nature communications

Article

https://doi.org/10.1038/s41467-024-54281-3

Dynamical regimes of diffusion models

Received: 7 May 2024

Accepted: 6 November 2024

Published online: 17 November 2024

Check for updates

Giulio Biroli[1], Tony Bonnaire[1] ✉, Valentin de Bortoli[2] & Marc Mézard[3]

We study generative diffusion models in the regime where both the data dimension and the sample size are large, and the score function is trained optimally. Using statistical physics methods, we identify three distinct dynamical regimes during the generative diffusion process. The generative dynamics, starting from pure noise, first encounters a speciation transition, where the broad structure of the data emerges, akin to symmetry breaking in phase transitions. This is followed by a collapse phase, where the dynamics is attracted to a specific training point through a mechanism similar to condensation in a glass phase. The speciation time can be obtained from a spectral analysis of the data's correlation matrix, while the collapse time relates to an excess entropy measure, and reveals the existence of a curse of dimensionality for diffusion models. These theoretical findings are supported by analytical solutions for Gaussian mixtures and confirmed by numerical experiments on real datasets.

Machine learning has recently witnessed thrilling advancements, especially in the realm of generative models. At the forefront of this progress are diffusion models (DMs), which have emerged as powerful tools for modeling complex data distributions and generating new realistic samples. They have become the state of the art in generating images, videos, audio or 3D scenes[1–8]. Although the practical success of generative DMs is widely recognized, their full theoretical understanding remains an open challenge. Rigorous results assessing their convergence on finite-dimensional data have been obtained in refs. 9–14. However, realistic data live in high-dimensional spaces, where interpolation between data points should face the curse of dimensionality[15]. A thorough understanding of how generative models escape this curse is still lacking. This requires approaches able to take into account that the number and the dimension of the data are simultaneously very large. In this work, we face this challenge using statistical physics methods which have been developed to study probability distributions, disordered systems and stochastic processes in high dimensions[16–18].

DMs work in two stages. In the forward diffusion, one starts from a data point (e.g., an image) and gradually adds noise to it, until the image has become pure noise. In the backward process, one gradually denoises the image using a diffusion in an effective force field (the score) which is learnt by leveraging techniques from score matching[19,20] and deep neural networks. In this study, we focus on DMs which are efficient enough to learn the exact empirical score, i.e., the one obtained by noising the empirical distribution of data. In practical implementations, this should happen when one performs a long training of a strongly over-parameterized deep network to learn the score, in the situation when the number of data is not too large.

Within this framework, we develop a theoretical approach able to characterize the dynamics of DMs in the simultaneous limit of large dimensions and large dataset. We show that the backward generative diffusion process consists of three subsequent dynamical regimes. The first one is basically pure Brownian motion. In the second one, the backward trajectory finds one of the main classes of the data (for instance if the data consist of images of horses and cars, a given trajectory will specialize towards one of these two categories). In the third regime, the diffusion "collapses" onto one of the examples of the dataset: a given trajectory commits to the attraction basin of one of the data points, and the backward evolution brings it back to that exact data point. Since DMs are defined as the time reversal of a forward noising process, the generative process has to collapse on the training dataset under the exact empirical score assumption[21,22]. We show, by performing a thorough analysis of the curse of dimensionality for DMs, that this memorization can be avoided at finite times only if the size of the dataset is exponentially large in the dimension. An alternative, which is the one used in practice, is to rely on regularization and approximate learning of the score, departing from its exact form.

[1]Laboratoire de Physique de l'Ecole Normale Supérieure, ENS, Université PSL, CNRS, Sorbonne Université, Université Paris Cité, Paris, France. [2]Computer Science Department, ENS, CNRS, PSL University, Paris, France. [3]Department of Computing Sciences, Bocconi University, Milano, Italy.
✉e-mail: tony.bonnaire@ens.fr

Understanding this crucial aspect of generative diffusion is a key open challenge[23–25] for which analyzing what happens when the exact empirical score is used represents a first step.

Separating these three dynamical regimes, we identify two characteristic cross-over times. The speciation time t_S is the transition from pure noise to the commitment of a trajectory towards one category. The collapse time t_C is the time where the backward trajectory falls into the attractor of one given data point. We provide mathematical tools to predict these times in terms of structure of data. We will first study simple models such as high-dimensional Gaussian mixtures, where we obtain a full analytical solution and hence a very detailed understanding. Within this setting, we show that in the simultaneous limit of large number and dimension of the data, the speciation and collapse cross-overs become sharp thresholds. Interestingly, both of them are related to phase transitions studied in physics. We then extend our results to more general settings and discuss the key role played by the dimensionality of data and the number of samples. Finally, we perform numerical experiments and confront the theory to real data such as CIFAR-10, ImageNet, and LSUN, showing that our main findings hold in realistic cases. We conclude by highlighting the consequences and the guidelines offered by our results, and discussing the next research steps, in particular how to go beyond the exact empirical score framework.

Results

We focus on cases in which the data can be organized in distinct classes. For simplicity, we consider below two classes, identified in the spectrum of the covariance matrix of the data: this spectrum is assumed to display a single large eigenvalue along the principal component which we will denote Λ. This is a simplifying assumption; our analysis can be extended to more than two classes and subclasses within classes. The data consist of n data points $\mathbf{a} \in \mathbb{R}^d$. We assume that there exists an underlying distribution $P_0(\mathbf{a})$ from which data are drawn, and we denote by $P_0^e(\mathbf{a}) = \sum_{\mu=1}^{n} \delta(\mathbf{a} - \mathbf{a}_\mu)/n$, the empirical distribution of the data. The components of $\mathbf{a}$ are normalized to be finite for large d, meaning we assume that the moments $\int d\mathbf{a} P_0(\mathbf{a}) a_i^p$ remain of order one for $d \to \infty$ and finite p. This implies in particular that the expectation of $|\mathbf{x}|^2$ grows linearly with d.

There exist many variants of DMs which are basically equivalent. We focus here on the diffusion process which consists in d independent Ornstein-Uhlenbeck Langevin equations,

$$d\mathbf{x}(t) = -\mathbf{x}(t)dt + d\mathbf{B}(t), \tag{1}$$

where $d\mathbf{B}(t)$ is square root of two times the standard Wiener process (a.k.a., Brownian motion) in $\mathbb{R}^d$. The exact empirical score is given by $\mathcal{F}_i(\mathbf{x}, t) = \partial \log P_t^e(\mathbf{x})/\partial x_i$ where $P_t^e(\mathbf{x})$ is the noisy empirical distribution at time t due to the process in (1)

$$P_t^e(\mathbf{x}) = \int d\mathbf{a} P_0^e(\mathbf{a}) \frac{1}{\sqrt{2\pi\Delta_t}^d} \exp\left(-\frac{1}{2}\frac{(\mathbf{x} - \mathbf{a}e^{-t})^2}{\Delta_t}\right). \tag{2}$$

This is the convolution of the empirical distribution of the data, $P_0^e(\mathbf{a})$, with a Gaussian law of variance $\Delta_t = (1 - e^{-2t})$. At long times $P_t^e(\mathbf{x})$ is a Gaussian distribution with zero mean and covariance equal to the identity. In DMs the generation of new data is obtained by time-reversing this process using the backward dynamics

$$-dy_i(t) = y_i dt + 2\mathcal{F}_i(y, t)dt + d\xi_i(t), \tag{3}$$

where the noise $d\xi_i(t)$ has the same distribution as in the forward process.

