across the previous diffusion process, resulting in a balanced coverage of the internal distribution.

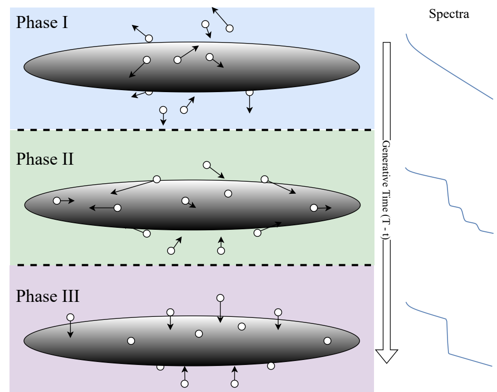


Figure 2: Visualization of the trivial (I), coverage (II) and consolidation phase (III).

6 Theoretical analysis of the spectral of linear diffusion models

In this section, we provide our main theoretical results concerning the spectral distribution for random linear subspaces and the relative spectral gaps formulas. We start by reviewing diffusion with data supported on linear manifolds, where the exact score function can be computed.

6.1 Linear manifolds

Normally the distribution $p_0(\boldsymbol{x})$ is unknown. It is however interesting to investigate tractable special case where the distribution is a multivariate Gaussian defined on a linear manifold:

$$\boldsymbol{y}^\mu = \frac{1}{\sqrt{m}} F \boldsymbol{z}^\mu \tag{10}$$

where $F \in \mathbb{R}^{d \times m}$ is an arbitrary projection matrix that implicitly define the structure of the latent manifold. and $\boldsymbol{z}^\mu \sim \mathcal{N}(0, I_m)$ the latent space vector. In this setting, the distribution can be explicitly written as

$$p_0(\mathbf{x}) = \int d\mathbf{z}\ \delta(\mathbf{x} - \frac{1}{\sqrt{m}} F\mathbf{z}). \tag{11}$$

Therefore, the density of the process at a given time t takes the form

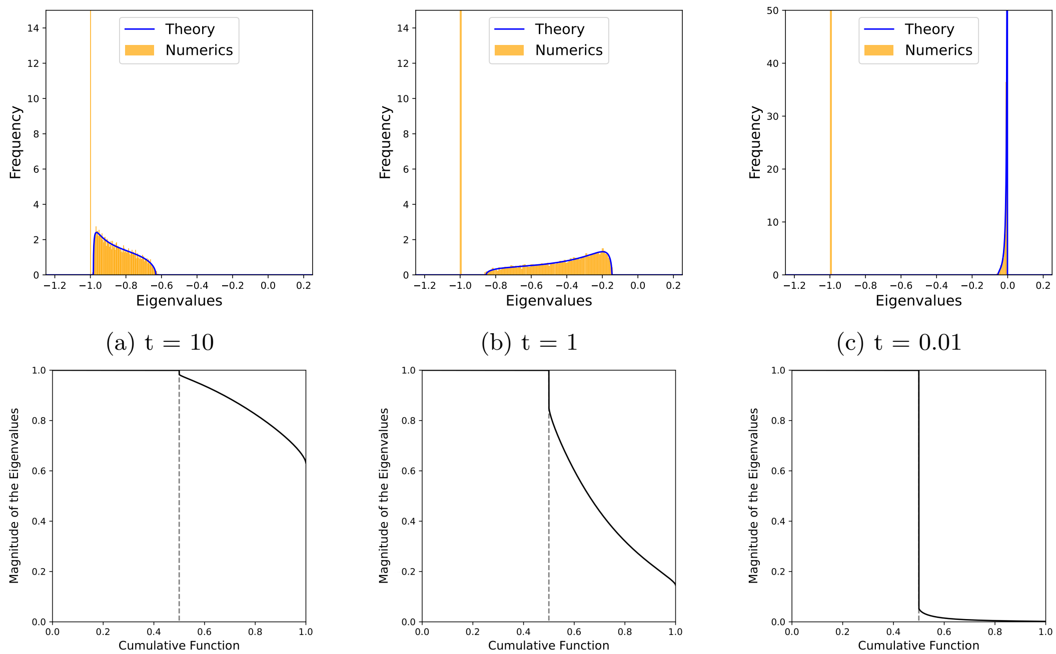
$$p_t(\boldsymbol{x}) = \int D\mathbf{z} \frac{1}{\sqrt{2\pi t}^d}\ e^{-\frac{1}{2t}\|\mathbf{x} - \frac{1}{\sqrt{m}} F\mathbf{z}\|^2}. , \tag{12}$$

where $\int D\mathbf{z}$ denotes standard Gaussian integration in $\mathbb{R}^m$.

While linear manifolds are very simple when compared with real data, they still exhibit a rich and non-trivial phenomenology that elucidate several universal phenomena of diffusion under the manifold hypothesis. In fact, these linear models capture the structure of tangent spaces of smooth manifolds.

The score function of the linear model is solvable analytically, since we only have to perform Gaussian integrals, from which we obtain a quadratic form in $\mathbf{x}$ that we can rewrite as

$$\log p_t(\mathbf{x}) = \frac{1}{2t} \mathbf{x}^\top J_t \mathbf{x} + const. \tag{13}$$



(a) t = 10 (b) t = 1 (c) t = 0.01

(d) t = 10 (e) t = 1 (f) t = 0.01

Figure 3: Spectrum of the eigenvalues of J_t and drop in the dimensionality of the data-manifold estimated from theory in the single-variance case, with $\alpha_m = 0.5, \sigma^2 = 1$. Numerical data are generated with $d = 100$ and collected over 100 realizations of the F matrix.

where the constant does not depend on $\mathbf{x}$ and

$$J_t = \frac{1}{t} F \left[I_m + \frac{1}{t} F^\top F \right]^{-1} F^\top - I_d. \tag{14}$$

The score function is thus derived as $\nabla_{\mathbf{x}} \log p_t(\mathbf{x}) = \frac{1}{t} J_t \mathbf{x}$. It is then useful to analyze the spectrum of the matrix J_t, since J_t is proportional to the Jacobian of the score function. Indeed, since the gradient of the score is orthogonal to the manifold sufficiently close to it, the number of null eigenvalues of J_t will correspond to the manifold dimension and we should expect to see a drop in the spectrum.

