FIELD-THEORETIC DIFFUSION MODELS AND SAMPLING AS SPATIAL PATTERN FORMATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Although images can be viewed as large vectors, such a representation fails to exploit nontrivial spatial structure: for example, nearby pixel locations tend to have similar colors and intensities. Similarly, theoretical analyses of image-based diffusion models tend to portray the input of denoisers and score functions as a grid of numbers rather than as a continuous object. In this work, we argue that the sampling dynamics of image-based diffusion models (and related concepts, like locality and translation equivariance) are usefully understood in terms of an idealization in which space is treated as continuous, which makes them a kind of spatial pattern formation process. To make this perspective concrete, we introduce and formalize field-theoretic diffusion models, whose central objects are fields: mappings from a real-valued pixel space to a real-valued color space. We consider field-theoretic diffusion models interesting both because they model the structure of architectures with strong spatial inductive biases, like convolutional neural networks, and because they provide a bridge between diffusion model theory and the theory of spatial pattern formation. We present theoretical results concerning sampling, including exact solutions for linear model sampling dynamics and brief results on the sampling dynamics of specific nonlinear models. The continuum description yields a novel picture of sampling in terms of reaction-diffusion-like dynamics, where different 'chemicals' drift, diffuse, and interact. We illustrate this intuition using a variety of examples. Our work demonstrates the theoretical benefits of reasoning about sampling via a continuous-space description.

1 INTRODUCTION

When used as input to neural networks, images are often represented as vectors or arrays of numbers. While this simple representational strategy is computationally useful and extremely successful, it also somewhat obscures important properties of images. Natural images are highly structured: the content of nearby pixels tend to be correlated, high-level features (e.g., a label like "dog") are often invariant to spatial geometric transformations like translations and rotations, and there is a characteristic distribution of spatial frequency content (Simoncelli & Olshausen, 2001).

In this paper, we leverage an explicitly spatial representation of images to better understand fundamental properties of image-based diffusion models (Sohl-Dickstein et al., 2015; Song & Ermon, 2019; Ho et al., 2020; Song et al., 2021). We model an image as a *field* $\phi(\boldsymbol{x})$ that maps a (generally two-dimensional) spatial coordinate $\boldsymbol{x}$ to a real-valued color space. Under this assumption, familiar objects like denoisers and score functions become *functionals* (i.e., functions of functions). We do not claim that it is impossible to reason about the consequences of spatial structure in the standard (discrete) setting; rather, our claim is that working in an idealized setting where pixel space is treated as continuous makes theorizing about this structure easier and more transparent.

The central observation of this paper is that sampling using an approximately local or translation-equivariant denoiser yields a spatial pattern formation process. For example, if we model locality as constraining a denoiser to depend only on ϕ and its first two (spatial) derivatives, each term of the partial differential equation (PDE) defined by sampling has a well-known interpretation: one term represents directed transport, one term represents diffusion, and additional terms may represent decay or interactions between field components. In this picture, generating an image from noise involves locally transporting, amplifying, and suppressing fluctuations until a pattern is formed.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

This theoretical reframing exercise is useful for two reasons. First, it is conceptually appealing—it is easier to reason about explicitly spatial properties like locality and translation equivariance in continuous spaces. Second, it provides a bridge between diffusion models and results on pattern formation and reaction-diffusion equations. This bridge is useful because it may help us reason about properties of nonlinear models. For example, linear stability analysis yields *dispersion relations* that describe how different spatial frequencies grow during sampling, while nonlinear pattern-forming equations show that noise can organize into domains, stripes, spots and other spatial structures.

Throughout, we focus on analytically tractable models, especially linear models (Wang & Vastola, 2024), since they are particularly well-behaved. Our main contributions are as follows.

- We introduce a **field-theoretic formulation of diffusion models** which makes it easier to mathematically study spatial inductive biases like locality and translation equivariance.
- We show that **probability-flow sampling using local/equivariant denoisers yields reaction-diffusion-like PDEs**, which we explicitly connect to pattern formation theory.
- We provide **analytic solutions to broad classes of linear models**.
- We show that concepts from pattern formation theory, like **symmetry breaking, finite-wavelength instabilities, and topological defects**, appear in the field-theoretic setting.

2 Formalism: field-theoretic models, locality, and equivariance

Fields. Motivated by the idea that we ought to take spatial structure in images more seriously, we introduce **field-theoretic diffusion models**, whose central objects are **fields**: mappings from a real-valued 'pixel' space $\mathbb{R}^{N_P}$ to a real-valued 'color' space $\mathbb{R}^{N_C}$ (i.e., $\boldsymbol{\phi} : \mathbb{R}^{N_P} \to \mathbb{R}^{N_C}$). These objects are intended to abstract the structure of images, but can model any data with spatial structure. The mathematics of fields is well-studied, and is foundational in subjects like physics (Kardar, 2007; Zinn-Justin, 2021; Schwartz, 2013) and pattern formation (Turing, 1952; Volpert & Petrovskii, 2009; Kondo & Miura, 2010). See Appendix A for a brief review of the mathematics of fields.

Forward/reverse diffusion. Let $p_{data}(\boldsymbol{\phi}_0)$ be a distribution over fields[1]. Field-theoretic diffusion models relate samples from $p_{data}(\boldsymbol{\phi}_0)$ to samples from the Gaussian process $p_{noise}(\boldsymbol{\phi}_T)$ via

$$\frac{\partial \boldsymbol{\phi}_t(\boldsymbol{x})}{\partial t} = -\beta_t \boldsymbol{\phi}_t(\boldsymbol{x}) + g_t \boldsymbol{\eta}_t \qquad t = 0 \to t = T \qquad \text{(forward, } p_{data} \text{ to } p_{noise}\text{)}$$

$$\frac{\partial \boldsymbol{\phi}_t(\boldsymbol{x})}{\partial t} = -\beta_t \boldsymbol{\phi}_t(\boldsymbol{x}) - \frac{g_t^2(\alpha_t \boldsymbol{D}_t[\phi_t](\boldsymbol{x}) - \boldsymbol{\phi}_t(\boldsymbol{x}))}{2\sigma_t^2} \qquad t = T \to t = \epsilon \qquad \text{(reverse, } p_{noise} \text{ to } p_{data}\text{)}$$

where $\boldsymbol{\eta}_t$ is a Gaussian white noise field. The forward process depends on $g_t \geq 0$, which controls the noise amplitude, and $\beta_t \geq 0$, which controls decay to the origin. These functions determine $\alpha_t := e^{-\int_0^t \beta_{t'}\, dt'}$ and $\sigma_t^2 := \alpha_t^2 \int_0^t (g_{t'}^2/\alpha_{t'}^2)\, dt'$, which govern the transition probability $p(\boldsymbol{\phi}_t|\boldsymbol{\phi}_0, t) = \mathcal{GP}(\alpha_t \boldsymbol{\phi}_0(\boldsymbol{x}), \sigma_t^2 \boldsymbol{\delta}(\boldsymbol{x} - \boldsymbol{x}'))$. The reverse process, which converts a sample from p_{noise} into a sample from p_{data}, depends on the time cutoff $\epsilon > 0$ and the (optimal) **denoiser** $\boldsymbol{D}_t$, which takes a field $\boldsymbol{\phi}$ as input and outputs another field with the same domain and codomain.

Let $p(\boldsymbol{\phi}_t|t) := \int p(\boldsymbol{\phi}_t|\boldsymbol{\phi}_0, t) p_{data}(\boldsymbol{\phi}_0)\ \mathcal{D}[\boldsymbol{\phi}_0]$, where $\mathcal{D}[\boldsymbol{\phi}_0]$ denotes the measure of a *functional integral* (see Appendix A). $\boldsymbol{D}_t$ is defined in terms of the **score functional** $\mathbf{s}_t[\boldsymbol{\phi}](\boldsymbol{x})$ via

$$\mathbf{s}_t[\boldsymbol{\phi}](\boldsymbol{x}) := \frac{\delta \log p(\boldsymbol{\phi}|t)}{\delta \boldsymbol{\phi}(\boldsymbol{x})} = \frac{\alpha_t \boldsymbol{D}_t[\boldsymbol{\phi}](\boldsymbol{x}) - \boldsymbol{\phi}(\boldsymbol{x})}{\sigma_t^2} \qquad \boldsymbol{D}_t[\boldsymbol{\phi}](\boldsymbol{x}) = \frac{\boldsymbol{\phi}(\boldsymbol{x}) + \sigma_t^2 \mathbf{s}_t[\boldsymbol{\phi}](\boldsymbol{x})}{\alpha_t} \qquad (1)$$

where $\delta/\delta\boldsymbol{\phi}(\boldsymbol{x})$ denotes a *functional derivative* (see Appendix A). Note that this relationship is just a continuous-space analogue of the usual denoiser-score relationship (see, e.g., Karras et al. (2022)). For a derivation of the reverse process and a proof of the above relationship, see Appendix B.

[1] Distributions over fields are more mathematically complex than distributions on finite-dimensional vector spaces, but these difficulties can be surmounted in various ways. See Appendix A.

Training objective. Here, training a diffusion model means learning a parameterized estimator $\hat{\boldsymbol{D}}_{\boldsymbol{\theta},t}$ of the denoiser $\boldsymbol{D}_t$. Assume this is done via a denoising objective (Karras et al., 2022)

$$J(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\phi},\boldsymbol{\phi}_0,t}\left\{\frac{\lambda_t}{2}\int \|\hat{\boldsymbol{D}}_{\boldsymbol{\theta},t}[\boldsymbol{\phi}](\boldsymbol{x}) - \boldsymbol{\phi}_0(\boldsymbol{x})\|_2^2 d\boldsymbol{x}\right\} \tag{2}$$

which is closely related to denoising score matching objectives (Vincent, 2011; Song & Ermon, 2019; Song et al., 2021). $\lambda_t > 0$ is a positive weighting function, and the time average is over a time-sampling distribution $p(t)$. $\mathcal{D}[\boldsymbol{\phi}]$ and $\mathcal{D}[\boldsymbol{\phi}_0]$ again denote functional integral measures. Without expressivity constraints, the global optimum of J is the optimal denoiser (Eq. 1; see Appendix C).

Locality. Information (e.g., object identity) in one small patch of an image tends to be unrelated to information in a distant patch, which motivates considering functions of fields that are *local*. Here, we consider a function $\boldsymbol{D}_t$ (e.g., a denoiser) local if its output at a location $\boldsymbol{x}$ only depends on the field's value near $\boldsymbol{x}$. In the extreme case, in which we call $\boldsymbol{D}_t$ *ultra-local*, it depends *only* on $\boldsymbol{\phi}(\boldsymbol{x})$:

$$\frac{\delta D[\phi](\boldsymbol{x})}{\delta\phi(\boldsymbol{y})} = \frac{\partial D[\phi](\boldsymbol{x})}{\partial\phi(\boldsymbol{x})}\delta(\boldsymbol{x}-\boldsymbol{y})\ . \tag{3}$$

We will slightly relax ultra-locality by considering functions which depend on the value of $\boldsymbol{\phi}(\boldsymbol{y})$ for $\boldsymbol{y}$ *near* $\boldsymbol{x}$; one can achieve this by assuming $\boldsymbol{D}_t[\boldsymbol{\phi}](\boldsymbol{x})$ depends on derivatives of $\boldsymbol{\phi}$ at $\boldsymbol{x}$, i.e.,

$$\boldsymbol{D}_t[\boldsymbol{\phi}](\boldsymbol{x}) = \boldsymbol{F}_t(\boldsymbol{x}, \boldsymbol{\phi}(\boldsymbol{x}), \partial_i\boldsymbol{\phi}(\boldsymbol{x}), \partial_{ij}\boldsymbol{\phi}(\boldsymbol{x}), ...)\ . \tag{4}$$

Why this makes sense is clearest when space is discretized. For example, when one uses a standard finite-difference approximation, the second derivative of a function at one lattice point depends on the values of that function at adjacent lattice points.

Equivariance. Another relevant inductive bias is that the output of a function of fields (e.g., an image classifier) should change in a specific way when its input is transformed. The idea of an *equivariant* model (Cohen & Welling, 2016; Bronstein et al., 2021) formalizes this intuition. A function $\boldsymbol{D}_t$ is *translation equivariant* if translations of the input field produce translations in output:

$$\boldsymbol{D}_t[\mathcal{T}_{\Delta\boldsymbol{x}}\boldsymbol{\phi}](\boldsymbol{x}) = \boldsymbol{D}_t[\boldsymbol{\phi}](\boldsymbol{x}-\Delta\boldsymbol{x})\ , \tag{5}$$

where the translation operator $\mathcal{T}_{\Delta\boldsymbol{x}}$ is defined via $[\mathcal{T}_{\Delta\boldsymbol{x}}\boldsymbol{\phi}](\boldsymbol{x}) = \boldsymbol{\phi}(\boldsymbol{x}-\Delta\boldsymbol{x})$ for any spatial shift $\Delta\boldsymbol{x} \in \mathbb{R}^{N_P}$. We are specifically interested in this kind of equivariance, since it is (approximately) baked into CNN architectures and has been claimed to play a role in generalization (Kamb & Ganguli, 2025). If a function is *both* linear and translation equivariant, it must have the form

$$\boldsymbol{D}_t[\boldsymbol{\phi}](\boldsymbol{x}) = \int \boldsymbol{W}(\boldsymbol{x}-\boldsymbol{y}, t)\boldsymbol{\phi}(\boldsymbol{y})\ d\boldsymbol{y} \tag{6}$$

for some filter $\boldsymbol{W}$. Interestingly, translation equivariance amounts to imposing ultra-locality in *wavevector space* for linear models. If $\tilde{\boldsymbol{D}}_t[\boldsymbol{\phi}](\boldsymbol{k})$ is the Fourier transform of $\boldsymbol{D}_t$ at wavevector $\boldsymbol{k}$,

$$\tilde{\boldsymbol{D}}_t[\boldsymbol{\phi}](\boldsymbol{k}) = \tilde{\boldsymbol{W}}(\boldsymbol{k}, t)\tilde{\boldsymbol{\phi}}(\boldsymbol{k})\ , \tag{7}$$

which means that it only depends on $\tilde{\boldsymbol{\phi}}$ at $\boldsymbol{k}$. More broadly, we find it useful to view locality and translation equivariance as *dual* to one another (e.g., spatial locality implies dispersion in Fourier space, and vice versa). This insight is well-known in quantum mechanics (Shankar, 2012; Griffiths & Schroeter, 2018), where it is the main content of the uncertainty principle (Heisenberg, 1927).