Our main contribution is the characterization of three dynamical regimes in the time-reversed process in the limit of large number of data and large dimension. In regime I, at the beginning of the backward process, the random dynamical trajectories generated by (3) have not committed to a particular class of data. They have roughly the same probability to end up in one of the two classes. Figure 1 illustrates this behavior by showing that trajectories corresponding to different final classes are within a single bundle (or tube). In regime II, instead, the dynamical trajectories have committed to a particular class. In this case, the trajectory will remain in the same class until the end of the backward process. During regime II, the backward dynamics generate the features needed to produce samples in a given class, but the fate of the trajectory in terms of class is sealed. In analogy with evolutionary dynamics, we call speciation the cross-over between regimes I and II, which has also important connections with the concept of symmetry breaking in physics[26,27] and with stochastic localization in probability theory[28,29]. As we shall show, the speciation cross-over takes place on a time scale t_S defined by

$$\Lambda e^{-2t_S} = 1, \qquad \text{Speciation: I} \to \text{II} \tag{4}$$

where Λ is the eigenvalue of the principal component of the covariance matrix of the data. Note that in the high-dimensional limit, if Λ diverges with d (typically one would expect $\Lambda \propto d$), then t_S diverges

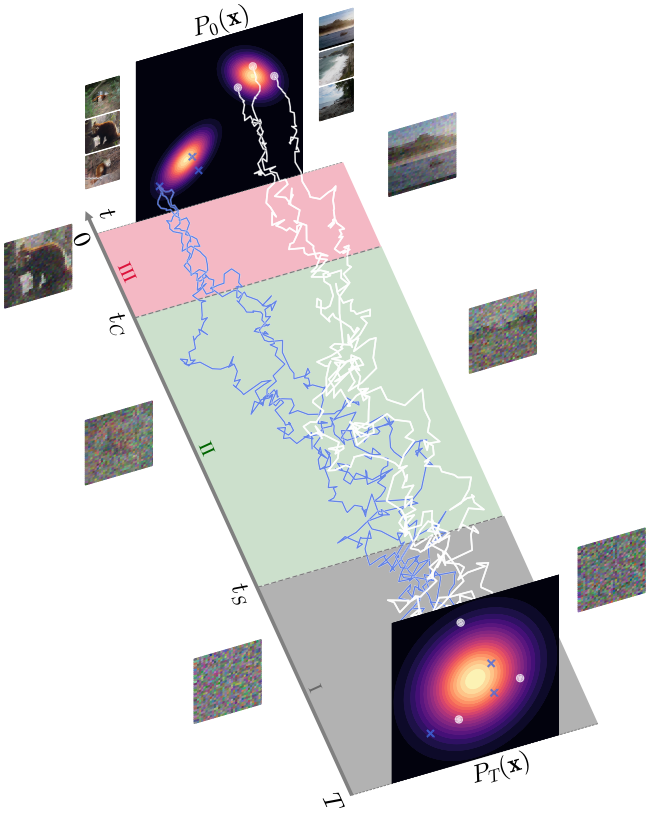

Fig. 1 | Illustration of the three regimes of the backward dynamics through an example corresponding to a Gaussian mixture in two dimensions. Trajectories are colored white and blue according to their class at the end of the backward dynamics. In regime I, blue and white trajectories are fluctuating within the same bundle and $\mathbf{x}$ is similar to white noise. At the speciation time t_S, the ensembles of blue and white trajectories divide and head towards the distribution associated to their respective class. Regime II is where the generative process constructs an $\mathbf{x}$ which resembles to one element of the class (e.g., a seashore in the illustration) without being linked to any data of the training set. At the collapse time t_C, trajectories start to be attracted by the data point on which they collapse at $t = 0$. Regime III corresponds to memorization, whereas in regime I and II, the diffusion model truly generalizes. The images on the right and on the left are illustrations obtained from our ImageNet numerical experiment (notice the collapse on the panda and seashore from the training set at $t = 0$).

logarithmically with d, and the speciation cross-over becomes a phase transition over time-scales of order t_S. Here, and in what follows, we measure time from the beginning of the forward process, i.e., large t corresponds to the beginning of the backward process. We find that in the regimes I and II the DM generalizes (if the number of data is large enough), which means that the empirical distribution at time t, $P_t^e(\mathbf{x})$ is basically the same as the true $P_t(\mathbf{x})$, which is the convolution of P_0 and a Gaussian of variance Δ_t. Therefore the distributions obtained by noising the desired P_0 and the empirical P_0^e are identical.

In regime III, the situation is completely different. There is no generalization; instead, the DM displays memorization of the training set. The probability distribution $P_t^e(\mathbf{x})$ no longer reflects $P_t(\mathbf{x})$. It actually decomposes in separated lumps around the points of the training set, and a given trajectory is committed to the attractor of the original data point, that is reached at $t = 0$. Figure 1 illustrates this behavior. We call collapse the cross-over between regimes II and III. As we shall show, the collapse takes place on a time scale t_C defined by

$$s(t_C) = s^{sep}(t_C), \qquad \text{Collapse: II} \to \text{III} \tag{5}$$

where $s(t) = -\frac{1}{d}\int d\mathbf{x}\, P_t(\mathbf{x}) \log P_t(\mathbf{x})$ is the Shannon entropy per variable of the distribution at time t, and $s^{sep}(t) = \frac{\log n}{d} + \frac{1}{2} + \frac{1}{2}\log(2\pi\Delta_t)$ is its counterpart for a mixture of n Gaussian distributions with variance Δ_t and centered on well-separated points. We shall argue below that criterion (5) is actually valid beyond the exact empirical score hypothesis, and therefore provides a way to characterize the collapse (or its absence) in practical applications.

For $t \gg 1$, the distribution $P_t(\mathbf{x})$ becomes a d-dimensional Gaussian, with entropy $\frac{d}{2} + \frac{d}{2}\log(2\pi)$. In consequence, the difference between the two entropies, $f(t) = s^{sep}(t) - s(t)$, that we call excess entropy density, equals $\frac{\log n}{d}$. On the other hand, for $t \to 0$, the entropy $s(t)$ goes to the one of $P_0(\mathbf{a})$, while $\log(2\pi\Delta_t) \to -\infty$. Therefore the excess entropy density goes to $-\infty$. When running the backward process, starting from large times where $f(t) = \frac{\log n}{d} > 0$, this excess entropy density decreases and crosses 0 at the collapse time t_C. Thus, by monitoring the time-dependence of $f(t)$, one can pinpoint when the collapse takes place. In numerical studies where P_0 is not known, one can approximate $f(t)$ as $f^e(t) = s^{sep}(t) - s^e(t)$, where $s^e(t)$ is the entropy per variable of the empirical distribution $P_t^e(\mathbf{x})$. This is a good approximation for $t \geq t_C$, where the two distributions P_t^e and P_t are identical and $f^e(t) = f(t)$. Instead, as we shall show, for $t < t_C$ the empirical excess entropy density vanishes while $f(t) < 0$. The parameter $\alpha = \frac{\log n}{d}$ plays a key role in the collapse, as it allows to tune the value of t_C. A very small α, implies a very small excess entropy density for $t \gg 1$. In this case, the collapse takes place at the very beginning of the backward dynamics. One needs $\alpha \sim O(1)$ in order to have $t_C \sim O(1)$. To diminish t_C (shrink regime III) and reduce the collapse, one has to increase α. These findings characterize a curse of dimensionality which is different but related to the one arising in supervised learning: in order to avoid memorization, the number of data has to increase exponentially with the dimension d. In practice, this unwanted phenomenon is avoided by using an approximate score function which is smoother than the exact one, together with a large enough dataset. We will come back to this point later in the discussion.