6.2 The linear random-manifold score in large dimensions: A random matrix analysis

In the following, we provide an outline of our theoretical results on the distribution of spectral gaps in the matrix J_t under random linear manifolds. This choice reflects the fact that the distribution and support of the data is usually not known in advance, and it is therefore important to quantify the statistical variability induced by this uncertainty. Specifically, we will consider random projection matrices F distributed according to different distributions. To ensure tractability, we perform the analysis in the limit of large d (visible) and m (latent) dimensions while keeping the ratio $\alpha_m = m/d$ constant.

6.3 The isotropic case

If $F \sim \mathcal{N}(0, \sigma^2)$ we are able to derive analytically the full expression of the distribution of the eigenvalues of J_t, as it is a simple transformation of the distribution of the eigenvalues of $F^\top F$, which is known to be the Marchenko-Pastur distribution reported in Appendix A.1. As observable from Fig. 3, which reports the shape of the spectrum in time, the bulk of the distribution, inherited from the density of the eigenvalues of $F^\top F$, gradually shifts from left to right in the support. By measuring the cumulative function of the spectrum, one can isolate a drop in the effective dimensionality of the manifold, as also plotted in Fig. 3. The step is present at any time in the process and it is implied by the gap between the left bound of the bulk and the spike in -1. The width of this gap evolves in time according to

$$\Delta_{\text{fin}}^{\text{GAP}}(t;\sigma) = \frac{\sigma^2(1+\alpha_m^{-1/2})^2}{t+\sigma^2(1+\alpha_m^{-1/2})^2}. \tag{15}$$

If we name $\gamma_+(\sigma)$ the left bound eigenvalue of the bulk in the spectrum of $F^\top F$ (see Fig. 8), one can recover a more general expression for the gap, being

$$\frac{\gamma_+(\sigma)}{t+\gamma_+(\sigma)} = \Delta. \tag{16}$$

Hence we can resolve the gap at a scale Δ at time

$$t_{\text{in}} = \gamma_+(\sigma)\left(\frac{1-\Delta}{\Delta}\right). \tag{17}$$

6.4 Intermediate gaps and subspaces with different variances

Another relevant case for our study is the one where we consider a manifold having multiple subspaces with different variances. We will here focus on the instance of two distinct variances. This scenario is reproduced by considering a number $f \cdot m$ of columns of F to be Gaussian distributed with zero mean and a variance σ_1^2, and the remaining $(1-f)\cdot m$ columns with variance σ_2^2. The spectrum of J_t can be computed also in this case, as explained in Appendix A.2. The density function of the eigenvalues shows a transient behaviour of the spectrum in the form of an intermediate drop in the estimated dimensionality of the hidden data manifold. This behaviour is reported in Fig. 4. Even though the expression of the spectral density has a not an explicit analytical form, but has to be computed numerically, one can adopt a special assumption on the behaviour of the density of the eigenvalues of $F^\top F$ to estimate the typical times at which the intermediate drop occurs. Generally speaking, the spectrum of $F^\top F$ can be composed by two separated bulks, as observable in Fig. (9). This happens when σ_2^2 and σ_1^2 are significantly different. In analogy with the single variance scenario, we name γ_+ the left bound of the bulk associated to higher eigenvalues, i.e. with the higher variance, and γ_- the right bound of the bulk associated with smaller eigenvalues, i.e. smaller variance. Most commonly, $\gamma_- = \gamma_-(\sigma_2)$ and $\gamma_+ = \gamma_+(\sigma_1)$. In this case the gap-forming phenomenology provides for two distinct processes: the emergence of intermediate gaps (i.e. steps in the dimensionality plot) between speareated bulks of the spectrum, the opening of a final gap that allows to infer the dimensionality of the full manifold. The width of the intermediate gap between two bulks can be obtained from Eq. (27) as

$$\Delta_{\text{inter}}^{\text{GAP}}(t;\sigma_1,\sigma_2) = \frac{t}{t+\gamma_-(\sigma_2)} - \frac{t}{t+\gamma_+(\sigma_1)}. \tag{18}$$

By imposing $\Delta_{\text{inter}}^{\text{GAP}}(t;\sigma_1,\sigma_2) = \Delta$ one finds the following quadratic form

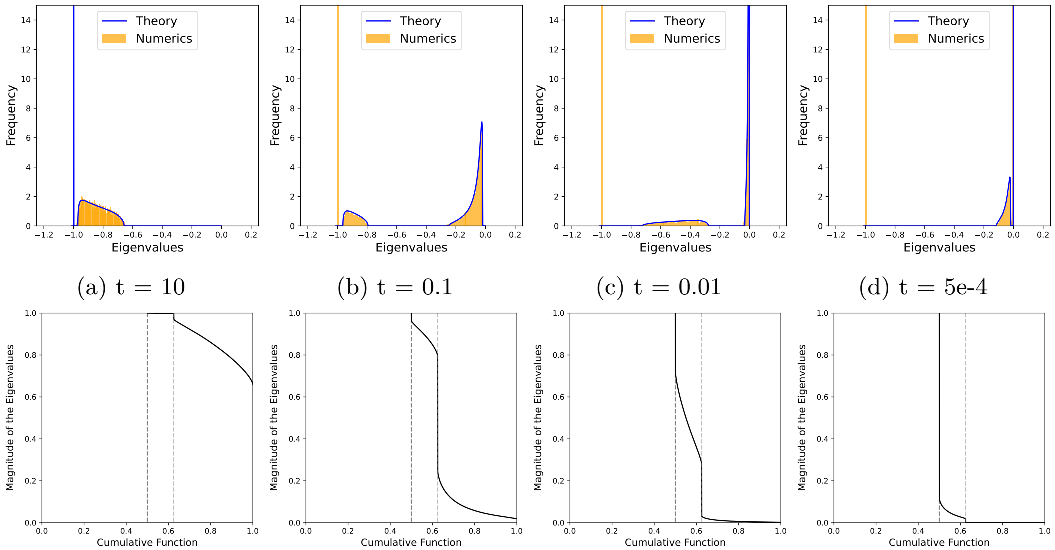
$$\Delta t^2 + \big[(\Delta-1)\gamma_- + (\Delta+1)\gamma_+\big]t + \Delta\gamma_-\gamma_+ = 0. \tag{19}$$