3 Sampling as spatial pattern formation

Since Turing (1952), it has been well-known that simple reaction-diffusion equations can give rise to complex spatial patterns (Cross & Hohenberg, 1993). A major reason we think field-theoretic diffusion models are worth studying is that, under certain conditions, **sampling can be viewed as a spatial pattern formation process**, and in particular, it is governed by similar PDEs. Consider, for example, a local linear denoiser $\boldsymbol{D}_t$ which depends only on low-order derivatives of $\boldsymbol{\phi}$:

$$\boldsymbol{D}_t[\boldsymbol{\phi}](\boldsymbol{x}) = \boldsymbol{b}(\boldsymbol{x}, t) + \boldsymbol{W}(\boldsymbol{x}, t)\boldsymbol{\phi}(\boldsymbol{x}) + \sum_i \boldsymbol{G}_i(\boldsymbol{x}, t)\partial_i\boldsymbol{\phi}(\boldsymbol{x}) + \frac{\xi}{2}\sum_{i,j}\boldsymbol{H}_{ij}(\boldsymbol{x}, t)\partial^2_{ij}\boldsymbol{\phi}(\boldsymbol{x})\ . \tag{8}$$

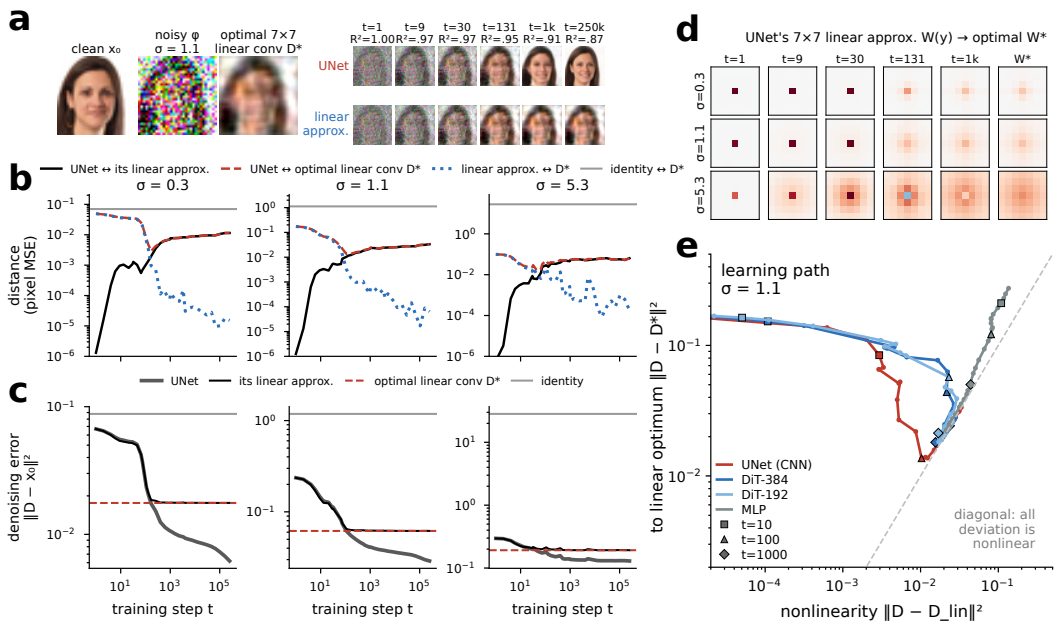


Figure 1: **Early in training, a CNN denoiser is approximately local and linear.** FFHQ 32×32, EDM UNet. (**a**) Outputs at $\sigma = 1.1$ of the UNet (top) and its best-fitting 7×7 linear convolutional model (*conv*; bottom) over training step t. Left: clean image, noisy input, and the data-optimal 7×7 linear conv D^*. (**b**) Distance from the UNet to its linear approximation (black), from the UNet to D^* (red), and from the linear approximation to D^* (blue). The UNet stays nearly linear while approaching D^*. Its linear part becomes D by $t \approx 100$. (**c**) Denoising error. The linear part reaches the D error, far below the identity map, and further gains are nonlinear. (**d**) Learned 7×7 filter vs. the optimal W^*. (**e**) Learning paths at $\sigma = 1.1$: the UNet reaches D^* while staying linear, whereas diffusion transformers (DiTs) and the multilayer perceptron (MLP) turn nonlinear earlier.

for some $\boldsymbol{b}$, $\boldsymbol{W}$, $\boldsymbol{G}$, and $\boldsymbol{H}$. Probability-flow sampling (Song et al., 2021) yields dynamics

$$\frac{\partial \boldsymbol{\phi}_t(\boldsymbol{x})}{\partial t} = -\frac{g_t^2 \alpha_t}{2\sigma_t^2}\boldsymbol{b}(\boldsymbol{x},t) - \boldsymbol{Q}(\boldsymbol{x},t)\boldsymbol{\phi}_t(\boldsymbol{x}) - \frac{g_t^2 \alpha_t}{2\sigma_t^2}\sum_i \boldsymbol{G}_i(\boldsymbol{x},t)\partial_i \boldsymbol{\phi}_t(\boldsymbol{x}) - \frac{g_t^2 \alpha_t}{2\sigma_t^2}\frac{\xi}{2}\sum_{i,j}\boldsymbol{H}_{ij}(\boldsymbol{x},t)\partial_{ij}^2 \boldsymbol{\phi}_t(\boldsymbol{x})$$

where we use $\boldsymbol{Q}(\boldsymbol{x},t)$ to abbreviate a longer, $\boldsymbol{W}$-dependent expression. Importantly, the components of $\boldsymbol{\phi}$ satisfy *reaction-diffusion equations*, so sampling defines a spatial pattern formation process. Their interpretation is as follows. Instead of having 'chemicals', we have 'colors'; each field component represents a different color channel of the final sample. These components can receive space- and time-dependent *input* (the first term); can be *suppressed or amplified* depending on their current values (the $\boldsymbol{\phi}_t$ term); can be *transported* at particular speeds in particular directions (the $\partial_i \boldsymbol{\phi}_t$ terms); and *spread throughout space*, possibly faster in some areas and slower in others (the $\partial_{ij}^2 \boldsymbol{\phi}_t$ terms). If the matrices $\boldsymbol{W}$, $\boldsymbol{G}$, and $\boldsymbol{H}$ are not diagonal, the different color channels are not independent, but can mix. They can interact in more nontrivial ways, including ones that recall Turing's original model of diffusion-driven instability (Turing, 1952), if the denoiser includes a term that depends nonlinearly on $\boldsymbol{\phi}$. Such terms represent the 'reactions' of reaction-diffusion dynamics.

Similarly, near-equivariant linear denoisers have the form

$$\boldsymbol{D}_t[\boldsymbol{\phi}](\boldsymbol{x}) = \boldsymbol{b}(\boldsymbol{x},t) + \int [\boldsymbol{W}(\boldsymbol{x}-\boldsymbol{y},t) + \sum_i \boldsymbol{G}_i(\boldsymbol{x}-\boldsymbol{y},t)y_i + \frac{\xi}{2}\sum_{i,j}\boldsymbol{H}_{ij}(\boldsymbol{x}-\boldsymbol{y},t)y_i y_j]\boldsymbol{\phi}(\boldsymbol{y})d\boldsymbol{y} \quad (9)$$

and their (spatial) *Fourier transform* has the same form as Eq. 8, except that the corresponding derivatives are with respect to the Fourier variable $\boldsymbol{k}$. This again produces a reaction-diffusion equation, except in Fourier space rather than real (pixel) space.

Sampling dynamics do not generally correspond *exactly* to reaction-diffusion dynamics for a few reasons. First, although locality and translation equivariance are significant inductive biases of CNN-

based architectures, real diffusion models are neither exactly local nor exactly translation equivariant (Fig. 1). Receptive fields (which can be estimated by numerically computing the gradient of learned denoisers with respect to their input) of real denoisers grow throughout training and can be large at high noise scales (Niedoba et al., 2024; Lukoianov et al., 2025) (Fig. 1d). Hence, spatial locality and equivariance are fairly special—they are more likely to hold for CNN-based architectures, early in training, at moderate-to-low noise scales. Second, sampling dynamics involves explicit time/noise-scale-dependence that makes the corresponding PDEs more complex than standard reaction-diffusion equations. Nonetheless, we find the connection between field-theoretic diffusion models and spatial pattern formation theoretically enlightening, and in the remainder of the paper explore it in the context of simple, analytically tractable examples.

4 Local linear models: modulation, transport, and diffusion

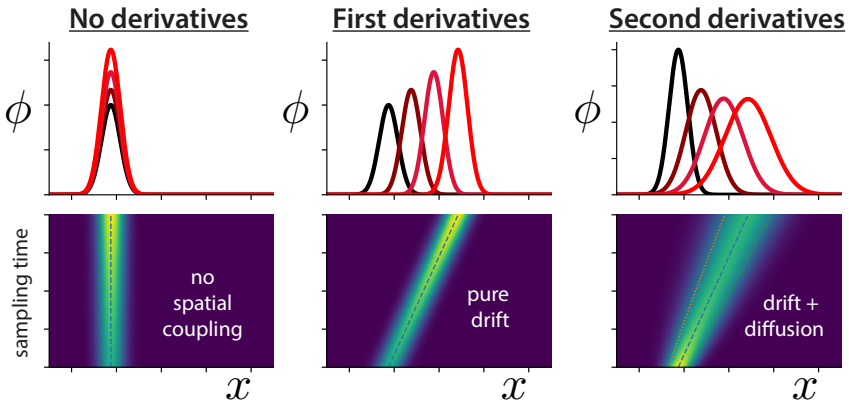


Figure 2: **Local linear denoisers produce amplification, transport, and diffusion in pixel space.** How local the denoiser D_t is affects the extent to which information at one pixel affects another during sampling. Pictured is a localized scalar field evolving under sampling dynamics. **Left:** an ultra-local denoiser yields dynamics that may involve amplification or suppression, but that do not couple different spatial locations. **Center:** if the denoiser depends on the first derivative of ϕ, information at one pixel can be 'transported' to affect another. **Right:** if the denoiser depends on first and second derivatives of ϕ, information at one pixel may be both 'transported' and 'spread'; spreading causes one pixel of the initial noise seed to affect many locations on the final sample.

Sampling dynamics are easiest to mathematically understand when the underlying denoiser is *linear* in the field $\boldsymbol{\phi}$. In the next two sections, we study the sampling dynamics of linear denoisers which are local in either real (pixel) space or Fourier space, which respectively correspond to weak violations of spatial locality and translation equivariance. Our analysis here is similar in spirit to Wang and Vastola's analysis of the Gaussian model (Wang & Vastola, 2023; Wang & Vastola, 2024); in some sense, what we consider here is its spatially local analogue. Although this model class is simple, it has the advantage that it is solvable, which allows us to precisely understand the effects of a strong inductive bias (here, spatial locality) on sampling.

For simplicity, in the next two sections we ignore color-mixing dynamics and treat ϕ as a scalar. One way to do this is via an 'eigencolor' assumption; see Appendix D. Consider a local linear denoiser

$$D_t[\phi](\boldsymbol{x}) = b(\boldsymbol{x},t) + W(\boldsymbol{x},t)\phi(\boldsymbol{x}) + \sum_i G_i(\boldsymbol{x},t)\partial_i\phi(\boldsymbol{x}) + \frac{\xi}{2}\sum_{i,j} H_{ij}(\boldsymbol{x},t)\partial^2_{ij}\phi(\boldsymbol{x}) \tag{10}$$

where b denotes a bias term, and W, the G_i, and the H_{ij} determine how the denoiser depends on ϕ. The scalar $\xi \geq 0$ is assumed to be small, which will allow sampling dynamics to be solved via the method of characteristics (Risken, 1996).

Ultra-local linear denoisers. The simplest version of Eq. 10 has

$$D_t[\phi](\boldsymbol{x}) := b(\boldsymbol{x},t) + W(\boldsymbol{x},t)\phi(\boldsymbol{x})\,. \tag{11}$$

We call this kind of denoiser *ultra-local* because it depends only on ϕ evaluated at the queried spatial location $\boldsymbol{x}$. Because the denoiser involves no coupling of different spatial locations, there is essentially no meaningful spatial component of sampling dynamics; the values of the field at different spatial locations evolve independently. Note that probability-flow sampling dynamics become

$$\frac{\partial \phi_t(\boldsymbol{x})}{\partial t} = -\frac{g_t^2 \alpha_t}{2\sigma_t^2} b(\boldsymbol{x}, t) - \left[\beta_t + \frac{g_t^2}{2\sigma_t^2}\left(\alpha_t W(\boldsymbol{x}, t) - 1\right)\right] \phi_t(\boldsymbol{x}) \,. \tag{12}$$

We essentially have a different first-order ODE for each value of $\boldsymbol{x}$. Solving these in the usual way,

$$\begin{aligned} \phi_t(\boldsymbol{x}) = \frac{\alpha_t}{\alpha_T} \exp\left\{\int_t^T \frac{g_s^2}{2\sigma_s^2}\left(\alpha_s W(\boldsymbol{x}, s) - 1\right) \, ds\right\} \phi_T(\boldsymbol{x}) \\ + \alpha_t \int_t^T \frac{g_s^2}{2\sigma_s^2} \exp\left\{\int_t^s \frac{g_u^2}{2\sigma_u^2}\left(\alpha_u W(\boldsymbol{x}, u) - 1\right) \, du\right\} b(\boldsymbol{x}, s) \, ds \,. \end{aligned} \tag{13}$$

This equation says that, at each location $\boldsymbol{x}$, the value of the final sample is obtained by scaling the initial noise seed $\phi_T(\boldsymbol{x})$, filtering the source term $b(\boldsymbol{x}, t)$, and adding them. The sample at any one location evolves completely independently of the sample at any other location (Fig. 2, left).

First-derivative terms produce transport. Now add a gradient-dependent term, so

$$D_t[\phi](\boldsymbol{x}) := b(\boldsymbol{x}, t) + W(\boldsymbol{x}, t)\phi(\boldsymbol{x}) + \sum_i G_i(\boldsymbol{x}, t)\partial_i \phi(\boldsymbol{x}) \,. \tag{14}$$

This denoiser is less local than the previous one, since the first-order term relates ϕ at one location $\boldsymbol{x}$ to ϕ at a (very close) neighboring location. The resulting sampling PDE is first-order in space and is directly amenable to the method of characteristics (Risken, 1996). Solving it involves the trajectory $\boldsymbol{y}_s$ defined via dynamics

$$\frac{d\boldsymbol{y}_s}{ds} = \frac{g_s^2 \alpha_s}{2\sigma_s^2} \boldsymbol{G}(\boldsymbol{y}_s, s) \tag{15}$$

for $s \in [t, T]$, subject to the boundary condition $\boldsymbol{y}_t = \boldsymbol{x}$. Up to this trajectory, we obtain almost the same solution as before:

$$\begin{aligned} \phi_t(\boldsymbol{x}) = \frac{\alpha_t}{\alpha_T} \exp\left\{\int_t^T \frac{g_s^2}{2\sigma_s^2}\left(\alpha_s W(\boldsymbol{y}_s, s) - 1\right) \, ds\right\} \phi_T(\boldsymbol{y}_T) \\ + \alpha_t \int_t^T \frac{g_s^2}{2\sigma_s^2} \exp\left\{\int_t^s \frac{g_u^2}{2\sigma_u^2}\left(\alpha_u W(\boldsymbol{y}_u, u) - 1\right) \, du\right\} b(\boldsymbol{y}_s, s) \, ds \,. \end{aligned} \tag{16}$$

The main difference is that the solution now depends on $\boldsymbol{x}$ through the curve $\boldsymbol{y}_s(\boldsymbol{x})$. We can interpret it as indicating that the value of the final sample at $\boldsymbol{x}$ depends on the values of the noise seed and source term at locations that are *transported* to $\boldsymbol{x}$ along the characteristic curve (Fig. 2, center).