Our characterization of the backward dynamics is obtained using statistical physics methods developed to study phase transitions, disordered systems, and out-of-equilibrium dynamics in physics. We provide a brief introduction in SI Appendix. The connection is not only methodological; in fact, our analysis unveils interesting relationships between the transition described above and phenomena intensively studied in physics in the last decades. In particular, the collapse transition is mutatis mutandis a glass transition in which the low energy glass states correspond to the training data. The results we obtain in Eqs. (4) and (5), which are at the level of rigor of theoretical physics, provide guidelines and testable predictions for realistic applications. Numerical experiments on several subsets of realistic datasets (MNIST, CIFAR-10, ImageNet, and LSUN) confirm their validity.

Discussion

In this work, we analyze the backward dynamics of DMs assumed to be efficient enough to learn the exact empirical score. We show that for large number of data n, large dimension d, and in the absence of regularization, the backward dynamics display three different regimes. We characterized the cross-overs between them, dubbed speciation and collapse, which become true transitions in the large n, d limit. Interestingly, both of them have physical counterparts in the theory of phase transitions. Speciation is a symmetry-breaking transition of $P_t(\mathbf{x})$ at which the most prominent classes are generated[26,27]. Collapse corresponds to a glass transition at which $P_t(\mathbf{x})$ fragments in an ensemble of lumps centered around the training data. Our approach characterizes the time at which speciation and collapse take place in terms of structure of data. The speciation time is determined by the eigenvalue of the principal component of the covariance matrix of the data, whereas the collapse time is governed by the entropy of the noised data distribution.

Although we focus on a specific DM, our results hold for the large variety of DMs which are based on inverting the noising process[30]. These methods use different procedures to implement the denoising process $P_\infty(\mathbf{x}) \to P_t(\mathbf{x}) \to P_0(\mathbf{x})$. However, the phenomena of speciation and collapse can be understood and analyzed focusing on the properties of the distribution $P_t(\mathbf{x})$ alone. In consequence, since all these different DMs lead to the same backward evolution of $P_t(\mathbf{x})$, the criteria from Eqs. (4) and (5) for speciation and collapse hold in general, whether the generative process is deterministic and flow-based, or stochastic and diffusion-based.

Our work is done within the exact empirical score hypothesis. However, the phenomena we analyze are relevant also when the score is learned approximately, as shown by our numerical experiments. Moving beyond this hypothesis opens up multiple avenues for further research. On the one hand, it would be interesting to analyze simple models of data and score in the limit of large number of data, dimension and parameters, as initiated in refs. 25,26. On the other hand, it would be important to develop a quantitative study of the role of regularization on the phenomena we presented in this work. As discussed in ref. 26, using a good model of the score is a key ingredient to generate data in the right proportions, i.e., reproducing the correct weights of the classes. In consequence, regularization methods could be detrimental in regime I and for speciation: by preventing the score to be close to the exact one they could lead to data with correct features (e.g., realistic images) but wrong proportions with respect to the training set. On the contrary, regularization is beneficial to avoid the collapse. As our work demonstrates, DMs are cursed: one needs an exponential number of data in d to avoid the collapse of training points. Regularization allows to circumvent this problem in practice if the number of data is large enough[23] (see also SI Appendix). The volume argument we develop in this work can be applied beyond the exact empirical score hypothesis and could offer a way to analyze quantitatively how the collapse depends on both n and d, and the capacity of the model used to learn the score.

Finally, our results also provide suggestions to improve and understand procedures used in practical applications. In particular, taking into account the existence of the three regimes of the backward dynamics we depict in practical implementations could lead to better performances.

Methods

Analytical analysis of Gaussian mixtures

An instructive simple example to study the backward dynamics in the large d and n limit is the Gaussian mixture model. In this case, the initial law $P_0(\mathbf{a})$ is the superposition of two Gaussian clusters of equal weight,

which we take for simplicity with means $\pm\mathbf{m}$, and the same variance σ^2. We shall assume that $|\mathbf{m}|^2 = d\tilde{\mu}^2$, with σ and $\tilde{\mu}$ of order 1.

Speciation time. In order to show the existence of regimes I and II, and compute the speciation time, we focus on the following protocol which consists of cloning trajectories. We consider a backward trajectory starting at time $t_f \gg 1$ from a point $\mathbf{x}_f$ drawn from a random Gaussian distribution where all components are independent with mean zero and unit variance. This trajectory evolves backward in time, through the backward process until time $t < t_f$. At this time the trajectory has reached the point $\mathbf{y}(t)$, at which cloning takes place. One generates for $\tau < t$ two clones, starting from the same $\mathbf{x}_1(t) = \mathbf{x}_2(t) = \mathbf{y}(t)$, and evolving as independent trajectories $\mathbf{x}_1(\tau)$ and $\mathbf{x}_2(\tau)$, i.e., with independent thermal noises. We compute the probability $\phi(t)$ that the two trajectories ending in $\mathbf{x}_1(0)$ and $\mathbf{x}_2(0)$ are in the same class. Defining $P(\mathbf{x}_1, 0|\mathbf{y}, t)$ as the probability that the backward process ends in $\mathbf{x}_1$, given that it was in $\mathbf{y}$ at time t, the joint probability of finding the trajectory in $\mathbf{y}$ at time t and the two clones in $\mathbf{x}_1$ and $\mathbf{x}_2$ at time 0 is obtained as $Q(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}, t) = P(\mathbf{x}_1, 0|\mathbf{y}, t)P(\mathbf{x}_2, 0|\mathbf{y}, t)P(\mathbf{y}, t)$. Then $\phi(t)$ is the integral of Q over $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{y}$ with the constraint $(\mathbf{x}_1 \cdot \mathbf{m})(\mathbf{x}_2 \cdot \mathbf{m}) > 0$. This simplifies into a one-dimensional integral (see SI Appendix):

$$\phi(t) = \frac{1}{2}\int_{-\infty}^{+\infty} \mathrm{d}y \frac{G(y, me^{-t}, \Gamma_t)^2 + G(y, -me^{-t}, \Gamma_t)^2}{G(y, me^{-t}, \Gamma_t) + G(y, -me^{-t}, \Gamma_t)}, \tag{6}$$

where $G(y, u, v)$ is a Gaussian probability density function for the real variable y, with mean u and variance v, and $m = |\mathbf{m}| = \tilde{\mu}\sqrt{d}$, $\Gamma_t = \Delta_t + \sigma^2 e^{-2t}$. The probability $\phi(t)$ that the two clones end up in the same cluster is a decreasing function of t, going from $\phi(0) = 1$ to $\phi(\infty) = 1/2$. In the large d limit, the scaling variable controlling the change of ϕ is $\tilde{\mu}\sqrt{d}e^{-t}$ which can be rewritten as $\tilde{\mu}e^{-(t-t_S)}$ by using $t_S = (1/2)\log d$. This explicitly shows that speciation takes place at the time scale t_S on a window of time of order one. As detailed in SI Appendix, this expression for t_S coincides with the one obtained from the general criterion (5). We show in Fig. 2 the analytical result from (6) and direct numerical results obtained for increasingly larger dimensions. This comparison shows that our analysis is accurate already for moderately large dimensions. In the limit of infinite d, the analytical curve in Fig. 2 suddenly jumps from one to zero at $t/t_S = 1$, corresponding to a symmetry-breaking phase transition (or a threshold phenomenon) on the time scale t_S. In the numerics, following finite size scaling theory[31], we have defined the speciation time as the crossing point of the curves for different d, which corresponds approximatively to $\phi(t_S) = 0.775$ and indeed is of the order $t_S = (1/2)\log d$ for $d \to \infty$. As it happens in mean-field theories of phase transitions[32], the large dimensional limit allows to obtain a useful limiting process. In our case, this leads to a full characterization of the asymptotic backward dynamics. At its beginning, i.e., in regime I, the overlap with the centers of the Gaussian model, $\pm\mathbf{m}\cdot\mathbf{x}(t)$, is of order $\sqrt{d}$. Defining $q(t) = \mathbf{m}\cdot\mathbf{x}(t)/\sqrt{d}$, one can obtain a closed stochastic Langevin equation on q in a potential $V(q, t)$ (see SI Appendix),