Considering $\Delta \ll 1$ and $\gamma_+ \ll \gamma_-$ the opening time for the intermediate gap can be found by

$$t_{\text{in}}(\Delta) \simeq \Delta^{-1}\gamma_-(\sigma_1), \tag{20}$$

that is a reference time at which the gap becomes visible. On the other hand, by assuming the closure time to be close to zero, it can be obtained as

$$t_{\text{fin}}(\Delta) \simeq \Delta\gamma_+(\sigma_2). \tag{21}$$



(a) t = 10 (b) t = 0.1 (c) t = 0.01 (d) t = 5e-4

(e) t = 10 (f) t = 0.1 (g) t = 0.01 (h) t = 5e-4

Figure 4: Spectrum of the eigenvalues of J_t and drop in the dimensionality of the data-manifold estimated from theory in the double-variance case, with $\alpha_m = 0.5, \sigma_1^2 = 1, \sigma_2^2 = 0.01$, $f = 0.75$. Numerical data are generated with $d = 100$ and collected over 100 realizations of the F matrix.

Furthermore, the time at which the gap is maximum in width, and so maximally visible, is located in between t_{in} and t_{fin}. This is the most important time scale for the problem, it is obtained by imposing $\partial\Delta^{\text{GAP}}/\partial t = 0$ and it measures

$$t_{\max} = \sqrt{\gamma_-(\sigma_1)\gamma_+(\sigma_2)}. \tag{22}$$

Indeed, when $\sigma_1^2 \gg \sigma_2^2$ the total spectrum can be approximated by a mixture of two separated Marchenko-Pastur distributions, with variances σ_1^2 and σ_2^2, and parameters α_m and γ to be rescaled with respect to f and $(1-f)$. This approximation becomes exact under a slight modification of F which does not imply any loss of the quality of the description. Now the relevant quantities for the gap become

$$\Delta^{\text{GAP}}_{\text{inter}}(t;\sigma_1,\sigma_2) = \frac{t\left[f\sigma_1^2(1-\sqrt{\frac{1}{f\alpha_m}})^2 - (1-f)\sigma_2^2(1+\sqrt{\frac{1}{(1-f)\alpha_m}})^2\right]}{\left[t+(1-f)\sigma_2^2(1+\sqrt{\frac{1}{(1-f)\alpha_m}})^2\right]\left[t+f\sigma_1^2(1-\sqrt{\frac{1}{f\alpha_m}})^2\right]} \tag{23}$$

$$t_{\text{in}}(\Delta) \simeq \Delta^{-1} f\left(1-\sqrt{\frac{1}{f\alpha_m}}\right)^2 \sigma_1^2, \tag{24}$$

$$t_{\text{fin}}(\Delta) \simeq \Delta(1-f)\left(1+\sqrt{\frac{1}{(1-f)\alpha_m}}\right)^2 \sigma_2^2. \tag{25}$$

$$t_{\max} = \sqrt{f(1-f)}\left(1-\sqrt{\frac{1}{f\alpha_m}}\right)\left(1+\sqrt{\frac{1}{(1-f)\alpha_m}}\right)\sigma_1\cdot\sigma_2. \tag{26}$$

This same analysis can be extended to the more general case where the spectral density is known to be formed by different detached bulks, associated to hierarchically smaller variances of the data. The evolution of the intermediate gaps in a double-variance diffusion model is reported in Fig. 4: notice that $t_{\max} = \mathcal{O}(\sigma_2)$ is consistent with Fig. 4b and 4f, where the gap was found to be maximum in width. It is worth noting that subspaces with higher variances are the first ones to be explored by diffusion, and to be learned by the model. This point suggests that the model is sensitive to the parameters of the probability distribution on the manifold as recently suggested by other works in the literature (see section 9 for further details).

7 Experiments with synthetic linear datasets

We first measure the spectrum of the singular values of the Jacobian of a score function trained through a neural network on a linear manifold data-model generated by two variances σ_1^2, σ_2^2. Results are reported in Figure 5 (left and central panels) for one arbitrary choice of the fraction f. The opening of the gaps is consistent with the theory for the exact score: an intermediate gap associated to the subspace with higher variance first opens; subsequently, the gap relative to the lower variance subspace, which here corresponds to the final gap, opens. We can infer the dimensions of the subspaces by subtracting the dimension signed by the dashed line to d. We underline the fact that higher-variance subspaces are learned first by repeating the experiment after swapping the values of the variances. Eventually, the right panel in Figure 5 reports the same experiments where variances are uniformly generated in the interval $[10^{-3}, 1]$: it is evident that the d intermediate gaps is now a continuous line, as it is expected to be in more realistic natural data-sets.