Second-derivative terms produce diffusion. The general case, where the denoiser depends on second derivatives, can be tackled (assuming the second-derivative term's coefficient ξ is small) using the WKB method (Evans, 2022), which reduces its solution to solving a system of ODEs. For simplicity, we assume no bias term. This time, there are two auxiliary variables ($\boldsymbol{y}_s, \boldsymbol{p}_s \in \mathbb{R}^{N_P}$) which follow Hamiltonian dynamics (Goldstein, 1980) defined by (see Appendix D)

$$H(\boldsymbol{y}, \boldsymbol{p}, t) := \frac{g_t^2 \alpha_t}{2\sigma_t^2} \left[\boldsymbol{G}(\boldsymbol{y}, t)^T \boldsymbol{p} - \frac{1}{2} \boldsymbol{p}^T \boldsymbol{H}(\boldsymbol{y}, t) \boldsymbol{p}\right] \,. \tag{17}$$

The extra degree of freedom represents the fact that the diffusion term allows many different locations to communicate, which causes there to be many initial locations $\boldsymbol{x}_T$ that contribute to the final sample at a given location $\boldsymbol{x}$. The WKB math indicates that we can quantify this influence via

$$q(\boldsymbol{x}_T | \boldsymbol{x}, t) := \frac{\exp\left\{\int_t^T -\frac{1}{\xi} \frac{g_s^2 \alpha_s}{2\sigma_s^2} \frac{1}{2} \boldsymbol{p}_s^T \boldsymbol{H}(\boldsymbol{x}_s, s) \boldsymbol{p}_s \, ds\right\}}{\int \exp\left\{\int_t^T -\frac{1}{\xi} \frac{g_s^2 \alpha_s}{2\sigma_s^2} \frac{1}{2} \boldsymbol{p}_s^T \boldsymbol{H}(\boldsymbol{x}_s, s) \boldsymbol{p}_s \, ds\right\} d\boldsymbol{x}_T} \,. \tag{18}$$

If we define

$$I_t(\boldsymbol{x};\boldsymbol{x}_T) := \int_t^T \beta_s + \frac{g_s^2}{2\sigma_s^2}\left(\alpha_s W(\boldsymbol{x}_s,s) - 1 - \frac{\alpha_s}{2}\sum_{i,j} H_{ij}(\boldsymbol{x}_s,s)\partial_{ij}^2 S_0(\boldsymbol{x}_s,s)\right) ds\ , \quad (19)$$

we can write the full solution, which involves integrating over all possible $\boldsymbol{x}_T$, as

$$\phi_t(\boldsymbol{x}) = \int e^{I_t(\boldsymbol{x};\boldsymbol{x}_T)}\phi_T(\boldsymbol{x}_T)\ q(\boldsymbol{x}_T|\boldsymbol{x},t)\ d\boldsymbol{x}_T\ . \quad (20)$$

Although this solution is extremely complicated, and for the most part implicit, we can make the following observations. First, if $\boldsymbol{H} \to \boldsymbol{0}$ or $\xi \to 0$, the solution reduces to what we found in the previous subsection, since q approaches a delta function. In this limit, the more complex Hamiltonian dynamics reduce to what we found earlier (Eq. 16). Second, for this type of denoiser, any pixel location in the initial noise seed can flow to *any other* pixel location in the final sample, although they are most likely to follow the characteristic curve defined by $\boldsymbol{G}$ (Fig. 2, right). The specific contribution of each possible flow to the final sample is determined by the distribution q.

5 Translation-equivariant linear models as Fourier-space local

Consider a linear denoiser which (up to an arbitrary bias b) is *nearly* translation equivariant, i.e.,

$$D_t[\phi](\boldsymbol{x}) = b(\boldsymbol{x},t) + \int [W(\boldsymbol{x}-\boldsymbol{y},t) + \sum_i G_i(\boldsymbol{x}-\boldsymbol{y},t)y_i + \frac{\xi}{2}\sum_{i,j} H_{ij}(\boldsymbol{x}-\boldsymbol{y},t)y_i y_j]\phi(\boldsymbol{y})d\boldsymbol{y} \quad (21)$$

where W, the G_i, and the H_{ij} determine how the denoiser depends on ϕ. This model is not exactly equivariant, since the function against which $\boldsymbol{\phi}$ is convolved does not depend solely on the difference $\boldsymbol{x}-\boldsymbol{y}$, but it is *almost* equivariant. We find this notion of relaxed equivariance useful, since it shows one can obtain similar behavior without strict equivariance. The last two terms quantify the extent of equivariance-breaking, and for mathematical convenience we again assume the scalar $\xi \geq 0$ is small. As in the previous section, we invoke an 'eigencolor' assumption to consider a scalar field ϕ.

As we pointed out in Sec. 3, this convolutional structure implies that the denoiser is local in wavevector space. More precisely, its (spatial) Fourier transform equals

$$\tilde{D}_t[\phi](\boldsymbol{k}) = \tilde{b}(\boldsymbol{k},t) + \tilde{W}(\boldsymbol{k},t)\tilde{\phi}(\boldsymbol{k}) - i\sum_a \tilde{G}_a(\boldsymbol{k},t)\frac{\partial\tilde{\phi}(\boldsymbol{k})}{\partial k_a} - \frac{\xi}{2}\sum_{a,b}\tilde{H}_{ab}(\boldsymbol{k},t)\frac{\partial^2\tilde{\phi}(\boldsymbol{k},t)}{\partial k_a \partial k_b} \quad (22)$$

where $\tilde{b}$, $\tilde{W}$, $\tilde{G}_i$ and $\tilde{H}_{ij}$ are the Fourier transforms of the bias and coefficients, and $\tilde{\phi}$ is the Fourier transform of the field. Hence, this model is in some sense identical to the local model we considered in the previous section, with the only difference being that it is local in Fourier rather than real space.

Sampling dynamics have the form of an advection-diffusion equation in Fourier space:

$$\begin{aligned}\frac{\partial\tilde{\boldsymbol{\phi}}_t(\boldsymbol{k})}{\partial t} = &-\frac{g_t^2\alpha_t}{2\sigma_t^2}\tilde{b}(\boldsymbol{k},t) - \left[\beta_t + \frac{g_t^2}{2\sigma_t^2}\left(\alpha_t\tilde{\boldsymbol{W}}(\boldsymbol{k},t) - 1\right)\right]\tilde{\boldsymbol{\phi}}_t(\boldsymbol{k}) \\ &+ i\frac{g_t^2\alpha_t}{2\sigma_t^2}\sum_a \tilde{\boldsymbol{G}}_a(\boldsymbol{k},t)\partial_a\tilde{\boldsymbol{\phi}}_t(\boldsymbol{k}) + \frac{g_t^2\alpha_t}{2\sigma_t^2}\frac{\xi}{2}\sum_{a,b}\tilde{\boldsymbol{H}}_{ab}(\boldsymbol{k},t)\partial_{ab}^2\tilde{\boldsymbol{\phi}}_t(\boldsymbol{k})\end{aligned} \quad (23)$$

As expected, the form of this equation is identical to the one for the local linear model, except it is in Fourier space. It can also be solved via the method of characteristics (see Appendix E), and the interpretation of the solution is similar to before: if one thinks in terms of 'wavevector patches' parameterized by a collection of similar $\boldsymbol{k}$ values, each patch of the initial sample's Fourier transform $\tilde{\boldsymbol{\phi}}_T$ flows to a patch of the final sample's Fourier transform $\tilde{\boldsymbol{\phi}}_0$ along the characteristic curves. For the most part, the strong translation-equivariance inductive bias causes **distant wavevector patches of the initial sample to contribute independently to different wavevector patches of the final sample**. But the story is quite different in 'pixel' space, where dynamics may be highly non-local.

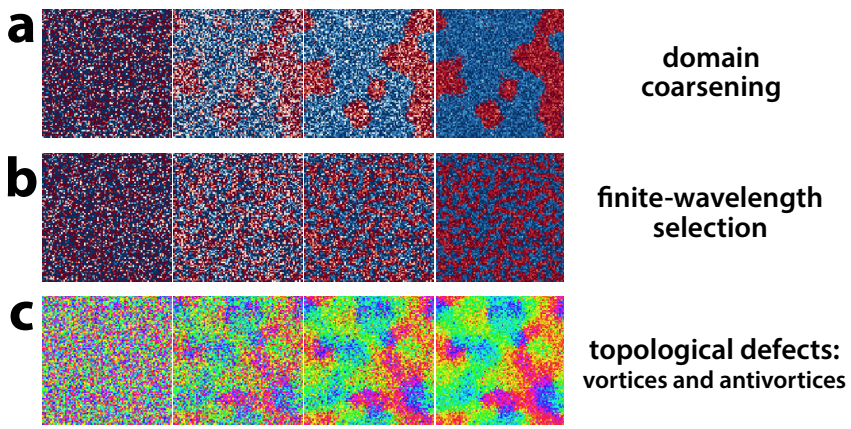


Figure 3: **Nonlinear field-theoretic diffusion models can exhibit canonical pattern-formation phenomena during sampling.** Each row shows a snapshot of the field at a different noise level (left to right: decreasing noise scale). (**a**) For the scalar bimodal model with a ferromagnetic interaction J_{F}, sampling produces symmetry breaking and domain coarsening. Same-sign regions grow as sampling proceeds, causing the number of coherent spatial regions to decrease. (**b**) Replacing J_F with a competing interaction J_{FW} produces finite-wavelength selection. The field develops a labyrinthine pattern with a characteristic spatial scale, and the radially averaged power spectrum becomes concentrated near a specific value. (**c**) For the two-component rotationally symmetric model with ferromagnetic coupling, sampling produces an XY-like phase field containing vortices and antivortices. Colors indicate the local phase $\theta(\mathbf{x}) := \tan^{-1}(\phi_2, \phi_1)$; the vortex density decreases as oppositely charged defects 'annihilate'.

6 Nonlinear models and spatial pattern formation

While interesting, linear denoisers are special. How does sampling change for a denoiser which depends nonlinearly on $\boldsymbol{\phi}$? We now go beyond linear denoisers, whose behavior is well-characterized by existing (discrete) formulations of diffusion models (Wang & Vastola, 2024; Wang & Pehlevan, 2025), and leverage insights from pattern formation theory to comment on nonlinear denoisers.

We make three broad claims and support each with an example. First, nonlinear denoisers can support **symmetry breaking behavior**; second, they can support **finite-wavelength selection**, which is the same mechanism responsible for spot and stripe formation in Turing patterns; and third, **topological defects** are possible when different color channels can interact. For concreteness, consider

$$p_{\mathrm{data}}[\boldsymbol{\phi}_0] \propto \exp\left\{-\int V(\boldsymbol{\phi}_0(\boldsymbol{x}))\, d\boldsymbol{x} + \frac{\epsilon}{2}\iint J(\boldsymbol{x}-\boldsymbol{x}')\boldsymbol{\phi}_0(\boldsymbol{x})^T\boldsymbol{\phi}_0(\boldsymbol{x}')\, d\boldsymbol{x}\, d\boldsymbol{x}'\right\} \tag{24}$$

where $\epsilon \geq 0$ determines the extent to which different spatial locations are coupled through the spatial kernel J. Without the interaction term (i.e., if $\epsilon = 0$), different spatial locations are independent. Let $\boldsymbol{M}_t(\boldsymbol{\phi})$ and $\boldsymbol{C}_t(\boldsymbol{\phi})$ denote the mean and covariance of the corresponding single-location posterior after adding Gaussian noise via the forward process. To first order in ϵ,

$$\boldsymbol{D}_t[\boldsymbol{\phi}](\boldsymbol{x}) = \boldsymbol{M}_t(\boldsymbol{\phi}(\boldsymbol{x})) + \epsilon\, \boldsymbol{C}_t(\boldsymbol{\phi}(\boldsymbol{x}))[J * \boldsymbol{M}_t(\boldsymbol{\phi})](\boldsymbol{x}) + \mathcal{O}(\epsilon^2)\,. \tag{25}$$

We can obtain three familiar pattern-forming systems by making specific choices for V and J.

Symmetry breaking and domain coarsening. For a scalar field, choose

$$V_{\mathrm{sb}}(\phi) = \frac{\phi^2}{2\nu^2} - \log\cosh\left(\frac{m\phi}{\nu^2}\right). \tag{26}$$

Up to normalization, $e^{-V_{\mathrm{sb}}}$ is an equal mixture of Gaussians centered at $\pm m$ with variance ν^2; for $m > \nu$, it is bimodal and explicitly favors two symmetry-related field values. Its pointwise Bayesian denoiser has the small-field expansion

$$M_t(\phi) = a_t\phi - b_t\phi^3 + \mathcal{O}(\phi^5), \qquad b_t = \frac{\sigma_t^2\alpha_t^3 m^4}{3(\sigma_t^2 + \alpha_t^2\nu^2)^4} > 0. \tag{27}$$

To encourage agreement between nearby positions, take a Gaussian interaction

$$J_{\mathrm{F}}(\boldsymbol{x}) = J_0 G_\ell(\boldsymbol{x}), \qquad G_\ell(\boldsymbol{x}) = \frac{e^{-\|\boldsymbol{x}\|^2/(2\ell^2)}}{(2\pi\ell^2)^{N_P/2}}. \tag{28}$$

Since $\widetilde{J}_{\mathrm{F}}(k) = J_0 e^{-\ell^2 k^2/2}$, $J_{\mathrm{F}} * \phi = J_0[\phi + (\ell^2/2)\nabla^2\phi + \cdots]$. Eq. 25 therefore gives

$$D_t[\phi] = A_t\phi - B_t\phi^3 + \kappa_t\nabla^2\phi + \cdots, \qquad \kappa_t > 0, \tag{29}$$

and probability-flow sampling has an *Allen–Cahn/Ginzburg–Landau* form (Allen & Cahn, 1979; Bray, 1994). The bimodal local distribution drives symmetry breaking, while the aligning interaction causes same-sign regions to grow and interfaces to disappear, yielding domain coarsening (Fig. 3a).

Finite-wavelength selection. Keeping exactly the same local distribution, instead choose an interaction with short-range attraction and longer-range repulsion:

$$J_{\mathrm{FW}}(\boldsymbol{x}) = J_0\left[G_\ell(\boldsymbol{x}) - \frac{1}{2}G_{2\ell}(\boldsymbol{x})\right]. \tag{30}$$

Its Fourier transform, $\widetilde{J}_{\mathrm{FW}}(k) = J_0[e^{-\ell^2k^2/2} - \frac{1}{2}e^{-2\ell^2k^2}]$, is maximized at the nonzero wavevector $k_0^2 = \frac{2\log 2}{3\ell^2}$. Near this maximum, $\widetilde{J}_{\mathrm{FW}}(k) = J_* - q(k^2 - k_0^2)^2 + \cdots$ with $q > 0$, so

$$D_t[\phi] = A_t\phi - B_t\phi^3 - q_t(\nabla^2 + k_0^2)^2\phi + \cdots. \tag{31}$$

This is the *Swift–Hohenberg* normal form (Swift & Hohenberg, 1977; Cross & Hohenberg, 1993). Hence, a simple competition between two interaction ranges converts coarsening into finite-wavelength selection, with the model predicting a characteristic spatial scale $2\pi/k_0$ (Fig. 3b).