$$-\mathrm{d}q = -\frac{\partial V(q,t)}{\partial q}\mathrm{d}t + \mathrm{d}\eta(t), \tag{7}$$

where $\eta(t)$ is square root of two times a Brownian motion, and

$$V(q,t) = \frac{1}{2}q^2 - 2\tilde{\mu}^2 \log\cosh\left(qe^{-t}\sqrt{d}\right). \tag{8}$$

At large d, this potential is quadratic at times $t \gg t_S = (1/2)\log d$, and it develops a double well structure, with a very large barrier, when $t \ll t_S = (1/2)\log d$. The trajectories of $q(t)$ are subjected to a force that drives them toward plus and minus infinity. The barrier between positive and negative values of q becomes so large that trajectories commit to a definite sign of q: this is how the symmetry breaking takes place dynamically at the time scale t_S, in agreement with the previous cloning results. Regime II corresponds to the scaling limit $q \to \infty$, where $\mathbf{m}\cdot\mathbf{x}(t)$ becomes of order d. In this regime, the rescaled overlap $\mathbf{m}\cdot\mathbf{x}(t)/d$ concentrates, and its sign depends on the set of trajectories one is focusing on. Moreover, the stochastic dynamics of the x_i correspond to the backward dynamics for a single Gaussian centered in $\pm\mathbf{m}$. This shows that the dynamics generalizes, see SI Appendix (and also[33] for similar results).

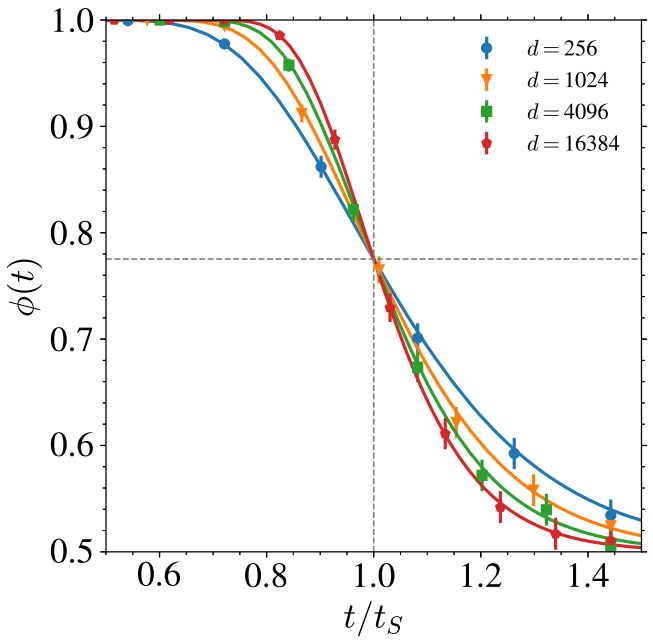

Fig. 2 | Speciation in Gaussian mixtures. Evolution of the probability $\phi(t)$ that the two clones end up in the same cluster as a function of t/t_S for several values of d at fixed $\tilde{\mu} = 1$ and $\sigma = 1$. The solid line corresponds to the evaluation of (6) while the dots are obtained by sampling 10,000 clone trajectories. The vertical (resp. horizontal) dashed line corresponds to $t/t_S = 1$ (resp. $\phi(t) = 0.775$). Error bars correspond to thrice standard error.

Collapse time. In order to study the collapse, we consider the probability distribution $P_t^e(\mathbf{x})$ given in (2) around a point $\mathbf{x}$ which has been obtained in the forward process starting at $t = 0$ from $\mathbf{x} = \mathbf{a}_1$. Our aim is to establish whether the score obtained from $P_t^e(\mathbf{x})$ imposes a force that pushes trajectories toward $\mathbf{a}_1$ in the backward process, corresponding to memorization, or instead allows for generalization. We shall consider that both n and d go to infinity, keeping $\alpha = \frac{\log n}{d}$ fixed (it will be clear from the analysis that this is the correct ratio for the asymptotic analysis).

The vector $\mathbf{x}$ we consider is equal to $\mathbf{a}_1 e^{-t} + \mathbf{z}\sqrt{\Delta_t}$ where $\mathbf{z}$ has iid Gaussian components with zero mean and unit variance. The probability can be written as $P_t^e(\mathbf{x}) = \left[Z_1 + Z_{2\ldots n}\right]/\sqrt{2\pi\Delta_t}^d$ where $Z_1 = e^{-\frac{1}{2}(\mathbf{x}-\mathbf{a}_1 e^{-t})^2/(2\Delta_t)} = e^{-\frac{\mathbf{z}^2}{2}}$ and

$$Z_{2\ldots n} = \sum_{\mu=2}^{n} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{a}_\mu e^{-t})^2/(2\Delta_t)}, \tag{9}$$

In the large d limit (to exponential accuracy), $Z_1 \simeq e^{-d/2}$. The computation of $Z_{2\ldots n}$ is instead tricky. Even though it is a sum of n uncorrelated random contributions, standard concentration methods, e.g., central limit theorem, do not apply as each term of the sum corresponds to the exponential of a random variable scaling as $\log n$[34]. Statistical physics

tools developed to study spin-glasses provide the method to solve this problem, see SI Appendix. In fact, given $\mathbf{x}$, $Z_{2\ldots n}$ is the partition function of a system with $n-1$ independent "random energy levels" $E^{\mu} = \frac{1}{2}[(\mathbf{a}_1 - \mathbf{a}_{\mu})e^{-t} + \mathbf{z}\sqrt{\Delta_t}]^2/(2\Delta_t)$ at thermal equilibrium at a temperature 1 (the randomness comes from $\mathbf{a}_{\mu}$). This is some elaboration of the "Random Energy Model" which was introduced originally in ref. 35 as a simple model of glass transition. A similar problem[36] was studied recently in the related context of dense associative memories, using large-deviations and replica methods (this connection was also noticed in ref. 37). Using a similar approach we can show (see SI Appendix) that in the large d and n limit the distribution of $(1/d)\log Z_{2\ldots n}$ concentrates on a value $\psi_+(t)$ which is an increasing function of the time t and of $\alpha = \lim_{d,n\to\infty}(1/d)\log n$. The analytical computation shows the existence of a collapse time t_C which separates two time regimes:

- Regime III: at small times, $t < t_C$, $\psi_+(t) < -(1/2)$ and the probability $P_t^e(\mathbf{x})$ is dominated by the term $Z_1 = e^{-(\mathbf{x}-\mathbf{a}_1 e^{-t})^2/(2\Delta_t)}$. When used in the backward diffusion, this gives a score that attracts $\mathbf{x}$ towards $\mathbf{a}_1$ at short times. With probability one, the backward trajectory, starting at time t from $\mathbf{x} = \mathbf{a}_1 e^{-t} + \mathbf{z}\sqrt{\Delta_t}$, collapses at the end of the backward process on data point $\mathbf{a}_1$. In this regime, the associated Random Energy Model is in a glass phase, which precisely corresponds to memorization.
- Regimes I and II: at large times, $t > t_C$, $\psi_+(t) > -(1/2)$ and the probability $P_t^e(\mathbf{x})$ is dominated by the term $Z_{2\ldots n}$. This is the regime which is not collapsed and corresponds to generalization (regime II) or Brownian motion (regime I): in a typical point $\vec{x}(t)$, drawn from the population distribution $P_t(\vec{x})$, the empirical distribution $P_t^e(\vec{x}, t)$ is equal to $P_t(\vec{x}, t)$ at leading order in d. In this regime the associated Random Energy Model is in the liquid (or high-temperature) phase.

The collapse time reads

$$t_C = \frac{1}{2}\log\left(1 + \frac{\sigma^2}{n^{2/d} - 1}\right). \tag{10}$$

This equation makes manifest the curse of dimensionality: t_C is of order one only when the number of data is exponential in the dimension. One needs $\frac{\log n}{d} \gg 1$ to push t_C to zero and avoid the collapse. One can check that this equation for t_C coincides with the one obtained from the general criterion (4). In Fig. 3 we plot the empirical excess entropy density $f^e(t)$ from which one deduces t_C as the largest time at which $f^e(t) = 0$. The numerical results compare well with the analytical prediction and confirm that the time scale for collapse is well captured by our approach, even at moderately large values of n. When $n, d \to \infty$ at fixed $\alpha = (1/d)\log n$, our analysis shows that regime III takes place on time scale of order one, and hence after the speciation transition, as illustrated in Fig. 1.

Generalization to realistic datasets

In the case of Gaussian mixtures, we could thoroughly characterize the backward dynamics for large n and d, using the knowledge of the initial distribution P_0. We now present a more general approach to the speciation and collapse transitions, for cases in which P_0 is not known or too complex to be analyzed exactly, but instead, a dataset drawn from P_0 is available. In particular, we present arguments to establish criteria from Eqs. (4) and (5) which can be directly applied to realistic datasets.

The speciation transition can be analyzed using the covariance matrix, C_0, of the initial data. Our assumption that data can be organized in two distinct and very different classes translates, in terms of C_0, in the existence of a strong principal component with eigenvalue Λ. The speciation transition can be generically understood in terms of the forward process: it corresponds to the time scale at which the noise added to the data blurs the principal component, and hence the connection to a given class. On this time scale, the trajectories coming from different classes in the forward process coalesce within the same bundle, as illustrated in Fig. 1. By time-reversal symmetry, it is therefore also on this time that the trajectories in the backward process separate and commit to the different classes. By evaluating the covariance matrix of the noised data $x_i(t)$, one finds

$$C(t) = C_0 e^{-2t} + \Delta_t \mathbf{I}. \tag{11}$$

The speciation time can be found by comparing the two contributions on the RHS in the direction of the principal component of C_0: Λe^{-2t} vs Δ_t. The first one is associated to the fluctuations between different classes, whereas the second corresponds to the broadening of $P_t(\mathbf{x})$ due to the noise. When the latter becomes of the same order as the former, the noise blurs the identification of trajectories in different classes. In consequence, one finds the criterion $\Lambda e^{-2t} \sim \Delta_t$, which shows that the variable controlling the speciation transition is indeed Λe^{-2t}, as we have shown for the Gaussian mixture model. Since for large d one expects a large Λ, the asymptotic window over which the speciation transition takes place is when Λe^{-2t} is of order one. This leads to the general result stated in (4). We have decided to associate the speciation transition with the time at which the variable Λe^{-2t} takes the value one, but this is a convention and the choice of another number of order one would work too when Λ is large. The argument above is obtained by studying the time at which the structure of the data, present at $t = 0$, is blurred by the noise. One can also tackle the problem starting by the other end, i.e., large times, and studying $P_t(\mathbf{x})$ perturbatively in e^{-t}, which is a small parameter at the beginning of the backward process. By performing an expansion of (2) in e^{-t}, one finds that at leading order the distribution $P_t(\mathbf{x})$ is a multivariate Gaussian with covariance $\frac{1}{\Delta_t}[\mathbf{I} - \frac{e^{-2t}}{\Delta_t}C_0]$. The expansion described above, and detailed in the SI, is similar to the Landau expansion of phase transitions[38], in which symmetry breaking can be understood in terms

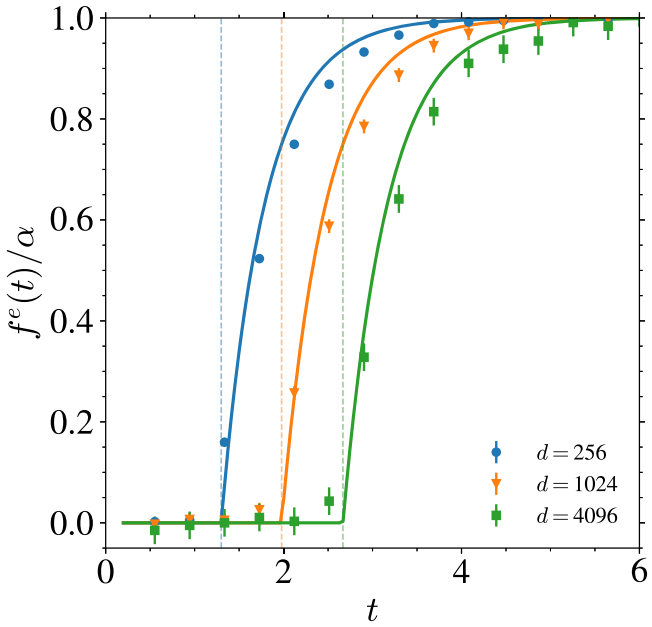

Fig. 3 | Collapse in Gaussian mixtures. Evolution of the excess entropy density $f^e(t)/\alpha$ as a function of time t for several values of d, at fixed $n = 20{,}000$. The solid lines are the theoretical predictions while the dots show the results of the numerical evaluation approximating the entropy from the dataset. The vertical dashed lines represent the collapse time t_C predicted analytically for Gaussian mixtures given in (10). Error bars correspond to thrice the standard error.

of the instability of the quadratic part of $\log P_t(\mathbf{x})$. The speciation transition is like a symmetry-breaking phase transition in $P_t(\mathbf{x})$[26,27], and this instability happens for $\Lambda e^{-2t} = \Delta_t \simeq 1$. The general result (4) can thus be obtained from two different complementary perspectives. We expect it to hold in a broad set of cases.