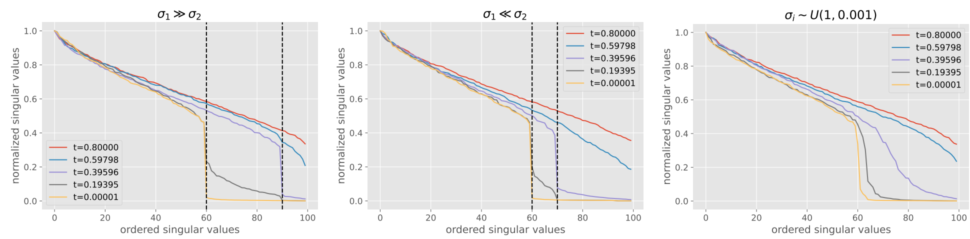


Figure 5: Ordered singular values obtained with the trained score model, for difference variances on the subspaces. Data are generated according to the linear-manifold model with $d = 100$ and $m = 40$. Left: $\sigma_1^2 = 1$, $\sigma_2^2 = 0.01$, $f = 0.75$; Center: $\sigma_1^2 = 0.01$, $\sigma_2^2 = 1$, $f = 0.75$; Right: variances sampled uniformly between 10^{-3} and 1.

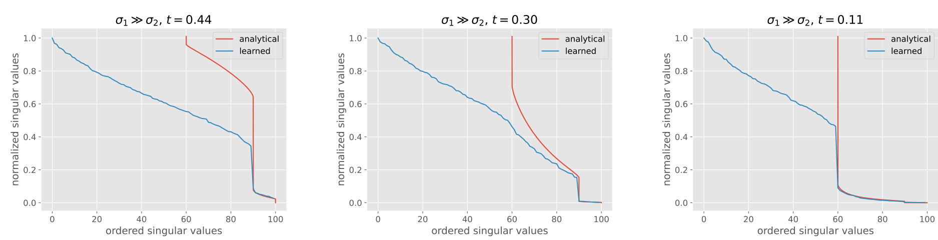


Figure 6: Comparison between spectra obtained with the trained score model and with the numerical analysis, for difference variances on the subspaces. Data are generated according to the linear-manifold model with $d = 100$ and $m = 40$, $\sigma_1^2 = 1$, $\sigma_2^2 = 0.01$, $f = 0.75$; from left to right, the spectra are evaluated a time $t \approx 0.45$, $t \approx 0.3$, $t \approx 0.11$.

We will now compare the gaps computed analytically with ones obtained from real neural networks trained on the same linear data-model. The results of such comparison are presented in figure 6, and they show a decent agreement between the ordered distribution of the singular values obtained through empirical methods, and the relative analytical counterpart, computed through the replica method. The opening of the predicted intermediate gaps signal the right dimension of the linear subspaces as verified from the experiments. One can notice from the figure that the analytical profile shows the shape of a sharp step between the zero value along the x-axis and the first appearing gap: this shape is related to the Dirac-delta spike that the spectrum of the eigenvalues presents at $r = -1$ (see Appendix A.2 for details about the spectrum); on the other hand, the numerical profile looks different in the same region, and this behaviour is associated to the absence of the spike in the distribution of the singular values, that leaves room to a separated bulk from the other ones. This evident discrepancy between theory and experiment is probably due to the final configuration of the trained neural network and leaves space for further investigations.

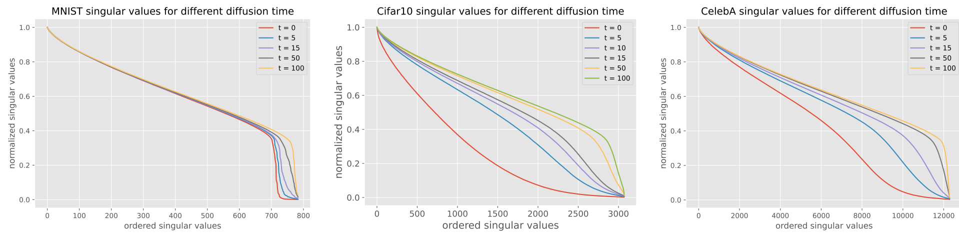


Figure 7: Jacobian spectra of diffusion models trained on MNIST, Cifar10 and CelebA.

8 Experiments with natural image datasets

While our theoretical analysis is limited to linear random-manifold models, several qualitative figures of its phenomenology can be observed in networks trained in natural images. Fig. 7 shows the temputal evolution of the spectrum estimated numerically from the Jacobian of models trained on MNIST, Cifar10 and CelebA and averaged over 100 images. The details of this experiment are given in Supp. C and Supp. B. We see that for low times, the network trained on MNIST shows a sharp total manifold gap at around the 700-th singular value. Consistently with our theory, for larger time the spectrum becomes smoother. The non averaged result obtained around single images also reveal the presence of several discernible small intermediate gaps (see left panel in Fig. 7). The results on Cifar10 and CelebA are more difficult to interpret, probably due to a less well-defined manifold and subspace structure. However, the inflection points of the spectra reveal a more suble gap structure, which recedes for small values of t as more local subspaces are revealed.

9 Related work & Discussion

The evolution of the fixed-points of the exact score was studied in (Raya and Ambrogioni, 2023) for the analysis spontaneous symmetry breaking and in (Biroli and Mézard, 2023) for the analysis of memorization and disordered phase transitions. The use of spectral gaps to quantify the dimensionality of the manifold was introduced in (Stanczuk et al., 2022). This work was concerned with the total manifold gap and did not investigate sub-gaps and temporal dynamics. Several recent studies investigated the local linear structure of trained diffusion models. For example, (Kadkhodaie et al., 2024) studied the expansion of the Jacobian of trained models and described it as an optimal geometry-adaptive Harmonic representation. Similarly, (Chen et al., 2024b) characterized the linear expansion of the Jacobian of trained networks and characterized the resulting components in terms of their frequency content. Our work can be seen as a theoretical complement with this more applied line of research, as we provide a comprehensive random-matrix analysis of the phenomenon in tractable models. The dynamic geometry of diffusion manifolds were also investigated in (Chen et al., 2024a) using techniques inspired by research on latent generators such as GANs and autoencoders. Another recent work (Wang et al., 2024) links the underlying structure of real data to the generalization abilities of diffusion models, showing an equivalence to subspace clustering. Finally, (Sakamoto et al., 2024) studies the dynamic geometry of tubular neighborhoods of the latent manifold and connect their singularities with spontaneous symmetry breaking events. Generative diffusion models exhibit rich geometric structures that have the potential to explain their impressive generative capabilities. Our work introduces the use of random-matrix theory techniques for the analysis of their dynamic local geometry and gives room to the use of advanced statistical physics techniques, which may in the future unveil more global, topological and non-linear aspects of the dynamic geometry of diffusion generative models.