Interacting channels and topological defects. Choose $N_C = 2$ and the rotationally-symmetric

$$V_{\mathrm{ring}}(\boldsymbol{\phi}) = \frac{\|\boldsymbol{\phi}\|^2}{2\nu^2} - \log I_0\left(\frac{m\|\boldsymbol{\phi}\|}{\nu^2}\right), \tag{32}$$

where I_0 is a modified Bessel function. Equivalently, a clean field value is generated by choosing a direction uniformly on the circle and adding isotropic Gaussian noise around the point $m\boldsymbol{n}$. Hence, the distribution favors a nonzero magnitude but no preferred direction in the two-dimensional color space. Combining this potential with J_F yields a denoiser

$$\boldsymbol{D}_t[\boldsymbol{\phi}] = A_t\boldsymbol{\phi} - B_t\|\boldsymbol{\phi}\|^2\boldsymbol{\phi} + \kappa_t\nabla^2\boldsymbol{\phi} + \cdots, \tag{33}$$

with a vector *Ginzburg–Landau* form. Sampling develops an oriented two-component field containing vortices, around which the local color-space direction winds; the aligning interaction subsequently drives vortex–antivortex annihilation and coarsening (Kosterlitz & Thouless, 1973; Yurke et al., 1993).

7 Discussion

We proposed and studied a continuous-space formulation of diffusion models in which generated samples are fields and probability-flow sampling is characterized by a PDE that describes a spatial pattern formation process. We view our work not as producing a new sampling or training innovation, but rather as providing a theoretically interesting perspective from which to interpret sampling behavior. Moreover, the field-theoretic formulation of diffusion models connects generative models to field theories relevant to physics (e.g., quantum (Schwartz, 2013) and statistical (Kardar, 2007; Zinn-Justin, 2021) field theory) and biological pattern formation (Turing, 1952; Volpert & Petrovskii, 2009; Kondo & Miura, 2010).

Limitations and future work. Our continuum field description is an idealization whose virtues are that it is theoretically interesting (e.g., due to its connections to pattern formation theory) and that it may be useful in interpreting the sampling dynamics of certain nonlinear denoisers. As mentioned in Sec. 3, a local continuum description is more likely to be reasonable for CNN-based architectures, early in training, and at moderate-to-low noise scales. Another limitation of our work is that, for mathematical clarity and simplicity, we have almost exclusively considered idealized models whose behavior is simpler than that of real diffusion models. There is evidence that quite basic (e.g., linear) models can explain diffusion model behavior, especially early in training (Wang & Vastola, 2024), but they are unquestionably simpler than the real thing. In future work, we hope to leverage our field-theoretic formulation to characterize the behavior of more complex nonlinear denoisers.

AI USE STATEMENT

In this work, we used generative AI tools for brainstorming, literature search, to assist with checking the mathematical correctness of our results, and to assist with code generation. AI tools were also used to draft Sec. 6, the most difficult section of the paper. The authors have checked all text and results and take full responsibility for them.

REFERENCES

Samuel M. Allen and John W. Cahn. A microscopic theory for antiphase boundary motion and its application to antiphase domain coarsening. *Acta Metallurgica*, 27(6):1085–1095, 1979. ISSN 0001-6160. doi: https://doi.org/10.1016/0001-6160(79)90196-2. URL https://www.sciencedirect.com/science/article/pii/0001616079901962.

A.J. Bray. Theory of phase-ordering kinetics. *Advances in Physics*, 43(3):357–459, 1994. doi: 10.1080/00018739400101505.

Michael M. Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges. *arXiv e-prints*, art. arXiv:2104.13478, April 2021. doi: 10.48550/arXiv.2104.13478.

Detlev Buchholz. Current trends in axiomatic quantum field theory. In Peter Breitenlohner and Dieter Maison (eds.), *Quantum Field Theory*, pp. 43–64, Berlin, Heidelberg, 2000. Springer Berlin Heidelberg. ISBN 978-3-540-44482-4.

Taco Cohen and Max Welling. Group equivariant convolutional networks. In Maria Florina Balcan and Kilian Q. Weinberger (eds.), *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pp. 2990–2999, New York, New York, USA, 20–22 Jun 2016. PMLR. URL https://proceedings.mlr.press/v48/cohenc16.html.

M. C. Cross and P. C. Hohenberg. Pattern formation outside of equilibrium. *Rev. Mod. Phys.*, 65: 851–1112, Jul 1993. doi: 10.1103/RevModPhys.65.851. URL https://link.aps.org/doi/10.1103/RevModPhys.65.851.

Lawrence C Evans. *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, 2nd edition, 2022.

Herbert Goldstein. *Classical mechanics*. Addison-Wesley, 1980.

David J Griffiths and Darrell F Schroeter. *Introduction to quantum mechanics*. Cambridge University press, 2018.

Martin Hairer. An Introduction to Stochastic PDEs. *arXiv e-prints*, art. arXiv:0907.4178, July 2009. doi: 10.48550/arXiv.0907.4178.

W. Heisenberg. Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Zeitschrift fur Physik*, 43(3-4):172–198, March 1927. doi: 10.1007/BF01397280.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020.

Mason Kamb and Surya Ganguli. An analytic theory of creativity in convolutional diffusion models. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 28795–28831. PMLR, 13–19 Jul 2025. URL https://proceedings.mlr.press/v267/kamb25a.html.

Mehran Kardar. *Statistical Physics of Fields*. Cambridge University Press, 2007.

Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho (eds.), *Advances in Neural Information Processing Systems*, 2022. URL https://openreview.net/forum?id=k7FuTOWMOc7.

Shigeru Kondo and Takashi Miura. Reaction-diffusion model as a framework for understanding biological pattern formation. *Science*, 329(5999):1616–1620, 2010. doi: 10.1126/science.1179047. URL https://www.science.org/doi/abs/10.1126/science.1179047.

J. M. Kosterlitz and D. J. Thouless. Ordering, metastability and phase transitions in two-dimensional systems. *Journal of Physics C: Solid State Physics*, 6(7):1181, April 1973. doi: 10.1088/0022-3719/6/7/010. URL https://doi.org/10.1088/0022-3719/6/7/010.

Finn Lindgren, Håvard Rue, and Johan Lindström. An Explicit Link between Gaussian Fields and Gaussian Markov Random Fields: The Stochastic Partial Differential Equation Approach. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 73(4):423–498, August 2011. ISSN 1369-7412. doi: 10.1111/j.1467-9868.2011.00777.x. URL https://doi.org/10.1111/j.1467-9868.2011.00777.x. _eprint: https://academic.oup.com/jrsssb/article-pdf/73/4/423/49163491/jrsssb_73_4_423.pdf.

Artem Lukoianov, Chenyang Yuan, Justin Solomon, and Vincent Sitzmann. Locality in Image Diffusion Models Emerges from Data Statistics. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 95121–95157. Curran Associates, Inc., 2025. doi: 10.52202/085713-3180. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/88ad9774ffcb7a272457e9396f793a07-Paper-Conference.pdf.

Matthew Niedoba, Berend Zwartsenberg, Kevin Murphy, and Frank Wood. Towards a Mechanistic Explanation of Diffusion Model Generalization. *arXiv e-prints*, art. arXiv:2411.19339, November 2024. doi: 10.48550/arXiv.2411.19339.

Hannes Risken. *The Fokker-Planck Equation: Methods of Solution and Applications*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1996. ISBN 978-3-642-61544-3.

Havard Rue and Leonhard Held. *Gaussian Markov random fields: theory and applications*. Chapman and Hall/CRC, 2005.

Matthew D. Schwartz. *Quantum Field Theory and the Standard Model*. Cambridge University Press, 2013.

R. Shankar. *Principles of Quantum Mechanics*. Springer US, 2012. ISBN 978-1-4757-0576-8. URL https://books.google.com/books?id=sDvrBwAAQBAJ.

Eero P Simoncelli and Bruno A Olshausen. Natural Image Statistics and Neural Representation. *Annual Review of Neuroscience*, 24(Volume 24, 2001):1193–1216, 2001. ISSN 1545-4126. doi: https://doi.org/10.1146/annurev.neuro.24.1.1193. URL https://www.annualreviews.org/content/journals/10.1146/annurev.neuro.24.1.1193. Type: Journal Article.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In Francis Bach and David Blei (eds.), *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 2256–2265, Lille, France, 07–09 Jul 2015. PMLR. URL https://proceedings.mlr.press/v37/sohl-dickstein15.html.

Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper/2019/file/3001ef257407d5a371a96dcd947c7d93-Paper.pdf.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=PxTIG12RRHS.

J. Swift and P. C. Hohenberg. Hydrodynamic fluctuations at the convective instability. *Phys. Rev. A*, 15:319–328, Jan 1977. doi: 10.1103/PhysRevA.15.319. URL https://link.aps.org/doi/10.1103/PhysRevA.15.319.

Alan Mathison Turing. The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 237(641):37–72, 1952. doi: 10.1098/rstb.1952.0012. URL https://royalsocietypublishing.org/doi/abs/10.1098/rstb.1952.0012.

Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural Computation*, 23(7):1661–1674, 2011. doi: 10.1162/NECO_a_00142.

V. Volpert and S. Petrovskii. Reaction–diffusion waves in biology. *Physics of Life Reviews*, 6(4):267–310, 2009. ISSN 1571-0645. doi: https://doi.org/10.1016/j.plrev.2009.10.002. URL https://www.sciencedirect.com/science/article/pii/S1571064509000347.

Binxu Wang and Cengiz Pehlevan. An Analytical Theory of Spectral Bias in the Learning Dynamics of Diffusion Models. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 95865–95963. Curran Associates, Inc., 2025. doi: 10.52202/085713-3207. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/8a0d3f77bb435817807d463c5dcef1ab-Paper-Conference.pdf.

Binxu Wang and John Vastola. The unreasonable effectiveness of gaussian score approximation for diffusion models and its applications. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. URL https://openreview.net/forum?id=I0uknSHM2j.

Binxu Wang and John J. Vastola. Diffusion Models Generate Images Like Painters: an Analytical Theory of Outline First, Details Later. *arXiv e-prints*, art. arXiv:2303.02490, March 2023. doi: 10.48550/arXiv.2303.02490.

B. Yurke, A. N. Pargellis, T. Kovacs, and D. A. Huse. Coarsening dynamics of the xy model. *Phys. Rev. E*, 47:1525–1530, Mar 1993. doi: 10.1103/PhysRevE.47.1525. URL https://link.aps.org/doi/10.1103/PhysRevE.47.1525.

Jean Zinn-Justin. *Quantum field theory and critical phenomena*, volume 171 of *International Series of Monographs on Physics*. Oxford university press, 5th edition, 2021.

A ELEMENTS OF FIELD THEORY: DERIVATIVES, INTEGRALS, AND DISTRIBUTIONS OVER FIELDS

In this appendix, we present a brief review of the mathematics of fields. This review is not intended to be exhaustive, but to give readers unfamiliar with fields a sense of how to work with them.

In this work, we largely treat fields the way a physicist would. For work by physicists covering how to work with fields, see texts on quantum field theory (e.g., Schwartz (2013)) and statistical field theory (e.g., Kardar (2007); Zinn-Justin (2021)). Fields are also discussed, albeit from a somewhat different point of view, in statistics work on Gaussian processes and Gaussian random fields. See Lindgren et al. (2011) for discussion of the relationship between Gaussian random fields and their discrete analogues (Gaussian Markov random fields (Rue & Held, 2005)), as well as references which discuss the use of fields in geostatistics.

Below, we discuss four critical items: (i) What is a field? (ii) How does one differentiate with respect to a field? (iii) How does one integrate with respect to a field? and (iv) How does one define a distribution over fields?

A.1 WHAT IS A FIELD?

For us, a **field** $\boldsymbol{\phi} := (\phi_1, ..., \phi_{N_C})^T$ is a function from $\mathbb{R}^{N_P}$ ('pixel' space) to $\mathbb{R}^{N_C}$ ('color' space). That is, given a 'spatial' input $\boldsymbol{x} \in \mathbb{R}^{N_P}$, $\boldsymbol{\phi}(\boldsymbol{x})$ is a vector in $\mathbb{R}^{N_C}$. If $N_C > 1$, we have a *vector field*; if $N_C = 1$, we have a *scalar field*.

Fields can also be viewed as the continuous limit of infinite-dimensional vectors, and from this perspective it is more obvious how to link them to diffusion models. Suppose $N_P = 1$ and $N_C = 1$, and consider a scalar field ϕ. If we discretize $\mathbb{R}$, ϕ is completely determined by its values on a 1D lattice, i.e.,

$$..., \phi(x_{i-2}), \phi(x_{i-1}), \phi(x_i), \phi(x_{i+1}), \phi(x_{i+2}), ... \tag{34}$$

where, for example, we might have $x_i := i\ \Delta x$ for $i \in \mathbb{Z}$ and some lattice spacing $\Delta x > 0$. Hence, we can consider ϕ as the limit (where we take the lattice spacing to zero) of this infinite-dimensional vector. We can also discretize higher-dimensional pixel spaces into a countably infinite number of lattice points, so the same argument holds for $N_P > 1$. To treat vector fields, we can simply consider block vectors where the vector containing all of the $\phi_{i+1}(\boldsymbol{x}_j)$ follows the vector containing all of the $\phi_i(\boldsymbol{x}_j)$. In all cases, fields are the continuous-space limit of infinite-dimensional vectors.

In what follows, we will go back and forth as necessary between a continuous-space description of fields, and a discrete approximation that treats them as large vectors. There are many mathematical subtleties here, and since our job is already hard enough, we will ignore them. How to rigorously understand the mathematics of (random) fields remains an open question, and relates (for example) to rigorously defining quantum field theories (Buchholz, 2000) and stochastic PDEs (Hairer, 2009).

A.2 FUNCTIONAL DERIVATIVES

Like the components of vectors, we can also take derivatives with respect to fields. For us, such derivatives become relevant in trying to define continuous-space analogues of the Fokker-Planck equation (which characterizes how the distribution of states evolves throughout sampling) and the score function.

Going back to our 1D example, if we view $\phi(x_i)$ as a component of a large vector, it is obvious how to differentiate with respect to it; the quantity

$$\frac{\partial}{\partial \phi(x_i)} F(..., \phi(x_{i-1}), \phi(x_i), \phi(x_{i+1}), ...) \tag{35}$$

is well-defined. For example,

$$\frac{\partial}{\partial \phi(x_i)} \phi(x_j) = \delta_{ij}\ , \tag{36}$$

where δ_{ij} is the Kronecker delta function. By analogy, we can define the derivative with respect to a scalar field ϕ, which we call a **functional derivative**, as

$$\frac{\delta}{\delta \phi(\boldsymbol{x})} F[\phi] := \frac{\partial F[\phi]}{\partial \phi(\boldsymbol{x})}\ . \tag{37}$$

An important example is that

$$\frac{\delta}{\delta\phi(\boldsymbol{x})}\phi(\boldsymbol{y}) = \delta(\boldsymbol{x}-\boldsymbol{y}) \tag{38}$$

where $\delta(\boldsymbol{x}-\boldsymbol{y})$ is the Dirac delta function, which equals zero if $\boldsymbol{x} \neq \boldsymbol{y}$ and infinity otherwise. This generalizes in the obvious way to vector fields. For example,

$$\frac{\delta}{\delta\phi_k(\boldsymbol{x})}\phi_\ell(\boldsymbol{y}) = \delta_{k\ell}\delta(\boldsymbol{x}-\boldsymbol{y})\ . \tag{39}$$

We may write $\frac{\delta}{\delta\boldsymbol{\phi}(\boldsymbol{x})}$, which is analogous to the notation $\frac{\partial}{\partial\boldsymbol{x}}$ sometimes used for the gradient, to ease notation.