We now turn to the collapse transition. Generalization can be understood as the regime in which, at leading order in d, $P_t^e(\mathbf{x})$ coincides with its population counterpart $P_t(\mathbf{x})$ on typical noised data $\mathbf{x}(t)$ (i.e., drawn from $P_t(\mathbf{x})$). This implies that at time t the training set has been forgotten. At large d, and for a given data $\mathbf{a}$, the random vectors $\mathbf{a}e^{-t} + \mathbf{z}\sqrt{\Delta_t}$, drawn from $P_t^e(\mathbf{x})$, lie with probability one on the ball of radius $\sqrt{d\Delta_t}$ around $\mathbf{a}e^{-t}$. Note that they are actually located close to the circumference, as this is the region with the largest volume in large d. At small time t, one can therefore envision the set $\mathcal{M}^e$ typically covered by the empirical distribution P_t^e as the union of the non-overlapping balls centered around the data vectors $\mathbf{a}_\mu e^{-t}$. In this regime, the set $\mathcal{M}$ typically covered by the population distribution $P_t(\mathbf{x})$ is clearly different. It is independent of the training set and not as singular for $t = 0$. In consequence, in this regime $P_t^e(\mathbf{x})$ is collapsed on the data and the DM is in the memorization phase (regime III). By increasing t, the balls progressively grow and at a certain time, t_C, they cover the set $\mathcal{M}$. Beyond this point, one expects that the empirical and population distribution coincides on $\mathcal{M}$ (at leading order in d). This picture suggests a volume argument to identify the collapse: one finds the time at which the volume of $\mathcal{M}^e$ coincides with the one of $\mathcal{M}$. The key point is that in large d, the volume V_P covered by typical configurations associated to a given distribution P scales as $V_P = e^S$ where $S = -\int d\mathbf{x} P(\mathbf{x}) \log P(\mathbf{x})$ is the Shannon entropy, which scales like d. The set $\mathcal{M}^e$ corresponds to n d-dimensional Gaussian distributions with mean zero and covariance matrix given by the identity times Δ_t. For $t \le t_C$, the volume of $\mathcal{M}^e$ is the one of n non-overlapping Gaussians, which therefore reads: $V_{\mathcal{M}^e} = ne^{S_G}$ where $S_G = (d/2)(1 + \log(2\pi\Delta_t))$ is the entropy of one of the Gaussian distributions. On the other hand, the volume of $\mathcal{M}$ is given by $e^{ds(t)}$. By requiring the equality of the two volumes, $V_{\mathcal{M}^e} = V_{\mathcal{M}}$, one finds the general result for t_C given in (5). Note that all the identities used in the volume argument are correct up to corrections exponentially small in d. For this reason, (5) is expressed in terms of intensive quantities, and hence valid up to vanishing corrections for $d \to \infty$. This criterion can also be obtained from the Random Energy Model method, generalizing our derivation for the Gaussian Mixture model. Note that the arguments above can also be applied to cases in which the score is learned by approximate models. In this case, the distribution, $P_t(\mathbf{x})$, and its associated entropy density $s(t)$ have to be replaced by their model-dependent counterparts.

The arguments above offer another way, besides the exact computation by the Random Energy Model method, to illustrate the relationship between the collapse and the glass transition studied in physics[39]. In fact, the ways in which $P_t^e(\mathbf{x})$ covers the space of $\mathbf{x}$ before and after the collapse is the exact counterpart of how the Boltzmann law covers the configuration space before and after the ideal glass transition[39]. In the glass phase, the Boltzmann law is formed by lumps centered around amorphous optimal configurations (ideal glasses). Whereas in the liquid phase, the Boltzmann law is spread over all configurations. In the case studied here, the elements of the training set are the counterparts of the ideal glass configurations.

Having shown the generality of our criteria for speciation and collapse, we now test them on realistic data.

Numerical experiments

In the following, we focus on various image datasets and we learn the score function using a state-of-the-art neural network with a finite number of samples n. We assume that, by using a heavily over-parameterized model, the resulting estimate $\hat{\mathcal{F}}$ of the score is close to the exact one.

Settings and datasets. We train a DM as first described in ref. 40, and corresponding to (1), with a time horizon $T = 5.05$ and a linear schedule for the variance. The denoiser is a U-Net[41] with an architecture similar to other traditional implementations of DMs[27,40,42]. Our training sets are constructed by focusing on data divided into two classes and extracted from the image datasets: MNIST[43], CIFAR-10[44], downsampled ImageNet[45], and LSUN[46]. For example, in the case of ImageNet, the two classes we use are panda and seashore (as illustrated in Fig. 1). The large variety of datasets allows us to explore different kinds of images and several values of n and d, see Table 1. We refer to SI Appendix for details on the processing of these datasets, the denoiser, and how the numerical experiments presented below are carried out.

Speciation time. To numerically extract the speciation time and compare it to the theoretical result given by (4), we use the same cloning procedure introduced for Gaussian mixtures. We generate two independent trajectories at time t of the backward process and numerically estimate the probability $\phi(t)$ that they end up in the same class by averaging over many initial conditions and backward processes. In order to recognize the class of $\mathbf{x}_1(0)$ and $\mathbf{x}_2(0)$, we use a classifier with a ResNet-18 architecture[47], which has a test accuracy larger than 95% on the datasets we focus on. The corresponding results are shown in Fig. 4. The time axis has been rescaled by the theoretically predicted $t_S = (1/2)\log\Lambda$ from (4). The values of Λ and t_S are listed in

Table 1 | Image datasets used for speciation experiments

Dataset name	n	d	Λ	t_S
1. MNIST	10,000	1024	7.66	1.02
2. CIFAR	3000	3072	16.72	1.41
3. ImageNet16	2000	768	3.05	0.56
4. ImageNet32	2000	3072	12.11	1.25
5. LSUN	40,000	12,288	60.52	2.05

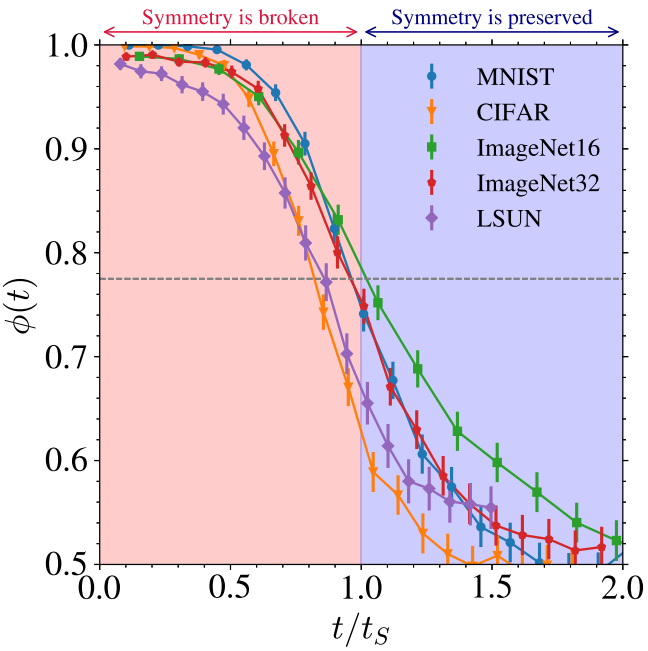

Fig. 4 | Speciation in realistic datasets. Evolution of $\phi(t)$, the probability that the two clones end up in the same class, as a function of t/t_S for several image datasets. The values of t_S are the theoretical prediction for the speciation time obtained using (4) and listed in Table 1. The dashed horizontal line indicate $\phi(t) = 0.775$, the error bars correspond to thrice the standard error and the solid lines linearly interpolate the experimental points.