References

Anderson, B. D. (1982). Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3):313–326.

Biroli, G., Bonnaire, T., de Bortoli, V., and Mézard, M. (2024). Dynamical regimes of diffusion models.

Biroli, G. and Mézard, M. (2023). Generative diffusion in very large dimensions. *Journal of Statistical Mechanics: Theory and Experiment*, 2023(9):093402.

Blattmann, A., Rombach, R., Ling, H., Dockhorn, T., Kim, S. W., Fidler, S., and Kreis, K. (2023). Align your latents: High-resolution video synthesis with latent diffusion models. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Brooks, T., Peebles, B., Holmes, C., DePue, W., Guo, Y., Jing, L., Schnurr, D., Taylor, J., Luhman, T., Luhman, E., Ng, C., Wang, R., and Ramesh, A. (2024). Video generation models as world simulators.

Chen, J., Li, M., Pan, Z., Gao, X., and Tu, C. (2024a). Varying manifolds in diffusion: From time-varying geometries to visual saliency. *arXiv preprint arXiv:2406.18588*.

Chen, S., Zhang, H., Guo, M., Lu, Y., Wang, P., and Qu, Q. (2024b). Exploring low-dimensional subspaces in diffusion models for controllable image editing. *arXiv preprint arXiv:2409.02374*.

Fefferman, C., Mitter, S., and Narayanan, H. (2016). Testing the manifold hypothesis. *Journal of the American Mathematical Society*, 29(4):983–1049.

Ho, J., Jain, A., and Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*.

Ho, J., Salimans, T., Gritsenko, A., Chan, W., Norouzi, M., and Fleet, D. J. (2022). Video diffusion models. *arXiv preprint arXiv:2204.03458*.

Hyvärinen, A. and Dayan, P. (2005). Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6(4).

Kadkhodaie, Z., Guth, F., Simoncelli, E. P., and Mallat, S. (2024). Generalization in diffusion models arises from geometry-adaptive harmonic representations. In *The Twelfth International Conference on Learning Representations*.

Kingma, D. P. and Gao, R. (2023). Understanding diffusion objectives as the elbo with simple data augmentation.

Loaiza-Ganem, G., Ross, B. L., Cresswell, J. C., and Caterini, A. L. (2022). Diagnosing and fixing manifold overfitting in deep generative models. *Transactions on Machine Learning Research*.

Peyré, G. (2009). Manifold models for signals and images. *Computer vision and image understanding*, 113(2):249–260.

Raya, G. and Ambrogioni, L. (2023). Spontaneous symmetry breaking in generative diffusion models. arXiv:2305.19693 [cs].

Sakamoto, K., Sakamoto, R., Tanabe, M., Akagawa, M., Hayashi, Y., Yaguchi, M., Suzuki, M., and Matsuo, Y. (2024). The geometry of diffusion models: Tubular neighbourhoods and singularities. In *ICML 2024 Workshop on Geometry-grounded Representation Learning and Generative Modeling*.

Salimans, T., Karpathy, A., Chen, X., and Kingma, D. P. (2017). Pixelcnn++: Improving the pixelcnn with discretized logistic mixture likelihood and other modifications. *arXiv preprint arXiv:1701.05517*.

Singer, U., Polyak, A., Hayes, T., Yin, X., An, J., Zhang, S., Hu, Q., Yang, H., Ashual, O., Gafni, O., et al. (2022). Make-a-video: Text-to-video generation without text-video data. *arXiv preprint arXiv:2209.14792*.

Song, Y. and Ermon, S. (2019). Generative modeling by estimating gradients of the data distribution. *Advances in Neural Information Processing Systems*.

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. (2021). Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*.

Stanczuk, J., Batzolis, G., Deveney, T., and Schönlieb, C.-B. (2022). Your diffusion model secretly knows the dimension of the data manifold. *arXiv preprint arXiv:2212.12611*.

Vincent, P. (2011). A connection between score matching and denoising autoencoders. *Neural computation*, 23(7):1661–1674.

Wang, P., Zhang, H., Zhang, Z., Chen, S., Ma, Y., and Qu, Q. (2024). Diffusion models learn low-dimensional distributions via subspace clustering.

A ANALYTICAL DERIVATION OF THE SPECTRUM OF J_t

A.1 SINGLE VARIANCE SCENARIO

We want to compute the spectrum of the matrix in 14. Let us first consider the case in which F is a $d \times m$ matrix with Gaussian entries, and call γ the eigenvalues of $FF^\top$.

The function that gives the eigenvalues r of J_t as function of γ is

$$r_j = \frac{1}{t}\frac{\gamma_j}{1+\frac{1}{t}\gamma_j} - 1 = -\frac{t}{t+\gamma_j} \tag{27}$$

Thus, knowing that the distribution of γ is Marchenko-Pastur, we can obtain the distribution of r

$$\rho_t(r) = -\frac{\alpha_m}{2\pi}\frac{1}{r(1+r)}\sqrt{(r_+ - r)(r - r_-)} + (1-\alpha_m)\,\delta(r+1)\,\theta\left(\alpha_m^{-1} - 1\right) \tag{28}$$

for $r \in [r_-(t), r_+(t)]$, with $r_\pm(t) = -\frac{t}{\left(1 \pm \frac{1}{\sqrt{\alpha_m}}\right)^2 + t}$.