A.3 Functional integrals

We can also integrate with respect to fields. To motivate this, consider integrating with respect to the 1D discretization of our scalar field ϕ:

$$\int F[\phi] \prod_{i\in\mathbb{Z}} d\phi(x_i)\ . \tag{40}$$

In the limit as the lattice spacing goes to zero, we have a **functional integral**

$$\int F[\phi]\ \mathcal{D}[\phi]\ . \tag{41}$$

Especially for fields defined on all of Euclidean space, one often runs into issues with integrals like the above being difficult to meaningfully define, because they seem to be formally infinite. An important class of functional integrals which *can* be rigorously defined on all of $\mathbb{R}^{N_P}$ are Gaussian integrals.

Let $J(\boldsymbol{x})$ be a fixed scalar field and $A(\boldsymbol{x};\boldsymbol{x}') > 0$ be a fixed scalar function of two fields. A well-known Gaussian integral has

$$\begin{aligned}&\int \mathcal{D}[\phi]\ \exp\left\{\int -\frac{1}{2}\phi(\boldsymbol{x})A(\boldsymbol{x};\boldsymbol{x}')\phi(\boldsymbol{x}')\ d\boldsymbol{x}d\boldsymbol{x}' + \int J(\boldsymbol{x})\phi(\boldsymbol{x})\ d\boldsymbol{x}\right\}\\ &= \mathcal{N}\exp\left\{\frac{1}{2}\int J(\boldsymbol{x})A^{-1}(\boldsymbol{x};\boldsymbol{x}')J(\boldsymbol{x}')\ d\boldsymbol{x}d\boldsymbol{x}'\right\}\end{aligned} \tag{42}$$

where $\mathcal{N}$ is a certain (formally infinite) constant, and the inverse A^{-1} of A is defined as the scalar field satisfying

$$\int A^{-1}(\boldsymbol{x};\boldsymbol{x}')A(\boldsymbol{x}';\boldsymbol{x}'')\ d\boldsymbol{x}' = \int A(\boldsymbol{x};\boldsymbol{x}')A^{-1}(\boldsymbol{x}';\boldsymbol{x}'')\ d\boldsymbol{x}' = \delta(\boldsymbol{x}-\boldsymbol{x}'')\ . \tag{43}$$

We can define an analogous Gaussian integral over vector fields $\boldsymbol{\phi}$. If $\boldsymbol{J}(\boldsymbol{x})$ is a fixed vector field and $\boldsymbol{A}(\boldsymbol{x};\boldsymbol{x}') > 0$ is a fixed vector function of two fields,

$$\begin{aligned}&\int \mathcal{D}[\phi]\ \exp\left\{\int -\frac{1}{2}\boldsymbol{\phi}(\boldsymbol{x})^T\boldsymbol{A}(\boldsymbol{x};\boldsymbol{x}')\boldsymbol{\phi}(\boldsymbol{x}')\ d\boldsymbol{x}d\boldsymbol{x}' + \int \boldsymbol{J}(\boldsymbol{x})^T\boldsymbol{\phi}(\boldsymbol{x})\ d\boldsymbol{x}\right\}\\ &= \mathcal{N}\exp\left\{\frac{1}{2}\int \boldsymbol{J}(\boldsymbol{x})^T\boldsymbol{A}^{-1}(\boldsymbol{x};\boldsymbol{x}')\boldsymbol{J}(\boldsymbol{x}')\ d\boldsymbol{x}d\boldsymbol{x}'\right\}\end{aligned} \tag{44}$$

where $\mathcal{N}$ is again a formally infinite 'normalizing' constant, and $\boldsymbol{A}^{-1}$ is defined via

$$\int \boldsymbol{A}^{-1}(\boldsymbol{x};\boldsymbol{x}')\boldsymbol{A}(\boldsymbol{x}';\boldsymbol{x}'')\ d\boldsymbol{x}' = \int \boldsymbol{A}(\boldsymbol{x};\boldsymbol{x}')\boldsymbol{A}^{-1}(\boldsymbol{x}';\boldsymbol{x}'')\ d\boldsymbol{x}' = \boldsymbol{I}\delta(\boldsymbol{x}-\boldsymbol{x}'')\ . \tag{45}$$

A.4 Distributions over fields

Being able to define distributions over fields depends on being able to define integrals over fields (so that we can normalize said distributions), so the preceding subsection takes us most of the way there.

The Gaussian integrals we just reported can be used to define field-theoretic analogues of normal distributions. Although the normalization of such distributions involves a factor $\mathcal{N}$ which is formally infinite, this is not usually an issue for two reasons. First, expectations do not involve this constant. Second, when we run into trouble, we can always work in discrete space, and consider some lattice regularization (whose spacing Δx is small but nonzero) of our integrals/distributions/etc.

We will define the field-theoretic analogue of a normal distribution via the density

$$p(\boldsymbol{\phi}) := \frac{\exp\left\{\int -\frac{1}{2}[\boldsymbol{\phi}(\boldsymbol{x}) - \boldsymbol{\mu}(\boldsymbol{x})]^T \boldsymbol{C}^{-1}(\boldsymbol{x};\boldsymbol{x}')[\boldsymbol{\phi}(\boldsymbol{x}') - \boldsymbol{\mu}(\boldsymbol{x}')]\, d\boldsymbol{x} d\boldsymbol{x}'\right\}}{\int \exp\left\{\int -\frac{1}{2}[\boldsymbol{\phi}(\boldsymbol{x}) - \boldsymbol{\mu}(\boldsymbol{x})]^T \boldsymbol{C}^{-1}(\boldsymbol{x};\boldsymbol{x}')[\boldsymbol{\phi}(\boldsymbol{x}') - \boldsymbol{\mu}(\boldsymbol{x}')]\, d\boldsymbol{x} d\boldsymbol{x}'\right\} \mathcal{D}[\boldsymbol{\phi}]} \tag{46}$$

where $\boldsymbol{\mu}(\boldsymbol{x})$ is the field's mean value at $\boldsymbol{x}$, and $\boldsymbol{C}(\boldsymbol{x};\boldsymbol{x}')$ is the covariance function that determines the relatedness of field values at different spatial locations. If $\boldsymbol{\phi}$ has this distribution, we will write $\boldsymbol{\phi} \sim \mathcal{GP}(\boldsymbol{\mu}(\boldsymbol{x}, \boldsymbol{C}(\boldsymbol{x};\boldsymbol{x}'))$.

It is worth noting that 'normally-distributed' random fields have similar properties to normally-distributed random variables. For example, if $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$ are independent normally-distributed random fields and c_1 and c_2 are real-valued constants,

$$c_1\boldsymbol{\phi}_1 + c_2\boldsymbol{\phi}_2 \sim \mathcal{GP}(c_1\boldsymbol{\mu}_1(\boldsymbol{x}) + c_2\boldsymbol{\mu}_2(\boldsymbol{x}), c_1^2\boldsymbol{C}_1(\boldsymbol{x};\boldsymbol{x}') + c_2^2\boldsymbol{C}_2(\boldsymbol{x};\boldsymbol{x}'))\,. \tag{47}$$

We can also define analogues of more complex distributions, like Gaussian mixtures. Gaussian mixtures are defined in the obvious way, e.g.,

$$p(\boldsymbol{\phi}) := \frac{1}{\mathcal{Z}} \sum_{i=1}^{M} w_i \exp\left\{\int -\frac{1}{2}[\boldsymbol{\phi}(\boldsymbol{x}) - \boldsymbol{\mu}_i(\boldsymbol{x})]^T \boldsymbol{C}_i^{-1}(\boldsymbol{x};\boldsymbol{x}')[\boldsymbol{\phi}(\boldsymbol{x}') - \boldsymbol{\mu}_i(\boldsymbol{x}')]\, d\boldsymbol{x} d\boldsymbol{x}'\right\} \tag{48}$$

where $M \geq 1$ is the number of mixture components, the weights w_i sum to one (i.e., $\sum_i w_i = 1$), and $\mathcal{Z}$ is a normalization constant.

Consider, for example, the case that the training set consists of a finite number M of fields $\boldsymbol{\mu}_i$, in which case p_{data} is a mixture of delta functions centered on those fields:

$$p_{data}(\boldsymbol{\phi}) = \frac{1}{\mathcal{Z}} \sum_{i=1}^{M} \delta(\boldsymbol{\phi} - \boldsymbol{\mu}_i)\,. \tag{49}$$

When we run the forward process (see Sec. 2) on this data distribution until a time $t > 0$, we obtain

$$p(\boldsymbol{\phi}|t) = \frac{1}{\mathcal{Z}} \sum_{i=1}^{M} \exp\left\{\int -\frac{1}{2\sigma_t^2}\|\boldsymbol{\phi}(\boldsymbol{x}) - \alpha_t\boldsymbol{\mu}_i(\boldsymbol{x})\|^2\, d\boldsymbol{x}\right\}\,. \tag{50}$$

B DERIVATION OF REVERSE PROCESS IN FIELD-THEORETIC SETTING

In this appendix, we derive the (PF-ODE-like) reverse process associated with the forward process

$$\frac{\partial \boldsymbol{\phi}_t(\boldsymbol{x})}{\partial t} = -\beta_t \boldsymbol{\phi}_t(\boldsymbol{x}) + g_t \boldsymbol{\eta}_t \qquad t = 0 \to t = T \text{ (forward process, } p_{data} \text{ to } p_{noise})$$

which is (see Sec. 2)

$$\frac{\partial \boldsymbol{\phi}_t(\boldsymbol{x})}{\partial t} = -\beta_t \boldsymbol{\phi}_t(\boldsymbol{x}) - \frac{g_t^2(\alpha_t \boldsymbol{D}_t[\phi_t](\boldsymbol{x}) - \boldsymbol{\phi}_t(\boldsymbol{x}))}{2\sigma_t^2} \quad t = T \to t = \epsilon \text{ (reverse process, } p_{noise} \text{ to } p_{data})$$

where $\boldsymbol{D}_t$ is the optimal denoiser. $\boldsymbol{D}_t$ is defined via

$$\frac{\delta \log p(\boldsymbol{\phi}|t)}{\delta \boldsymbol{\phi}(\boldsymbol{x})} = \frac{(\alpha_t \boldsymbol{D}_t[\boldsymbol{\phi}](\boldsymbol{x}) - \boldsymbol{\phi}(\boldsymbol{x}))}{\sigma_t^2} \qquad \boldsymbol{D}_t[\boldsymbol{\phi}](\boldsymbol{x}) = \frac{\boldsymbol{\phi}(\boldsymbol{x}) + \sigma_t^2 \frac{\delta \log p(\boldsymbol{\phi}|t)}{\delta \boldsymbol{\phi}(\boldsymbol{x})}}{\alpha_t} \tag{51}$$

where $\frac{\delta \log p(\boldsymbol{\phi}|t)}{\delta \boldsymbol{\phi}(\boldsymbol{x})}$ is what we call the **score functional**.

B.1 THE FIELD-THEORETIC ANALOGUE OF THE FOKKER-PLANCK EQUATION AND PF-ODE

We will follow the argument from, e.g., Song et al. (2021), which invokes the Fokker-Planck equation (Risken, 1996), a PDE that characterizes how the distribution of states evolves throughout the sampling process. Consider a standard (i.e., non-field-theoretic) forward process

$$\dot{\boldsymbol{x}}_t = \boldsymbol{f}(\boldsymbol{x}_t, t) + g_t \boldsymbol{\eta}_t \tag{52}$$

where $\boldsymbol{f}$ is a drift term, g_t is a time-dependent noise amplitude, and $\boldsymbol{\eta}_t$ is a vector of Gaussian white noise terms. The associated Fokker-Planck equation is

$$\frac{\partial p(\boldsymbol{x}|t)}{\partial t} = -\frac{\partial}{\partial \boldsymbol{x}} \cdot [\boldsymbol{f}(\boldsymbol{x}, t) p(\boldsymbol{x}|t)] + \frac{g_t^2}{2} \nabla_{\boldsymbol{x}}^2 p(\boldsymbol{x}|t) = -\frac{\partial}{\partial \boldsymbol{x}} \cdot \left[\boldsymbol{f}(\boldsymbol{x}, t) p(\boldsymbol{x}|t) - \frac{g_t^2}{2} \frac{\partial}{\partial \boldsymbol{x}} p(\boldsymbol{x}|t) \right] . \tag{53}$$

The continuous-space, field analogue of this equation, following the analogies presented in Appendix A, is

$$\begin{aligned} \frac{\partial p(\boldsymbol{\phi}|t)}{\partial t} &= -\int \frac{\delta}{\delta \boldsymbol{\phi}(\boldsymbol{x})} [\boldsymbol{f}_t[\boldsymbol{\phi}](\boldsymbol{x}) p(\boldsymbol{\phi}|t)] \; d\boldsymbol{x} + \frac{g_t^2}{2} \int \frac{\delta^2}{\delta \boldsymbol{\phi}(\boldsymbol{x})^2} p(\boldsymbol{\phi}|t) d\boldsymbol{x} \\ &= -\int \frac{\delta}{\delta \boldsymbol{\phi}(\boldsymbol{x})} \left[\boldsymbol{f}_t[\boldsymbol{\phi}](\boldsymbol{x}) p(\boldsymbol{\phi}|t) - \frac{g_t^2}{2} \frac{\delta}{\delta \boldsymbol{\phi}(\boldsymbol{x})} p(\boldsymbol{\phi}|t) \right] \; d\boldsymbol{x} \end{aligned} \tag{54}$$

where $\boldsymbol{f}_t[\boldsymbol{\phi}](\boldsymbol{x})$ is a drift function. To obtain a valid process when time is reversed, Eq. 53 indicates that we ought to consider an ODE with a drift term

$$\boldsymbol{f}_{rev}(\boldsymbol{x}, t) := \boldsymbol{f}(\boldsymbol{x}, t) - \frac{g_t^2}{2} \frac{1}{p(\boldsymbol{x}|t)} \frac{\partial}{\partial \boldsymbol{x}} p(\boldsymbol{x}|t) = \boldsymbol{f}(\boldsymbol{x}, t) - \frac{g_t^2}{2} \frac{\partial}{\partial \boldsymbol{x}} \log p(\boldsymbol{x}|t) . \tag{55}$$

Similarly, our field-theoretic Fokker-Planck equation (Eq. 54) says that we can obtain an analogue of the PF-ODE by using the drift term

$$\boldsymbol{f}_{rev,t}[\boldsymbol{\phi}] := \boldsymbol{f}_t[\boldsymbol{\phi}] - \frac{g_t^2}{2} \frac{1}{p(\boldsymbol{\phi}|t)} \frac{\delta}{\delta \boldsymbol{\phi}(\boldsymbol{x})} p(\boldsymbol{\phi}|t) = \boldsymbol{f}_t[\boldsymbol{\phi}] - \frac{g_t^2}{2} \frac{\delta}{\delta \boldsymbol{\phi}(\boldsymbol{x})} \log p(\boldsymbol{\phi}|t) \tag{56}$$

where $\frac{\delta \log p(\boldsymbol{\phi}|t)}{\delta \boldsymbol{\phi}(\boldsymbol{x})}$ is the score functional. This completes the argument.