https://doi.org/10.1038/s41467-024-54281-3

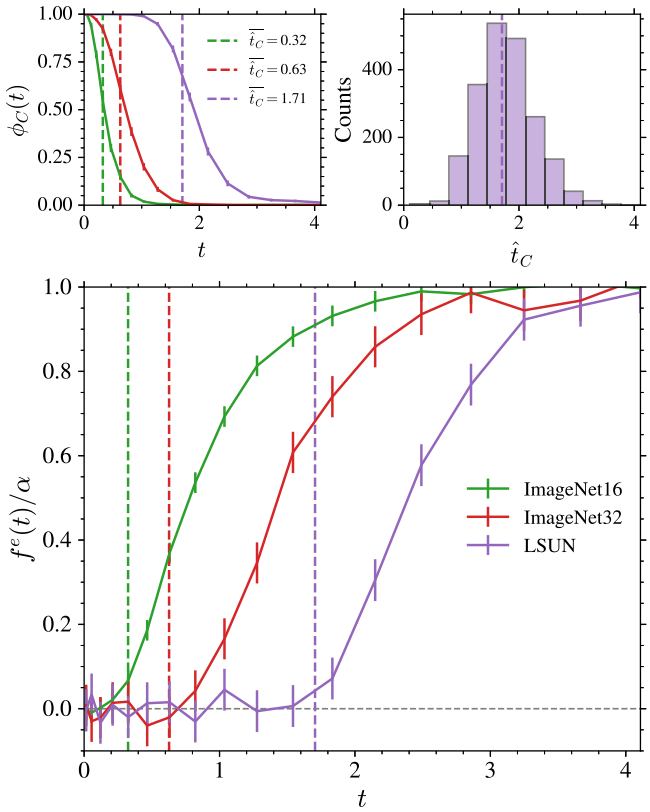

Fig. 5 | Collapse in realistic datasets (ImageNet16, ImageNet32 and LSUN). (Top-left) Evolution of $\phi_C(t)$, the probability that two cloned trajectories collapse on the same data of the training set at time zero. (Top-right) Histograms of $\hat{t}_c$ derived from the last-changing indices $\mu_\star$ on 4000 generated samples for the LSUN dataset trained with $n = 200$. (Bottom) Evolution of the empirical excess entropy $f(t)/\alpha$. In all panels, the colored vertical dashed lines indicate the average of $\hat{t}_C$. The error bars correspond to thrice the standard error.

Table 1. Figure 4 shows that indeed the speciation phenomenon is at play in realistic DMs: $\phi(t)$ goes from 0.5 when $t \gg t_S$ to one when $t \ll t_S$, in a way which is qualitatively analogous to Fig. 2. When rescaled with respect to t_S, the time-dependence exhibits remarkable similarity across vastly different image sets, suggesting evidence of the common underlying phenomenon of speciation. Moreover, our prediction for the speciation time, t_S, captures well the behavior found in these numerical experiments.

Collapse time. We focus on the datasets ImageNet16, ImageNet32, and LSUN. In order to be able to study thoroughly the collapse phenomenon, we have to keep the number of training data small ($n = 200$ for LSUN and $n = 2000$ for ImageNet). Otherwise, the model is not expressive enough to represent the singular behavior of the exact score at small times and collapse would not be observed (see also Fig. 6 of SI Appendix). To estimate the collapse time numerically, and compare it to the theoretical result from (5), we employ two distinct strategies. First, we use again the cloning procedure where a backward dynamics is cloned at a time t. The original and cloned dynamics are then evolved independently with separate noise. For sufficiently low values of n, all trajectories collapse onto a specific training point. We then estimate the probability $\phi_C(t)$ that the two cloned trajectories converge to the exact same data point in the training set at time zero. The results of this first experiment are shown in the top-left panel of Fig. 5. The cross-over time where $\phi_C(t)$ goes from zero (indicating the two clones collapse onto different training points) to one (meaning the clones collapse on the same training points) provides a first numerical estimate of t_C. An alternative estimation is found by tracking, during the backward process, the index of the closest neighbor in the training set, noted $\mu_\star(t) \in \{1, \cdots, n\}$, in the sense of the L2-norm. Considering $x(t)$ a generated sample at time t, this index therefore reads

$$\mu_\star(t) = \underset{\mu \in \{1,\cdots,n\}}{\operatorname{argmin}} \|\mathbf{a}_\mu e^{-t} - \mathbf{x}(t)\|_2^2. \tag{12}$$

At the beginning of the backward process (large t), this index fluctuates at each timestep. However, once the features determining the attraction to a given training point appear, the index remains fixed to that data point. For an individual trajectory, we estimate the collapse time $\hat{t}_C$ as the last timestep during the backward process at which this index changes—meaning that for $t < \hat{t}_C$, the nearest neighbor of $\mathbf{x}(t)$ stays the same. The top-right panel of Fig. 5 illustrates the distribution of $\hat{t}_C$ measured on the LSUN dataset. By averaging this distribution for each dataset, we obtain the estimates $\bar{\hat{t}}_C$ for the collapse time. They are marked as vertical dashed lines in all panels of Fig. 5, with colors corresponding to the three different training sets. Notably, the top-left panel shows that the two different methods to estimate the collapse time agree well with $\bar{\hat{t}}_C$ crossing $\phi_C(t)$ at around 0.60 for all datasets. We can now test the theoretical criterion for the collapse time given by (5), or equivalently by the largest time at which the empirical excess entropy density $f^e(t)$ is equal to zero, see the bottom panel of Fig. 5. The shape of $f^e(t)/\alpha$ is remarkably similar to the one obtained analytically for the Gaussian Model (Fig. 3). It clearly shows evidence of a transition. The theoretical estimate, $f^e(t_C) = 0$, also compares very well with the numerical ones $\bar{\hat{t}}_C$.

In conclusion, the numerical experiments presented in this section demonstrate the presence of speciation and collapse in realistic image datasets, and validate our theory, in particular the criteria from Eqs. (4) and (5) to identify the times at which these phenomena take place.

Data availability

All the datasets used to produce this research are publicly accessible online: MNIST, CIFAR, downsampled ImageNet, and LSUN.

Code availability

The code used to obtain the figures and run the experiments is freely accessible at this address.

References

1. Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N. & Ganguli, S. Deep unsupervised learning using nonequilibrium thermodynamics. In *Proc. International Conference on Machine Learning*. (PMLR, 2015).
2. Song, Y. & Ermon, S. Generative modeling by estimating gradients of the data distribution. In *Proc. Advances in Neural Information Processing Systems*. (Curran Associates Inc., 2019).
3. Song, Y. et al. Score-based generative modeling through stochastic differential equations. In *Proc. International Conference on Learning Representations* (2021).
4. Guth, F., Coste, S., De Bortoli, V. & Mallat, S. Wavelet score-based generative modeling. *Adv. Neural Inf. Process. Syst.* **35**, 478–491 (2022).
5. Yang, L. et al. Diffusion models: a comprehensive survey of methods and applications. *ACM Comput. Surv.* **56**, 1–39 (2023).
6. Saharia, C. et al. Photorealistic text-to-image diffusion models with deep language understanding. *Adv. Neural Inf. Process. Syst.* **35**, 36479–36494 (2022).
7. Bar-Tal, O. et al. Lumiere: a space-time diffusion model for video generation. *arXiv preprint arXiv:2401.12945* (2024).
8. Poole, B., Jain, A., Barron, J. T. & Mildenhall, B. Dreamfusion: text-to-3D using 2D diffusion. In *Proc. Eleventh International Conference on Learning Representations* (2023).