One could ask whether the bulk of J_t separates from $r = -1$ at a discrete time. This separation corresponds to a drop in the histogram of eigenvalues. According to Eq. (28), the bulk is always separated from the spike at finite time t, because $(1+\alpha_m^{-1/2})^2 + t$ for every t, and the width of the gap is given by $\Delta^{\text{GAP}}(t) = r_-(t) + 1$

$$\Delta^{\text{GAP}}(t) = \frac{(1+\alpha_m^{-1/2})^2}{t + (1+\alpha_m^{-1/2})^2}. \tag{29}$$

With respect to the starting spectrum of $F^\top F$, this condition reads

$$\frac{\gamma_+}{t+\gamma_+} = \Delta \tag{30}$$

so the time when we see the drop at a scale Δ is $t = \gamma_+^2 \frac{(1-\Delta)}{\Delta}$.

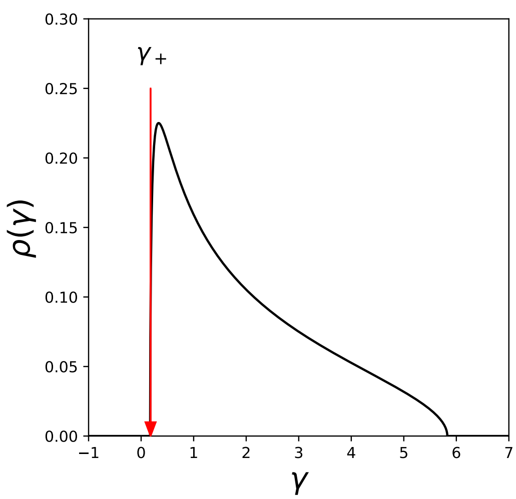


Figure 8: Spectrum of the eigenvalues of $F^\top F$ as obtained from random matrix theory with eigenvalue γ_+ indicated by red arrow. γ_+ is provided by the Marchenko-Pastur density function. Control parameters are chosen to be $\alpha_m = 0.5$, $\sigma^2 = 1$.

A.2 Double variance scenario

We want to compute the spectrum of J_t when $F_{i\mu} \sim \mathcal{N}(0, \sigma_1^2)$ for $\mu < fm/2$ and $F_{i\mu} \sim \mathcal{N}(0, \sigma_2^2)$ for $\mu > (1-f)m/2$, with $f \in [0,1]$. We use the replica method to compute the spectrum of $A = \frac{1}{m}FF^\top$, then with a transform we obtain the spectrum of J_t. In order to obtain the spectrum we need to compute the expectation of the resolvent of A in the $d \to +\infty$ limit, and to do this we will rely on the replica method

$$\mathbb{E}\left[g_A(z)\right] = -\frac{2}{d}\frac{\partial}{\partial z}\mathbb{E}\left[\log\frac{1}{\sqrt{\det\left(zI_d - A\right)}}\right] \tag{31}$$

$$= -\frac{2}{d}\frac{\partial}{\partial z}\lim_{n\to 0}\mathbb{E}\left[\frac{Z^n - 1}{n}\right] \tag{32}$$

with

$$Z^n = \det\left(zI_d - A\right)^{-n/2} \tag{33}$$

$$= \int\prod_{a=1}^{n}\prod_{i=1}^{d}\frac{d\phi_i^a}{\sqrt{2\pi}}\, e^{-\frac{1}{2}\sum_{a=1}^{n}\sum_{i,j=1}^{d}\phi_i^a\left(z\delta_{ij} - \frac{1}{m}\sum_\mu F_{i\mu}F_{j\mu}\right)\phi_j^a} \tag{34}$$

and taking the expectation

$$\mathbb{E}\left[Z^n\right] = \int\prod_{a,i}\frac{d\phi_i^a}{\sqrt{2\pi}}\, e^{-\frac{z}{2}\sum_a\sum_i(\phi_i^a)^2}\mathbb{E}\left[e^{\frac{1}{2m}\sum_a\sum_\mu\left(\sum_i\phi_i^a F_{i\mu}\right)^2}\right] \tag{35}$$

$$= \int\prod_{a,\mu}\frac{d\eta_\mu^a}{\sqrt{2\pi}}\, e^{-\frac{1}{2}\sum_a\sum_\mu(\eta_\mu^a)^2}\int\prod_{a,i}\frac{d\phi_i^a}{\sqrt{2\pi}}\, e^{-\frac{z}{2}\sum_a\sum_i(\phi_i^a)^2}\prod_\mu\mathbb{E}\left[e^{\frac{1}{\sqrt{m}}\sum_a\left(\sum_i\phi_i^a F_{i\mu}\right)\eta_\mu^a}\right] \tag{36}$$

where in the last step we have used the independence of the rows of F and applied a Hubbard-Stratonovic transform.

We can separate the product over μ and integrate over the distribution of F

$$\mathbb{E}\left[Z^n\right] = \int\prod_{a,\mu}\frac{d\eta_\mu^a}{\sqrt{2\pi}}e^{-\frac{1}{2}\sum_{a,\mu}(\eta_\mu^a)^2}\int\prod_{a,i}\frac{d\phi_i^a}{\sqrt{2\pi}}e^{-\frac{z}{2}\sum_{a,i}(\phi_i^a)^2} \tag{37}$$

$$\times\prod_{\mu=1}^{fm-1}\mathbb{E}\left[e^{\frac{1}{2\sqrt{m}}\sum_a\left(\sum_i\phi_i^a F_{i\mu}\right)\eta_\mu^a}\right]\prod_{\mu=fm}^{m}\mathbb{E}\left[e^{\frac{1}{2\sqrt{m}}\sum_a\left(\sum_i\phi_i^a F_{i\mu}\right)\eta_\mu^a}\right] \tag{38}$$