B.2 MORE ON THE RELATIONSHIP BETWEEN THE DENOISER AND SCORE FUNCTIONAL

Both in the main text and above, we defined the (optimal) denoiser $\boldsymbol{D}_t$ via its relationship with the score functional. We can also define the denoiser in a different way, as the expectation

$$\boldsymbol{D}_t[\boldsymbol{\phi}](\boldsymbol{x}) := \mathbb{E}[\boldsymbol{\phi}_0 | \boldsymbol{\phi}, t] , \tag{57}$$

which is more in line with the denoiser's interpretation as the best (Bayesian) estimate of the noise-free initial sample $\boldsymbol{\phi}_0$ given the noise-corrupted sample $\boldsymbol{\phi}$ and time t. (This interpretation suggests

the *Bayesian guessing game* interpretation of denoising and sampling dynamics; see, e.g., Kamb & Ganguli (2025).)

If we define the denoiser this way, we can derive its relationship with the score functional. Note,

$$\begin{aligned}\boldsymbol{D}_t[\boldsymbol{\phi}](\boldsymbol{x}) &= \int \boldsymbol{\phi}_0 \, p(\boldsymbol{\phi}_0|\boldsymbol{\phi}, t) \, \mathcal{D}[\boldsymbol{\phi}_0] \\ &= \int \boldsymbol{\phi}_0 \, \frac{p(\boldsymbol{\phi}|\boldsymbol{\phi}_0, t) p_{data}(\boldsymbol{\phi}_0)}{p(\boldsymbol{\phi}|t)} \, \mathcal{D}[\boldsymbol{\phi}_0] \\ &= \frac{1}{p(\boldsymbol{\phi}|t)} \int \boldsymbol{\phi}_0 \, \frac{e^{-\int \frac{1}{2\sigma_t^2} \|\boldsymbol{\phi}(\boldsymbol{x}) - \alpha_t \boldsymbol{\phi}_0(\boldsymbol{x})\|^2 \, d\boldsymbol{x}}}{\mathcal{Z}} \, p_{data}(\boldsymbol{\phi}_0) \, \mathcal{D}[\boldsymbol{\phi}_0]\end{aligned} \tag{58}$$

where $\mathcal{Z}$ is a normalization factor. From the definition of $p(\boldsymbol{\phi}|\boldsymbol{\phi}_0, t)$, we can show that

$$\begin{aligned}\frac{\delta}{\delta \boldsymbol{\phi}(\boldsymbol{x})} \log p(\boldsymbol{\phi}|\boldsymbol{\phi}_0, t) &= -\int \frac{\delta}{\delta \boldsymbol{\phi}(\boldsymbol{x})} \frac{1}{2\sigma_t^2} \|\boldsymbol{\phi}(\boldsymbol{y}) - \alpha_t \boldsymbol{\phi}_0(\boldsymbol{y})\|^2 \, d\boldsymbol{y} \\ &= \int \frac{1}{\sigma_t^2} [\alpha_t \boldsymbol{\phi}_0(\boldsymbol{y}) - \boldsymbol{\phi}(\boldsymbol{y})] \delta(\boldsymbol{x} - \boldsymbol{y}) \, d\boldsymbol{y} \\ &= \frac{1}{\sigma_t^2} [\alpha_t \boldsymbol{\phi}_0(\boldsymbol{x}) - \boldsymbol{\phi}(\boldsymbol{x})] \, .\end{aligned} \tag{59}$$

Using this fact, we find

$$\begin{aligned}\boldsymbol{D}_t[\boldsymbol{\phi}](\boldsymbol{x}) &= \frac{1}{p(\boldsymbol{\phi}|t)} \int \left[\frac{\sigma_t^2 \frac{\delta}{\delta \boldsymbol{\phi}} \log p(\boldsymbol{\phi}|\boldsymbol{\phi}_0, t) + \boldsymbol{\phi}}{\alpha_t} \right] \, p(\boldsymbol{\phi}|\boldsymbol{\phi}_0, t) p_{data}(\boldsymbol{\phi}_0) \, \mathcal{D}[\boldsymbol{\phi}_0] \\ &= \frac{1}{p(\boldsymbol{\phi}|t)} \frac{1}{\alpha_t} \int \left[\sigma_t^2 \frac{\delta}{\delta \boldsymbol{\phi}} p(\boldsymbol{\phi}|\boldsymbol{\phi}_0, t) \right] \, p_{data}(\boldsymbol{\phi}_0) \, \mathcal{D}[\boldsymbol{\phi}_0] + \frac{1}{\alpha_t} \boldsymbol{\phi}(\boldsymbol{x}) \\ &= \frac{1}{p(\boldsymbol{\phi}|t)} \frac{\sigma_t^2}{\alpha_t} \frac{\delta}{\delta \boldsymbol{\phi}(\boldsymbol{x})} p(\boldsymbol{\phi}|t) + \frac{1}{\alpha_t} \boldsymbol{\phi}(\boldsymbol{x}) \\ &= \frac{\sigma_t^2}{\alpha_t} \frac{\delta}{\delta \boldsymbol{\phi}(\boldsymbol{x})} \log p(\boldsymbol{\phi}|t) + \frac{1}{\alpha_t} \boldsymbol{\phi}(\boldsymbol{x}) \, ,\end{aligned} \tag{60}$$

which is the desired relationship.

C GLOBAL MINIMUM OF DENOISING OBJECTIVE

In this appendix, we show that the global minimum of the denoising objective

$$\begin{aligned} J(\boldsymbol{\theta}) &= \int \lambda_t \left[\int \frac{\|\hat{\boldsymbol{D}}_{\boldsymbol{\theta},t}[\boldsymbol{\phi}](\boldsymbol{x}) - \boldsymbol{\phi}_0(\boldsymbol{x})\|_2^2}{2} d\boldsymbol{x} \right] \; p(\boldsymbol{\phi}|\boldsymbol{\phi}_0, t)\; p_{data}(\boldsymbol{\phi}_0)\; p(t)\; \mathcal{D}[\boldsymbol{\phi}]\mathcal{D}[\boldsymbol{\phi}_0] dt \\ &= \mathbb{E}_{\boldsymbol{\phi},\boldsymbol{\phi}_0,t} \left\{ \frac{\lambda_t}{2} \int \|\hat{\boldsymbol{D}}_{\boldsymbol{\theta},t}[\boldsymbol{\phi}](\boldsymbol{x}) - \boldsymbol{\phi}_0(\boldsymbol{x})\|_2^2 d\boldsymbol{x} \right\} \end{aligned} \tag{61}$$

is the 'optimal' denoiser $\boldsymbol{D}_t[\boldsymbol{\phi}](\boldsymbol{x}) := \mathbb{E}[\boldsymbol{\phi}_0|\boldsymbol{\phi}, t]$ (see Appendix B).

First, note that J can be rewritten as

$$J(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\phi},\boldsymbol{\phi}_0,t} \left\{ \frac{\lambda_t}{2} \int \|\hat{\boldsymbol{D}}_{\boldsymbol{\theta},t}[\boldsymbol{\phi}](\boldsymbol{x}) - \boldsymbol{D}_t[\boldsymbol{\phi}](\boldsymbol{x}) + \boldsymbol{D}_t[\boldsymbol{\phi}](\boldsymbol{x}) - \boldsymbol{\phi}_0(\boldsymbol{x})\|_2^2 d\boldsymbol{x} \right\}, \tag{62}$$

which can be decomposed into three terms:

$$\begin{aligned} &\mathbb{E}_{\boldsymbol{\phi},\boldsymbol{\phi}_0,t} \left\{ \frac{\lambda_t}{2} \int \|\hat{\boldsymbol{D}}_{\boldsymbol{\theta},t}[\boldsymbol{\phi}](\boldsymbol{x}) - \boldsymbol{D}_t[\boldsymbol{\phi}](\boldsymbol{x})\|_2^2 d\boldsymbol{x} \right\} \\ &+ 2\, \mathbb{E}_{\boldsymbol{\phi},\boldsymbol{\phi}_0,t} \left\{ \frac{\lambda_t}{2} \int [\hat{\boldsymbol{D}}_{\boldsymbol{\theta},t}[\boldsymbol{\phi}](\boldsymbol{x}) - \boldsymbol{D}_t[\boldsymbol{\phi}](\boldsymbol{x})] \cdot [\boldsymbol{D}_t[\boldsymbol{\phi}](\boldsymbol{x}) - \boldsymbol{\phi}_0(\boldsymbol{x})] d\boldsymbol{x} \right\} \\ &+ \mathbb{E}_{\boldsymbol{\phi},\boldsymbol{\phi}_0,t} \left\{ \frac{\lambda_t}{2} \int \|\boldsymbol{D}_t[\boldsymbol{\phi}](\boldsymbol{x}) - \boldsymbol{\phi}_0(\boldsymbol{x})\|_2^2 d\boldsymbol{x} \right\}. \end{aligned} \tag{63}$$

The third term does not depend on $\boldsymbol{\theta}$, and hence does not affect our optimization, so we can ignore it. The second term equals

$$\begin{aligned} &\mathbb{E}_{\boldsymbol{\phi},\boldsymbol{\phi}_0,t} \left\{ \lambda_t \int [\hat{\boldsymbol{D}}_{\boldsymbol{\theta},t}[\boldsymbol{\phi}](\boldsymbol{x}) - \boldsymbol{D}_t[\boldsymbol{\phi}](\boldsymbol{x})] \cdot [\boldsymbol{D}_t[\boldsymbol{\phi}](\boldsymbol{x}) - \boldsymbol{\phi}_0(\boldsymbol{x})] d\boldsymbol{x} \right\} \\ &= \int \lambda_t \left(\int [\hat{\boldsymbol{D}}_{\boldsymbol{\theta},t}[\boldsymbol{\phi}](\boldsymbol{x}) - \boldsymbol{D}_t[\boldsymbol{\phi}](\boldsymbol{x})] \cdot [\boldsymbol{D}_t[\boldsymbol{\phi}](\boldsymbol{x}) - \boldsymbol{\phi}_0(\boldsymbol{x})] d\boldsymbol{x} \right) \; p(\boldsymbol{\phi}|\boldsymbol{\phi}_0, t)\; p_{data}(\boldsymbol{\phi}_0)\; p(t)\; \mathcal{D}[\boldsymbol{\phi}]\mathcal{D}[\boldsymbol{\phi}_0] dt \\ &= \int \lambda_t \left(\int [\hat{\boldsymbol{D}}_{\boldsymbol{\theta},t}[\boldsymbol{\phi}](\boldsymbol{x}) - \boldsymbol{D}_t[\boldsymbol{\phi}](\boldsymbol{x})] \cdot [\boldsymbol{D}_t[\boldsymbol{\phi}](\boldsymbol{x}) - \boldsymbol{\phi}_0(\boldsymbol{x})] d\boldsymbol{x} \right) \; p(\boldsymbol{\phi}_0|\boldsymbol{\phi}, t)\; p(\boldsymbol{\phi}|t)\; p(t)\; \mathcal{D}[\boldsymbol{\phi}]\mathcal{D}[\boldsymbol{\phi}_0] dt \\ &= 0\,, \end{aligned}$$

since the only factor that depends on $\boldsymbol{\phi}_0$ is $[\boldsymbol{D}_t[\boldsymbol{\phi}](\boldsymbol{x}) - \boldsymbol{\phi}_0(\boldsymbol{x})]$, and since $\boldsymbol{D}_t[\boldsymbol{\phi}](\boldsymbol{x}) := \mathbb{E}[\boldsymbol{\phi}_0|\boldsymbol{\phi}, t]$ by definition. This means that

$$\begin{aligned} J(\boldsymbol{\theta}) &= \mathbb{E}_{\boldsymbol{\phi},\boldsymbol{\phi}_0,t} \left\{ \frac{\lambda_t}{2} \int \|\hat{\boldsymbol{D}}_{\boldsymbol{\theta},t}[\boldsymbol{\phi}](\boldsymbol{x}) - \boldsymbol{D}_t[\boldsymbol{\phi}](\boldsymbol{x})\|_2^2 d\boldsymbol{x} \right\} + \text{const.} \\ &= \mathbb{E}_{\boldsymbol{\phi},t} \left\{ \frac{\lambda_t}{2} \int \|\hat{\boldsymbol{D}}_{\boldsymbol{\theta},t}[\boldsymbol{\phi}](\boldsymbol{x}) - \boldsymbol{D}_t[\boldsymbol{\phi}](\boldsymbol{x})\|_2^2 d\boldsymbol{x} \right\} + \text{const.} \end{aligned} \tag{64}$$

The part which depends on $\boldsymbol{\theta}$ has a global minimum when $\hat{\boldsymbol{D}}_{\boldsymbol{\theta},t} = \boldsymbol{D}_t$, since it cannot be smaller than zero. Hence, if the denoiser model $\hat{\boldsymbol{D}}_{\boldsymbol{\theta},t}$ used is sufficiently expressive, the global minimum of J corresponds to the optimal denoiser.

D SAMPLING DYNAMICS OF LOCAL LINEAR MODELS

In this appendix, we solve the sampling dynamics of models whose denoisers have the form

$$\boldsymbol{D}_t[\boldsymbol{\phi}](\boldsymbol{x}) = \boldsymbol{b}(\boldsymbol{x},t) + \boldsymbol{W}(\boldsymbol{x},t)\boldsymbol{\phi}(\boldsymbol{x}) + \sum_i \boldsymbol{G}_i(\boldsymbol{x},t)\partial_i\boldsymbol{\phi}(\boldsymbol{x}) + \frac{\xi}{2}\sum_{i,j}\boldsymbol{H}_{ij}(\boldsymbol{x},t)\partial^2_{ij}\boldsymbol{\phi}(\boldsymbol{x}) \tag{65}$$

where we leave multiplication over the color indices implicit, and where the parameter $\xi \geq 0$ that controls the degree of spatial locality is assumed to be small. More explicitly, we have

$$D^c_t[\boldsymbol{\phi}](\boldsymbol{x}) = b^c(\boldsymbol{x},t) + \sum_{c'} W^{cc'}(\boldsymbol{x},t)\phi^{c'}(\boldsymbol{x}) + \sum_{c',i} G^{cc'}_i(\boldsymbol{x},t)\partial_i\phi^{c'}(\boldsymbol{x}) + \frac{\xi}{2}\sum_{c',i,j} H^{cc'}_{ij}(\boldsymbol{x},t)\partial^2_{ij}\phi^{c'}(\boldsymbol{x})\,. \tag{66}$$

For this denoiser, sampling dynamics have the form

$$\frac{\partial\boldsymbol{\phi}_t(\boldsymbol{x})}{\partial t} = -\beta_t\boldsymbol{\phi}_t(\boldsymbol{x}) - \frac{g_t^2}{2\sigma_t^2}\left[\alpha_t\boldsymbol{D}_t[\boldsymbol{\phi}_t](\boldsymbol{x}) - \boldsymbol{\phi}_t(\boldsymbol{x})\right] \tag{67}$$

$$= -\frac{g_t^2\alpha_t}{2\sigma_t^2}\boldsymbol{b}(\boldsymbol{x},t) - \left[\beta_t + \frac{g_t^2}{2\sigma_t^2}\left(\alpha_t\boldsymbol{W}(\boldsymbol{x},t) - 1\right)\right]\boldsymbol{\phi}_t(\boldsymbol{x}) \tag{68}$$

$$- \frac{g_t^2\alpha_t}{2\sigma_t^2}\sum_i \boldsymbol{G}_i(\boldsymbol{x},t)\partial_i\boldsymbol{\phi}_t(\boldsymbol{x}) - \frac{g_t^2\alpha_t}{2\sigma_t^2}\frac{\xi}{2}\sum_{i,j}\boldsymbol{H}_{ij}(\boldsymbol{x},t)\partial^2_{ij}\boldsymbol{\phi}_t(\boldsymbol{x})\,.$$

In order to solve these sampling dynamics analytically, and hence write the final sample $\boldsymbol{\phi}_0$ in terms of the initial noise seed $\boldsymbol{\phi}_T$, we must confront two types of mathematical difficulties. The first is that we have to solve a PDE which is first order in time and second order in space. As long as ξ is small, this type of PDE can be solved using the method of characteristics (Evans, 2022), or equivalently the WKB method (Risken, 1996), which reduces the problem of solving a PDE to the problem of solving a system of ODEs.