9. De Bortoli, V., Thornton, J., Heng, J. & Doucet, A. Diffusion schrödinger bridge with applications to score-based generative modeling. *Adv. Neural Inf. Process. Syst.* **34**, 17695–17709 (2021).
10. Lee, H., Lu, J. & Tan, Y. Convergence for score-based generative modeling with polynomial complexity. *Adv. Neural Inf. Process. Syst*. **35**, 22870–22882 (2022).
11. De Bortoli, V. Convergence of denoising diffusion models under the manifold hypothesis. *Trans. Mach. Learn. Res.* https://openreview.net/forum?id=MhK5aXo3gB (2022).
12. Benton, J., De Bortoli, V., Doucet, A. & Deligiannidis, G. Nearly d-Linear Convergence Bounds for Diffusion Models via Stochastic Localization. In *Proc. International Conference on Learning Representations* (2024).
13. Conforti, G., Durmus, A. & Silveri, M. G. Score diffusion models without early stopping: finite fisher information is all you need. *arXiv preprint arXiv:2308.12240* (2023).
14. Chen, S. et al. Sampling is as easy as learning the score: theory for diffusion models with minimal data assumptions. In *Proc. International Conference on Learning Representations* (2023).
15. Donoho, D. L. et al. High-dimensional data analysis: the curses and blessings of dimensionality. *AMS Math Chall. Lect.* **1**, 32 (2000).
16. Mezard, M. & Montanari, A. *Information, Physics, and Computation* (Oxford University Press, 2009).
17. Charbonneau, P. et al. *Spin Glass Theory and Far Beyond: Replica Symmetry Breaking after 40 Years* (World Scientific, 2023).
18. Bonnaire, T. et al. High-dimensional non-convex landscapes and gradient descent dynamics. *J. Stat. Mech. Theory Exp.* **2024**, 104004 (2024).
19. Hyvärinen, A. & Dayan, P. Estimation of non-normalized statistical models by score matching. *J. Mach. Learn. Res.* **6**, 24 (2005).
20. Vincent, P. A connection between score matching and denoising autoencoders. *Neural Comput.* **23**, 1661–1674 (2011).
21. Cattiaux, P., Conforti, G., Gentil, I. & Léonard, C. Time reversal of diffusion processes under a finite entropy condition. in *Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques* Vol. 59, 1844–1881 (Institut Henri Poincaré, 2023).
22. Haussmann, U. G. & Pardoux, E. Time reversal of diffusions. *Ann. Probab.* **14**, 1188–1205 (1986).
23. Kadkhodaie, Z., Guth, F., Simoncelli, E. P. & Mallat, S. Generalization in diffusion models arises from geometry-adaptive harmonic representations. In *Proc. Twelfth International Conference on Learning Representations* (2024).
24. Yoon, T., Choi, J. Y., Kwon, S. & Ryu, E. K. Diffusion probabilistic models generalize when they fail to memorize. In *Proc. ICML 2023 Workshop on Structured Probabilistic Inference & Generative Modeling* (2023).
25. Cui, H., Krzakala, F., Vanden-Eijnden, E. & Zdeborová, L. Analysis of learning a flow-based generative model from limited sample complexity. In *Proc. ICLR* (2024).
26. Biroli, G. & Mézard, M. Generative diffusion in very large dimensions. *J. Stat. Mech.* **2023**, 093402 (2023).
27. Raya, G. & Ambrogioni, L. Spontaneous symmetry breaking in generative diffusion models. In *Proc. Thirty-seventh Conference on Neural Information Processing Systems* (2023).
28. Eldan, R. Taming correlations through entropy-efficient measure decompositions with applications to mean-field approximation. *Probab. Theory Relat. Fields* **176**, 737–755 (2020).
29. El Alaoui, A., Montanari, A. & Sellke, M. Sampling from the Sherrington-Kirkpatrick Gibbs measure via algorithmic stochastic localization. In *Proc. IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)* 323–334 (IEEE, 2022).
30. Albergo, M. S., Boffi, N. M. & Vanden-Eijnden, E. Stochastic interpolants: a unifying framework for flows and diffusions. *arXiv preprint arXiv:2303.08797* (2023).
31. Privman, V. *Finite Size Scaling and Numerical Simulation of Statistical Systems* (World Scientific, 1990).
32. Opper, M. & Saad, D. *Advanced Mean Field Methods: Theory and Practice* (MIT Press, 2001).
33. Ghio, D., Dandi, Y., Krzakala, F. & Zdeborová, L. Sampling with flows, diffusion, and autoregressive neural networks from a spin-glass perspective. *Proc. Natl. Acad. Sci. USA* **121**, e2311810121 (2024).
34. Ben Arous, G., Bogachev, L. V. & Molchanov, S. A. Limit theorems for sums of random exponentials. *Probab. Theory Relat. fields* **132**, 579–612 (2005).
35. Derrida, B. Random-energy model: an exactly solvable model of disordered systems. *Phys. Rev. B* **24**, 2613 (1981).
36. Lucibello, C. & Mézard, M. The exponential capacity of dense associative memories. *Phys. Rev. Lett.* **132**, 077301 (2024).
37. Ambrogioni, L. The statistical thermodynamics of generative diffusion models. *arXiv preprint arXiv:2310.17467* (2023).
38. Chaikin, P. M., Lubensky, T. C. & Witten, T. A. *Principles of Condensed Matter Physics* Vol. 10 (Cambridge University Press, 1995).
39. Berthier, L. & Biroli, G. Theoretical perspective on the glass transition and amorphous materials. *Rev. Mod. Phys.* **83**, 587 (2011).
40. Ho, J., Jain, A. & Abbeel, P. Denoising diffusion probabilistic models. *Adv. Neural Inf. Process. Syst.* **33**, 6840–6851 (2020).
41. Ronneberger, O., Fischer, P. & Brox, T. U-net: convolutional networks for biomedical image segmentation. In *Proc. 18th International Conference, Munich, Germany, October 5–9, 2015, Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: Part III 18* 234–241 (Springer, 2015).
42. Song, J., Meng, C. & Ermon, S. Denoising diffusion implicit models. In *Proc. International Conference on Learning Representations* (2021).
43. Lecun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **86**, 2278–2324 (1998).
44. Krizhevsky, A., Nair, V. & Hinton, G. *Learning Multiple Layers of Features from Tiny Images* (Canadian Institute for Advanced Research, 2009).
45. Chrabaszcz, P., Loshchilov, I. & Hutter, F. A downsampled variant of imagenet as an alternative to the cifar datasets. *arXiv preprint arXiv:1707.08819* (2017). 1707.08819.
46. Yu, F. et al. Lsun: construction of a large-scale image dataset using deep learning with humans in the loop. *arXiv preprint arXiv:1506.03365* (2016). 1506.03365.
47. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 770–778 (IEEE, 2016).

Acknowledgements

The authors thank F. Bach, S. Mallat, and A. Montanari for helpful discussions. G.B. acknowledges support from the French government under the management of the Agence Nationale de la Recherche as part of the "Investissements d'avenir" program, reference ANR-19-P3IA0001 (PRAIRIE 3IA Institute). M.M. acknowledges the support of the PNRR-PE-AI FAIR project funded by the NextGeneration EU program.

Author contributions

G.B., V.d.B. and M.M. conceived the project, G.B. and M.M. did the analytical studies, and T.B. performed the numerical simulations and their analysis. All four authors wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-54281-3.

Correspondence and requests for materials should be addressed to Tony Bonnaire.

Peer review information *Nature Communications* thanks Georgios Batzolis, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.