$$= \int\prod_{a,\mu}\frac{d\eta_\mu^a}{\sqrt{2\pi}}\, e^{-\frac{1}{2}\sum_{a,\mu}(\eta_\mu^a)^2}\int\prod_{a,i}\frac{d\phi_i^a}{\sqrt{2\pi}}e^{-\frac{z}{2}\sum_{a,i}(\phi_i^a)^2} \tag{39}$$

$$\times\, e^{\frac{\sigma_1^2}{2m}\sum_i\sum_{\mu<fm}\left(\sum_a\phi_i^a\eta_\mu^a\right)^2 + \frac{\sigma_2^2}{2m}\sum_i\sum_{\mu\geq fm}\left(\sum_a\phi_i^a\eta_\mu^a\right)^2} \tag{40}$$

$$= \int\prod_{a,\mu}\frac{d\eta_\mu^a}{\sqrt{2\pi}}\, e^{-\frac{1}{2}\sum_{a,\mu}(\eta_\mu^a)^2}\ \int\prod_{a,i}\frac{d\phi_i^a}{\sqrt{2\pi}}\, e^{-\frac{z}{2}\sum_{a,i}(\phi_i^a)^2}) \tag{41}$$

$$\times\, e^{\frac{\sigma_1^2}{2m}\sum_{ab}\left(\sum_i\phi_i^a\phi_i^b\right)\left(\sum_{\mu<fm}\eta_\mu^a\eta_\mu^b\right) + \frac{\sigma_2^2}{2m}\sum_{ab}\left(\sum_i\phi_i^a\phi_i^b\right)\left(\sum_{\mu>fm}\eta_\mu^a\eta_\mu^b\right)}. \tag{42}$$

Introducing $q_{ab} = \frac{1}{d}\sum_i\phi_i^a\phi_i^b$

$$\mathbb{E}\left[Z^n\right] = \int \prod_{a,b} \frac{dq_{ab} d\hat{q}_{ab}}{2\pi} \int \prod_{ai} \frac{d\phi_i^a}{\sqrt{2\pi}}\, e^{-\sum_{ab} \frac{1}{2}\hat{q}_{ab}\left(dq_{ab} - \sum_i \phi_i^a \phi_i^b\right) - \frac{z}{2}\sum_a \sum_i (\phi_i^a)^2} \tag{43}$$

$$\times \left[\int \prod_{a=1}^{n} \frac{d\eta^a}{\sqrt{2\pi}}\, e^{-\frac{1}{2}\sum_a (\eta^a)^2 + \frac{\sigma_1^2}{2\alpha_m}\sum_{ab} q_{ab}\eta^a\eta^b}\right]^{fm} \tag{44}$$

$$\times \left[\int \prod_{a=1}^{n} \frac{d\eta^a}{\sqrt{2\pi}}\, e^{-\frac{1}{2}\sum_a (\eta^a)^2 + \frac{\sigma_2^2}{2\alpha_m}\sum_{ab} q_{ab}\eta^a\eta^b}\right]^{(1-f)m} \tag{45}$$

$$= \int \prod_{a,b} \frac{dq_{ab} d\hat{q}_{ab}}{2\pi}\, e^{nd\Phi(q,\hat{q})} \tag{46}$$

with

$$\Phi(q,\hat{q}) = -\frac{1}{2n}\sum_{a,b} q_{ab}\hat{q}_{ab} + G_S(\hat{q}) + f\alpha_m G_E(q,\sigma_1) + (1-f)\alpha_m G_E(q,\sigma_2) \tag{47}$$

where

$$G_S(\hat{q}) = \frac{1}{n}\log \int \prod_{a=1}^{n} \frac{d\phi^a}{\sqrt{2\pi}}\, e^{-\frac{z}{2}\sum_a (\phi^a)^2 + \frac{1}{2}\sum_{ab}\hat{q}_{ab}\phi^a\phi^b} \tag{48}$$

$$G_E(q,\sigma) = \frac{1}{n}\log \int \prod_{a=1}^{n} \frac{d\eta^a}{\sqrt{2\pi}}\, e^{-\frac{1}{2}\sum_a (\eta^a)^2 + \frac{\sigma^2}{2\alpha_m}\sum_{ab} q_{ab}\eta^a\eta^b} \tag{49}$$

Using the replica symmetric ansatz $q_{ab} = \delta_{ab} q$, $\hat{q}_{ab} = -\delta_{ab}\hat{q}$

$$G_S(\hat{q}) = -\frac{1}{2}\log(z+\hat{q}) \tag{50}$$

$$G_E(q,\sigma) = -\frac{1}{2}\log(1 - \frac{\sigma^2 q}{\alpha_m}). \tag{51}$$

Putting all together we have

$$\Phi(z) = \frac{1}{2}\hat{q}q - \frac{1}{2}\log(z+\hat{q}) - f\frac{\alpha_m}{2}\log\left(1 - \frac{\sigma_1^2 q}{\alpha_m}\right) - (1-f)\frac{\alpha_m}{2}\log\left(1 - \frac{\sigma_2^2 q}{\alpha_m}\right). \tag{52}$$

The integral can be evaluated by the saddle point method

$$q = \frac{1}{z+\hat{q}} \tag{53}$$

$$\hat{q} = -f\frac{\alpha_m\sigma_1^2}{\alpha_m - \sigma_1^2 q} - (1-f)\frac{\alpha_m\sigma_2^2}{\alpha_m - \sigma_2^2 q}. \tag{54}$$

We can find the Stieltjes transform

$$\mathbb{E}[g_A(z)] = -2\alpha_m \frac{\partial \Phi(z)}{\partial z} \tag{55}$$

$$= \alpha_m q^*(z) \tag{56}$$

where q^* is found by solving the saddle point equation

$$zq^3 + q^2\left(\alpha_m - 1 - \frac{z\alpha_m}{\sigma_1^2} - \frac{z\alpha_m}{\sigma_2^2}\right) + \\ + q\left(\frac{\alpha_m^2}{\sigma_1^2\sigma_2^2}(z - f\sigma_1^2 - (1-f)\sigma_2^2) + \frac{\alpha_m}{\sigma_1^2} + \frac{\alpha_m}{\sigma_2^2}\right) - \frac{\alpha_m^2}{\sigma_1^2\sigma_2^2} = 0. \quad (57)$$