The second difficulty is that, since the denoiser is not necessarily diagonal in color space, we actually have a *system* of PDEs—one for each color—that interact via $\boldsymbol{W}$, $\boldsymbol{G}$, and $\boldsymbol{H}$.

D.1 THE EIGENCOLOR ASSUMPTION

To address the second difficulty, we make the following **eigencolor assumption**: there exists a color by color ($N_C \times N_C$) matrix $\boldsymbol{Q}$ that diagonalizes each matrix in color space. Concretely, suppose the original matrices, which we relabel as $\tilde{\boldsymbol{W}}$, $\tilde{\boldsymbol{G}}$, and $\tilde{\boldsymbol{H}}$, can be written

$$\tilde{\boldsymbol{W}}(\boldsymbol{x},t) = \boldsymbol{Q}\boldsymbol{W}(\boldsymbol{x},t)\boldsymbol{Q}^{-1} \qquad \tilde{\boldsymbol{G}}_i(\boldsymbol{x},t) = \boldsymbol{Q}\boldsymbol{G}_i(\boldsymbol{x},t)\boldsymbol{Q}^{-1} \qquad \tilde{\boldsymbol{H}}_{ij}(\boldsymbol{x},t) = \boldsymbol{Q}\boldsymbol{H}_{ij}(\boldsymbol{x},t)\boldsymbol{Q}^{-1} \tag{69}$$

for all values of $\boldsymbol{x}$ and t. Given this assumption, we can treat each color independently. Slightly abusing notation, we can just write the denoiser as

$$D_t[\phi](\boldsymbol{x}) = b(\boldsymbol{x},t) + W(\boldsymbol{x},t)\phi(\boldsymbol{x}) + \sum_i G_i(\boldsymbol{x},t)\partial_i\phi(\boldsymbol{x}) + \frac{\xi}{2}\sum_{i,j} H_{ij}(\boldsymbol{x},t)\partial^2_{ij}\phi(\boldsymbol{x})\,, \tag{70}$$

i.e., we can act as if there is only one color, so that D_t and ϕ are scalar fields. (Note that we have also relabeled the transformed $\boldsymbol{b}$ for the eigencolor of interest as just b.) Sampling dynamics becomes

$$\frac{\partial\phi_t(\boldsymbol{x})}{\partial t} = -\beta_t\phi_t(\boldsymbol{x}) - \frac{g_t^2}{2\sigma_t^2}\left[\alpha_t D_t[\phi_t](\boldsymbol{x}) - \phi_t(\boldsymbol{x})\right] \tag{71}$$

$$= -\frac{g_t^2\alpha_t}{2\sigma_t^2}b(\boldsymbol{x},t) - \left[\beta_t + \frac{g_t^2}{2\sigma_t^2}\left(\alpha_t W(\boldsymbol{x},t) - 1\right)\right]\phi_t(\boldsymbol{x}) \tag{72}$$

$$- \frac{g_t^2\alpha_t}{2\sigma_t^2}\sum_i G_i(\boldsymbol{x},t)\partial_i\phi_t(\boldsymbol{x}) - \frac{g_t^2\alpha_t}{2\sigma_t^2}\frac{\xi}{2}\sum_{i,j} H_{ij}(\boldsymbol{x},t)\partial^2_{ij}\phi_t(\boldsymbol{x})\,.$$

Given fields $\phi^c_{eig}(\boldsymbol{x})$ associated with each eigencolor q, we can always obtain the solution for one of the original colors c via

$$\phi^c_{OG}(\boldsymbol{x}) = \sum_q Q^{cq}\phi^q_{eig}(\boldsymbol{x})\,. \tag{73}$$

Taking this fact for granted, in the rest of this appendix we work exclusively with scalar fields and denoisers, which model sampling dynamics for a single eigencolor.

Since deriving the solution of Eq. 71 is somewhat complex, we will illustrate how to solve simpler versions of sampling dynamics first.

D.2 Ultra-local denoiser

The simplest version of Eq. 65 has

$$D_t[\phi](\boldsymbol{x}) := b(\boldsymbol{x}, t) + W(\boldsymbol{x}, t)\phi(\boldsymbol{x}) \tag{74}$$

where we remind the reader that we can treat D_t and ϕ as scalars via the eigencolor assumption. We will call this kind of denoiser *ultra-local*, since (ignoring the source term) it depends only on ϕ evaluated at the queried spatial location $\boldsymbol{x}$. Because the denoiser involves no coupling of different spatial locations, there is essentially no meaningful spatial component of sampling dynamics; the values of the field at different spatial locations evolve independently.

To see this explicitly, note that Eq. 67 becomes

$$\frac{\partial \phi_t(\boldsymbol{x})}{\partial t} = -\frac{g_t^2 \alpha_t}{2\sigma_t^2} b(\boldsymbol{x}, t) - \left[\beta_t + \frac{g_t^2}{2\sigma_t^2}\left(\alpha_t W(\boldsymbol{x}, t) - 1\right)\right] \phi_t(\boldsymbol{x}) \,. \tag{75}$$

We essentially have a different first-order ODE for each value of $\boldsymbol{x}$. Solving these ODEs in the usual way, we find

$$\begin{aligned} \phi_t(\boldsymbol{x}) = \frac{\alpha_t}{\alpha_T} \exp\left\{\int_t^T \frac{g_s^2}{2\sigma_s^2}\left(\alpha_s W(\boldsymbol{x}, s) - 1\right) ds\right\} \phi_T(\boldsymbol{x}) \\ + \alpha_t \int_t^T \frac{g_s^2}{2\sigma_s^2} \exp\left\{\int_t^s \frac{g_u^2}{2\sigma_u^2}\left(\alpha_u W(\boldsymbol{x}, u) - 1\right) du\right\} b(\boldsymbol{x}, s)\, ds \,. \end{aligned} \tag{76}$$

This equation says that the value of the final sample ϕ_0 at location $\boldsymbol{x}$ is simply obtained by scaling the value of the initial noise seed ϕ_T at the same location. Importantly, the final sample at location $\boldsymbol{x}$ *only* depends on the initial sample at the same location.

D.3 Gradient-dependent denoiser

Now add a gradient-dependent term, so that the denoiser has the form

$$D_t[\phi](\boldsymbol{x}) := b(\boldsymbol{x}, t) + W(\boldsymbol{x}, t)\phi(\boldsymbol{x}) + \sum_i G_i(\boldsymbol{x}, t)\partial_i \phi(\boldsymbol{x}) \,. \tag{77}$$

Sampling dynamics now follow

$$\frac{\partial \phi_t(\boldsymbol{x})}{\partial t} = -\frac{g_t^2 \alpha_t}{2\sigma_t^2} b(\boldsymbol{x}, t) - \left[\beta_t + \frac{g_t^2}{2\sigma_t^2}(\alpha_t W(\boldsymbol{x}, t) - 1)\right] \phi_t(\boldsymbol{x}) - \frac{g_t^2 \alpha_t}{2\sigma_t^2} \sum_i G_i(\boldsymbol{x}, t)\partial_i \phi_t(\boldsymbol{x}) \,. \tag{78}$$

Equivalently,

$$\frac{\partial \phi_t(\boldsymbol{x})}{\partial t} + \frac{g_t^2 \alpha_t}{2\sigma_t^2} \sum_i G_i(\boldsymbol{x}, t)\partial_i \phi_t(\boldsymbol{x}) = -\frac{g_t^2 \alpha_t}{2\sigma_t^2} b(\boldsymbol{x}, t) - \left[\beta_t + \frac{g_t^2}{2\sigma_t^2}(\alpha_t W(\boldsymbol{x}, t) - 1)\right] \phi_t(\boldsymbol{x}) \,. \tag{79}$$

This type of PDE is directly amenable to the method of characteristics. Assume that the $\boldsymbol{x}$-dependence of solutions follows a parameterized curve $\boldsymbol{x}_t$. Moreover, writing

$$\frac{d\phi_t(\boldsymbol{x}_t)}{dt} = \frac{\partial \phi_t(\boldsymbol{x}_t)}{\partial t} + \frac{\partial \phi_t(\boldsymbol{x}_t)}{\partial \boldsymbol{x}_t} \cdot \frac{d\boldsymbol{x}_t}{dt} = \frac{\partial \phi_t(\boldsymbol{x}_t)}{\partial t} + \frac{g_t^2 \alpha_t}{2\sigma_t^2} \sum_i G_i(\boldsymbol{x}_t, t)\partial_i \phi_t(\boldsymbol{x}_t) \tag{80}$$

implies that $\boldsymbol{x}_s$ ought to satisfy

$$\frac{d\boldsymbol{x}_s}{ds} = \frac{g_s^2 \alpha_s}{2\sigma_s^2} \boldsymbol{G}(\boldsymbol{x}_s, s) \tag{81}$$

for $s \in [t, T]$, subject to the boundary condition $\boldsymbol{x}_t = \boldsymbol{x}$. (This means, if we want to know the value of the field ϕ_t at location $\boldsymbol{x}$, make sure the characteristic curve ends at that point.) Whether this equation is solvable depends on the specific choice of $\boldsymbol{G}$; for example, it is solvable if $\boldsymbol{G}$ is linear in $\boldsymbol{x}_s$. Regardless, we can write

$$\frac{d\phi_t(\boldsymbol{x}_t)}{dt} = -\frac{g_t^2 \alpha_t}{2\sigma_t^2} b(\boldsymbol{x}, t) - \left[\beta_t + \frac{g_t^2}{2\sigma_t^2}(\alpha_t W(\boldsymbol{x}_t, t) - 1)\right] \phi_t(\boldsymbol{x}_t) \, . \tag{82}$$

This is the same linear ODE we solved earlier, and the answer is the same as before (except now, additional t dependence is smuggled in through the dependence of the $\boldsymbol{x}$ characteristic on t):

$$\begin{aligned} \phi_t(\boldsymbol{x}) = \frac{\alpha_t}{\alpha_T} \exp\left\{\int_t^T \frac{g_s^2}{2\sigma_s^2}\left(\alpha_s W(\boldsymbol{x}_s, s) - 1\right) ds\right\} \phi_T(\boldsymbol{x}_T) \\ + \alpha_t \int_t^T \frac{g_s^2}{2\sigma_s^2} \exp\left\{\int_t^s \frac{g_u^2}{2\sigma_u^2}\left(\alpha_u W(\boldsymbol{x}_s, u) - 1\right) du\right\} b(\boldsymbol{x}_s, s) \, ds \, . \end{aligned} \tag{83}$$

Here, $\boldsymbol{x}_s(\boldsymbol{x})$ denotes the value of the characteristic at time $s \in [t, T]$, given that it takes value $\boldsymbol{x}$ at time t. This solution tells us that the value of the final sample at a given spatial location can depend on the value of the noise seed at multiple spatial locations, since different initial locations can 'flow' to the same final point. In this sense, pixel dynamics are still fairly decoupled, but not totally decoupled. This is more likely to happen if said initial pixel locations are near one another, since solutions of Eq. 81 depend smoothly on their initial condition.

D.4 Laplacian-dependent denoiser

Setup. Now assume that the denoiser depends on ϕ, its first derivative, and its second derivative in a local fashion:

$$D_t[\phi](\boldsymbol{x}) = W(\boldsymbol{x}, t)\phi(\boldsymbol{x}) + \sum_i G_i(\boldsymbol{x}, t)\partial_i \phi(\boldsymbol{x}) + \frac{\xi}{2}\sum_{i,j} H_{ij}(\boldsymbol{x}, t)\partial_{ij}^2 \phi(\boldsymbol{x}) \, . \tag{84}$$

Because this case is already fairly complicated, we will assume that there is no bias term.

Now sampling dynamics has the form of an advection-diffusion equation:

$$\frac{\partial \phi_t(\boldsymbol{x})}{\partial t} = -\left[\beta_t + \frac{g_t^2}{2\sigma_t^2}(\alpha_t W(\boldsymbol{x}, t) - 1)\right] \phi_t(\boldsymbol{x}) - \frac{g_t^2 \alpha_t}{2\sigma_t^2}\sum_i G_i(\boldsymbol{x}, t)\partial_i \phi_t(\boldsymbol{x}) - \frac{g_t^2 \alpha_t}{2\sigma_t^2}\frac{\xi}{2}\sum_{i,j} H_{ij}(\boldsymbol{x}, t)\partial_{ij}^2 \phi_t(\boldsymbol{x}) \, . \tag{85}$$

The method of characteristics can no longer be directly applied, but there is a sense in which it can still be used. If ξ is small, we expect that sampling dynamics should be *approximately* describable in terms of characteristic curves. But the diffusion-like term means that, in principle, any location can flow to *any other* location (although they are most likely to flow to locations near the characteristic curve endpoints).