The asymptotic distribution of eigenvalues can be obtained from the Stieltjes transform as

$$\rho(\gamma) = \frac{1}{\pi}\lim_{\epsilon\to 0^+} \mathrm{Im}\left(g_A(\gamma - i\epsilon)\right) \quad (58)$$

$$= \frac{1}{\pi}\lim_{\epsilon\to 0^+} \mathrm{Im}\Phi'(\gamma - i\epsilon) \quad (59)$$

$$= \frac{1}{\pi}\alpha_m \lim_{\epsilon\to 0^+} \mathrm{Im}[q^*] \quad (60)$$

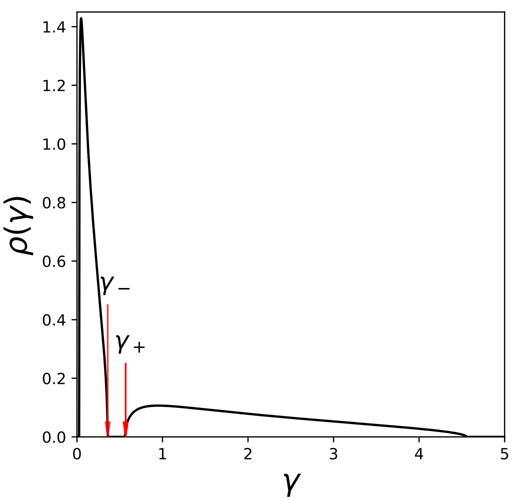


Figure 9: Spectrum of the eigenvalues of $F^\top F$ as obtained from random matrix theory with eigenvalues γ_- and γ_+ indicated by red arrows. Control parameters are chosen to be $\alpha_m = 0.5$, $f = 0.5$, $\sigma_1^2 = 1, \sigma_2^2 = 0.1$.

Once the density of the eigenvalues is computed one can perform the same change of variables described in Appendix A.1 for the case of single variance, and obtain the density $\rho_t(r)$ for the eigenvalues of J_t.
Figures (6.4) and (4) report the evolution in time of the spectral density, as well as its cumulative function, when $f = 0.25$ and $f = 0.75$. The cumulative function has been used to estimate the formation of the intermediate gaps to compare with the experiments for the estimation of the data-manifold dimension.

B Network training and Model Architecture Details

Dataset	Image Size	Latent Dim.	Channel Mult.	Param. Count	Batch size	Iterations
Cifar10	32	128	(1, 2, 2, 2)	35.7M	128	500,000
MNIST	28	128	(1, 2, 2)	24.5M	128	400,000
CelebA	64	64	(1, 1, 2, 2, 4, 4)	27.4M	64	800,000

Table 1: Table displaying both model and training configurations for each dataset.

For the image datasets, we used the diffusion setting in (Ho et al., 2020). We use the variance scheduler with $\beta_{\min} = 10^{-4}$ and $\beta_{\min} = 2 \times 10^{-2}$, $T = 1000$ time steps, and score model backbone (PixelCNN++ (Salimans et al., 2017)). Furthermore, for each of the datasets, we adjusted the partameters to account for the different complexity see Table 1.

For the linear models, we used a Variance Exploding continuous score model trained with 2M steps (batch size 128). The model had a Residual architecture with size 128 hidden channels in each layer, two residual blocks comprised by two linear layers with SiLu.

We primarily utilized NVIDIA Tesla V100 GPUs with 32 GB of memory.

C Experimental methodology: Computing the Singular Values of the Jacobian of the Score Function

For computing the singular values, we use the procedure from (Stanczuk et al., 2022) reported in algorithm 1. For the linear models and MNIST models we used a symmetrized version which we empirically found to be more stable, reported in algorithm 2.

Algorithm 1 Estimate singular values at x_0

Require: s_θ - trained diffusion model (score), t_0 - sampling time, K - number of score vectors.

1: Sample $x_0 \sim p_0(x)$ from the data set
2: $d \leftarrow \dim(x_0)$
3: $S \leftarrow$ empty matrix
4: **for** $i = 1, ..., K$ **do**
5: Sample $x_{t_0}^{(i)} \sim \mathcal{N}(x_{t_0} | x_0, \sigma_{t_0}^2 I)$
6: Append $s_\theta(x_{t_0}^{(i)}, t_0)$ as a new column to S
7: **end for**
8: $(s_i)_{i=1}^d, (v_i)_{i=1}^d, (w_i)_{i=1}^d \leftarrow \mathrm{SVD}(S)$

Algorithm 2 Estimate singular values at x_0 with central difference

Require: s_θ - trained diffusion model (score), t_0 - sampling time, K - number of score vectors.

1: Sample $x_0 \sim p_0(x)$ from the data set
2: $d \leftarrow \dim(x_0)$
3: $S \leftarrow$ empty matrix
4: **for** $i = 1, ..., K$ **do**
5: Sample $x_{t_0}^{+(i)} \sim \mathcal{N}(x_{t_0} | x_0, \sigma_{t_0}^2 I)$
6: Sample $x_{t_0}^{-(i)} \sim \mathcal{N}(x_{t_0} | x_0, -\sigma_{t_0}^2 I)$
7: Append $\frac{s_\theta(x_{t_0}^{+(i)}, t_0) - s_\theta(x_{t_0}^{-(i)}, t_0)}{2}$ as a new column to S
8: **end for**
9: $(s_i)_{i=1}^d, (v_i)_{i=1}^d, (w_i)_{i=1}^d \leftarrow \mathrm{SVD}(S)$