This motivates considering not just $\phi_t(\boldsymbol{x})$, but the auxiliary field $\phi(\boldsymbol{x}, t; \boldsymbol{x}_T, T)$, which corresponds to $\phi_t(\boldsymbol{x})$ given the specific initial condition $\phi_T(\boldsymbol{x}) = \delta(\boldsymbol{x} - \boldsymbol{x}_T)$. Let us make a WKB-type ansatz

$$\phi(\boldsymbol{x}, t|\boldsymbol{x}_T, T) \approx \frac{\exp(-S_0(\boldsymbol{x}, t; \boldsymbol{x}_T, T)/\xi - S_1(\boldsymbol{x}, t; \boldsymbol{x}_T, T))}{\int e^{-S_0(\boldsymbol{x}, t; \boldsymbol{x}_T, T)/\xi} d\boldsymbol{x}_T} \tag{86}$$

for some functions S_0 and S_1. The denominator helps ensure good behavior in the $\xi \to 0$ limit. We can construct a full solution via

$$\phi_t(\boldsymbol{x}) = \int \phi(\boldsymbol{x}, t|\boldsymbol{x}_T, T) \, d\boldsymbol{x}_T = \int \frac{\exp(-S_0(\boldsymbol{x}, t; \boldsymbol{x}_T, T)/\xi - S_1(\boldsymbol{x}, t; \boldsymbol{x}_T, T))}{\int e^{-S_0(\boldsymbol{x}, t; \boldsymbol{x}_T, T)/\xi} d\boldsymbol{x}_T} \, d\boldsymbol{x}_T \, . \tag{87}$$

We will now proceed by substituting this ansatz into the Fokker-Planck equation to obtain equations for S_0 and S_1 (for a fixed endpoint $\boldsymbol{x}_T$). Since the first and second derivatives of this ansatz are

$$\begin{aligned} \partial_i \phi_t(\boldsymbol{x}) &= -\left(\frac{1}{\xi}\partial_i S_0 + \partial_i S_1\right) \phi_t(\boldsymbol{x}) \\ \partial_{ij}^2 \phi_t(\boldsymbol{x}) &= \left[\left(\frac{1}{\xi}\partial_i S_0 + \partial_i S_1\right)\left(\frac{1}{\xi}\partial_j S_0 + \partial_j S_1\right) - \left(\frac{1}{\xi}\partial_{ij}^2 S_0 + \partial_{ij}^2 S_1\right)\right] \phi_t(\boldsymbol{x}) \\ &\approx \left[\frac{1}{\xi^2}(\partial_i S_0)(\partial_j S_0) + \frac{1}{\xi}(\partial_i S_0 \partial_j S_1 + \partial_j S_0 \partial_i S_1 - \partial_{ij}^2 S_0)\right] \phi_t(\boldsymbol{x}) \, , \end{aligned} \tag{88}$$

where we have dropped the subleading terms (including those generated by the denominator), self-consistency requires that

$$-\frac{\partial S_0}{\partial t} = \frac{g_t^2\alpha_t}{2\sigma_t^2}\boldsymbol{G}^T\nabla S_0 - \frac{g_t^2\alpha_t}{2\sigma_t^2}\frac{1}{2}(\nabla S_0)^T\boldsymbol{H}(\nabla S_0) = \frac{g_t^2\alpha_t}{2\sigma_t^2}\left[\boldsymbol{G}^T\nabla S_0 - \frac{1}{2}(\nabla S_0)^T\boldsymbol{H}(\nabla S_0)\right]$$
$$-\frac{\partial S_1}{\partial t} = -\left[\beta_t + \frac{g_t^2}{2\sigma_t^2}(\alpha_t W - 1)\right] + \frac{g_t^2\alpha_t}{2\sigma_t^2}\boldsymbol{G}^T\nabla S_1 - \frac{g_t^2\alpha_t}{2\sigma_t^2}\frac{1}{2}\sum_{i,j}H_{ij}(\partial_i S_0\partial_j S_1 + \partial_j S_0\partial_i S_1 - \partial_{ij}^2 S_0).$$

Note that the S_1 equation can be further simplified to

$$\begin{aligned}-\frac{\partial S_1}{\partial t} &= -\left[\beta_t + \frac{g_t^2}{2\sigma_t^2}(\alpha_t W - 1)\right] + \frac{g_t^2\alpha_t}{2\sigma_t^2}\boldsymbol{G}^T\nabla S_1 - \frac{g_t^2\alpha_t}{2\sigma_t^2}\left[(\boldsymbol{H}\nabla S_0)^T(\nabla S_1) - \frac{1}{2}\sum_{i,j}H_{ij}\partial_{ij}^2 S_0)\right] \\ &= -\left[\beta_t + \frac{g_t^2}{2\sigma_t^2}\left(\alpha_t W - 1 - \frac{\alpha_t}{2}\sum_{i,j}H_{ij}\partial_{ij}^2 S_0)\right)\right] + \frac{g_t^2\alpha_t}{2\sigma_t^2}\left[\boldsymbol{G} - \boldsymbol{H}\nabla S_0\right]^T\nabla S_1.\end{aligned}$$

We have reduced the Fokker-Planck equation to two coupled PDEs. This does not at first appear to be an improvement, but two features make this reduction useful. First, the equation for S_0 is independent of S_1. Second, the S_0 equation only depends on first derivatives of S_0, and the S_1 equation only depends on first derivatives of S_1. These features suggest the following strategy: solve the S_0 equation using the method of characteristics, and then solve the S_1 equation (using S_0 as input) also using the method of characteristics.

Method of characteristics. The S_0 equation

$$-\frac{\partial S_0}{\partial t} = \frac{g_t^2\alpha_t}{2\sigma_t^2}\left[\boldsymbol{G}^T\nabla S_0 - \frac{1}{2}(\nabla S_0)^T\boldsymbol{H}(\nabla S_0)\right] \tag{89}$$

has the form of a Hamilton-Jacobi equation. According to a standard method of characteristics approach, we can look for solutions along characteristic curves parameterized by t. In particular, we have $\boldsymbol{x}(s)$ and $\boldsymbol{p}(s) := \nabla_{\boldsymbol{x}} S_0(\boldsymbol{x}, s)$ for all $s \in [t, T]$.

The curves follow Hamiltonian dynamics with a Hamiltonian

$$H(\boldsymbol{x}, \boldsymbol{p}, t) := \frac{g_t^2\alpha_t}{2\sigma_t^2}\left[\boldsymbol{G}(\boldsymbol{x}, t)^T\boldsymbol{p} - \frac{1}{2}\boldsymbol{p}^T\boldsymbol{H}(\boldsymbol{x}, t)\boldsymbol{p}\right], \tag{90}$$

which means that

$$\begin{aligned}\frac{\partial H}{\partial \boldsymbol{p}_s} &= \frac{g_s^2\alpha_s}{2\sigma_s^2}\left[\boldsymbol{G}(\boldsymbol{x}_s, s) - \boldsymbol{H}(\boldsymbol{x}_s, s)\boldsymbol{p}_s\right] = \dot{\boldsymbol{x}}_s \\ \frac{\partial H}{\partial \boldsymbol{x}_s} &= \frac{\partial}{\partial \boldsymbol{x}_s}\frac{g_s^2\alpha_s}{2\sigma_s^2}\left[\boldsymbol{G}(\boldsymbol{x}_s, s)^T\boldsymbol{p}_s - \frac{1}{2}\boldsymbol{p}_s^T\boldsymbol{H}(\boldsymbol{x}_s, s)\boldsymbol{p}_s\right] = -\dot{\boldsymbol{p}}_s\,.\end{aligned} \tag{91}$$

As before, these ODEs are generally too complicated to solve analytically—to do so, we would have to make specific assumptions about the $\boldsymbol{x}$-dependence of $\boldsymbol{G}$ and $\boldsymbol{H}$. We can instead take for granted that these ODEs *can* be solved, and write a solution for S_0 and S_1 in terms of them.

S_0 depends on this solution through the relationship

$$\frac{dS_0}{ds} = \boldsymbol{p}_s^T\dot{\boldsymbol{x}}_s - H(\boldsymbol{x}_s, \boldsymbol{p}_s, s) = -\frac{g_s^2\alpha_s}{2\sigma_s^2}\frac{1}{2}\boldsymbol{p}_s^T\boldsymbol{H}(\boldsymbol{x}_s, s)\boldsymbol{p}_s \tag{92}$$

which implies that

$$S_0(t) = S_0(T) + \int_t^T \frac{g_s^2\alpha_s}{2\sigma_s^2}\frac{1}{2}\boldsymbol{p}_s^T\boldsymbol{H}(\boldsymbol{x}_s, s)\boldsymbol{p}_s\, ds\,. \tag{93}$$

Inspecting the form of the S_1 equation, which reads

$$-\frac{\partial S_1}{\partial t} - \frac{g_t^2\alpha_t}{2\sigma_t^2}\left[\boldsymbol{G} - \boldsymbol{H}\nabla S_0\right]^T\nabla S_1 = -\left[\beta_t + \frac{g_t^2}{2\sigma_t^2}\left(\alpha_t W - 1 - \frac{\alpha_t}{2}\sum_{i,j}H_{ij}\partial_{ij}^2 S_0)\right)\right], \tag{94}$$

tells us that it can be solved using the *same* characteristic curves we used for S_0. (This is clear from the form of the left-hand side.) This implies

$$\frac{dS_1}{ds} = \beta_s + \frac{g_s^2}{2\sigma_s^2}\left(\alpha_s W - 1 - \frac{\alpha_s}{2}\sum_{i,j} H_{ij}\partial_{ij}^2 S_0)\right) \tag{95}$$

and hence that

$$S_1(\boldsymbol{x},t) = S_1(\boldsymbol{x}_T,T) - \int_t^T \left[\beta_t + \frac{g_t^2}{2\sigma_t^2}\left(\alpha_t W(\boldsymbol{x}_s,s) - 1 - \frac{\alpha_t}{2}\sum_{i,j} H_{ij}(\boldsymbol{x}_s,s)\partial_{ij}^2 S_0(\boldsymbol{x}_s,s)\right)\right] ds\,. \tag{96}$$

Putting our results for S_0 and S_1 together, we finally find that, assuming a *fixed initial condition* x_T at time T, the field at time t is

$$\phi(\boldsymbol{x},t|\boldsymbol{x}_T,T) = q(\boldsymbol{x}_T|\boldsymbol{x},t)e^{I_t(\boldsymbol{x};\boldsymbol{x}_T)}\phi_T(\boldsymbol{x}_T)\,. \tag{97}$$

This looks like our result for the gradient-dependent denoiser, but involves an extra part: the distribution $q(\boldsymbol{x}_T|\boldsymbol{x},t)$ which quantifies the probability of a given endpoint $\boldsymbol{x}_T$. We have

$$\begin{aligned} q(\boldsymbol{x}_T|\boldsymbol{x},t) &= \frac{\exp\left\{\int_t^T -\frac{1}{\xi}\frac{g_s^2\alpha_s}{2\sigma_s^2}\frac{1}{2}\boldsymbol{p}_s^T\boldsymbol{H}(\boldsymbol{x}_s,s)\boldsymbol{p}_s\,ds\right\}}{\int \exp\left\{\int_t^T -\frac{1}{\xi}\frac{g_s^2\alpha_s}{2\sigma_s^2}\frac{1}{2}\boldsymbol{p}_s^T\boldsymbol{H}(\boldsymbol{x}_s,s)\boldsymbol{p}_s\,ds\right\}d\boldsymbol{x}_T} \\ I_t(\boldsymbol{x};\boldsymbol{x}_T) &= \int_t^T \beta_t + \frac{g_s^2}{2\sigma_s^2}\left(\alpha_s W(\boldsymbol{x}_s,s) - 1 - \frac{\alpha_s}{2}\sum_{i,j} H_{ij}(\boldsymbol{x}_s,s)\partial_{ij}^2 S_0(\boldsymbol{x}_s,s)\right) ds\,. \end{aligned} \tag{98}$$

The final solution, which involves integrating over all possible $\boldsymbol{x}_T$, is

$$\phi_t(\boldsymbol{x}) = \int e^{I_t(\boldsymbol{x};\boldsymbol{x}_T)}\phi_T(\boldsymbol{x}_T)\; q(\boldsymbol{x}_T|\boldsymbol{x},t)d\boldsymbol{x}_T\,. \tag{99}$$

Although this solution is extremely complicated, and for the most part implicit, we can make the following observations. First, if $\boldsymbol{H} \to \boldsymbol{0}$ or $\xi \to 0$, the solution reduces to what we found in the previous subsection, since q approaches a delta function. In this limit, the more complex Hamiltonian dynamics (Eq. 91) reduce to what we found earlier (Eq. 81).

Second, for this type of denoiser, any pixel location in the initial noise seed can flow to *any other* pixel location in the final sample, although they are most likely to follow the characteristic curve defined by $\boldsymbol{G}$ (Eq. 81). The specific contribution of each possible flow to the final sample is determined by the distribution q.

E SAMPLING DYNAMICS OF TRANSLATION EQUIVARIANT LINEAR MODELS

In this appendix, we analytically solve the sampling dynamics of approximately translation equivariant models whose denoisers have the form

$$\boldsymbol{D}_t[\boldsymbol{\phi}](\boldsymbol{x}) = \boldsymbol{b}(\boldsymbol{x},t) + \int \left[\boldsymbol{W}(\boldsymbol{x}-\boldsymbol{y},t) + \sum_i \boldsymbol{G}_i(\boldsymbol{x}-\boldsymbol{y},t) y_i + \frac{\xi}{2} \sum_{i,j} \boldsymbol{H}_{ij}(\boldsymbol{x}-\boldsymbol{y},t) y_i y_j \right] \boldsymbol{\phi}(\boldsymbol{y})\, d\boldsymbol{y} \,. \tag{100}$$

It is useful to observe that the Fourier transform of this quantity looks like a *spatially local* denoiser:

$$\tilde{\boldsymbol{D}}_t[\boldsymbol{\phi}](\boldsymbol{k}) = \tilde{\boldsymbol{b}}(\boldsymbol{k},t) + \tilde{\boldsymbol{W}}(\boldsymbol{k},t)\tilde{\boldsymbol{\phi}}(\boldsymbol{k}) - i \sum_a \tilde{\boldsymbol{G}}_a(\boldsymbol{k},t) \frac{\partial \tilde{\boldsymbol{\phi}}(\boldsymbol{k})}{\partial k_a} - \frac{\xi}{2} \sum_{a,b} \tilde{\boldsymbol{H}}_{ab}(\boldsymbol{k},t) \frac{\partial^2 \tilde{\boldsymbol{\phi}}(\boldsymbol{k},t)}{\partial k_a \partial k_b} \,. \tag{101}$$

Because of this fact, in wavevector/Fourier space, the form of sampling dynamics is almost the same as what we solved in Appendix D:

$$\begin{aligned} \frac{\partial \tilde{\boldsymbol{\phi}}_t(\boldsymbol{k})}{\partial t} = & -\frac{g_t^2 \alpha_t}{2\sigma_t^2} \tilde{\boldsymbol{b}}(\boldsymbol{k},t) - \left[\beta_t + \frac{g_t^2}{2\sigma_t^2} \left(\alpha_t \tilde{\boldsymbol{W}}(\boldsymbol{k},t) - 1 \right) \right] \tilde{\boldsymbol{\phi}}_t(\boldsymbol{k}) \\ & + i \frac{g_t^2 \alpha_t}{2\sigma_t^2} \sum_a \tilde{\boldsymbol{G}}_a(\boldsymbol{k},t) \partial_a \tilde{\boldsymbol{\phi}}_t(\boldsymbol{k}) + \frac{g_t^2 \alpha_t}{2\sigma_t^2} \frac{\xi}{2} \sum_{a,b} \tilde{\boldsymbol{H}}_{ab}(\boldsymbol{k},t) \partial_{ab}^2 \tilde{\boldsymbol{\phi}}_t(\boldsymbol{k}) \,. \end{aligned} \tag{102}$$

In fact, up to the identifications $\boldsymbol{b} \to \tilde{\boldsymbol{b}}$, $\boldsymbol{W} \to \tilde{\boldsymbol{W}}$, $\boldsymbol{G} \to -i\tilde{\boldsymbol{G}}$, and $\boldsymbol{H} \to -\tilde{\boldsymbol{H}}$, they are *exactly* the same. We can hence use exactly the same strategy—an eigencolor assumption, plus the method of characteristics—to analytically solve sampling dynamics.

We will not reproduce the result from Appendix D, but merely note that its form is identical, and its interpretation is almost the same. Instead of *spatial* locations of the initial noise seed 'flowing' along characteristic curves to spatial locations of the final sample, something identical happens in Fourier/wavevector space: wavevector values of the initial noise seed flow along characteristic curves to wavevector values of the final sample.