Title: Renormalization Group Flows: from Optimal Transport to Diffusion Models

Speakers: Jordan Cotler

Collection/Series: Theory + AI Workshop: Theoretical Physics for AI

Date: April 10, 2025 - 1:30 PM

URL: https://pirsa.org/25040095

Abstract:

We show that Polchinski’s equation for exact renormalization group flow is equivalent to the optimal transport gradient flow of a field-theoretic relative entropy. This gives a surprising information-theoretic formulation of the exact renormalization group, expressed in the language of optimal transport. We will provide reviews of both the exact renormalization group, as well as the theory of optimal transportation. Our techniques generalize to other RG flow equations beyond Polchinski's. Moreover, we establish a connection between this more general class of RG flows and stochastic Langevin PDEs, enabling us to construct ML-based adaptive bridge samplers for lattice field theories. Finally, we will discuss forthcoming work on related methods to variationally approximate ground states of quantum field theories.

[Cotler, Rezchikov 2202.11737]

[Cotler, Rezchikov 2308.12355]

[Albergo, Cotler WIP]

Renormalization Group Flows: from Diffusion Models to Optimal Transport

JORDAN COTLER

HARVARD UNIVERSITY

Setting the scene

Me: Can you make an image of a scientist giving a talk about latent diffusion models and RG flow at Perimeter?

GPT-4o + DALL-E 3:

Setting the scene

Me: Can you make an image of a scientist giving a talk about latent diffusion models and RG flow at Perimeter?

GPT-4o + DALL-E 3:

How does this work?

How can we leverage it for physics?

Three Parts

Part I

Latent Diffusion Models

Part II

Exact RG as Stochastic Langevin Dynamics

Part III

Synthesis: Renormalizing Diffusion Models

Part I

Latent Diffusion Models

Basic problem

Want to sample $\text{image} \leftarrow p(\text{image} \mid \text{text})$

How do we learn $p(\text{image} \mid \text{text})$ and then efficiently sample from it?

Idea: Use training set $\{(\text{image}_i, \text{text}_i)\}_{i=1}^N$ and optimize p_θ over θ so that it approximates p

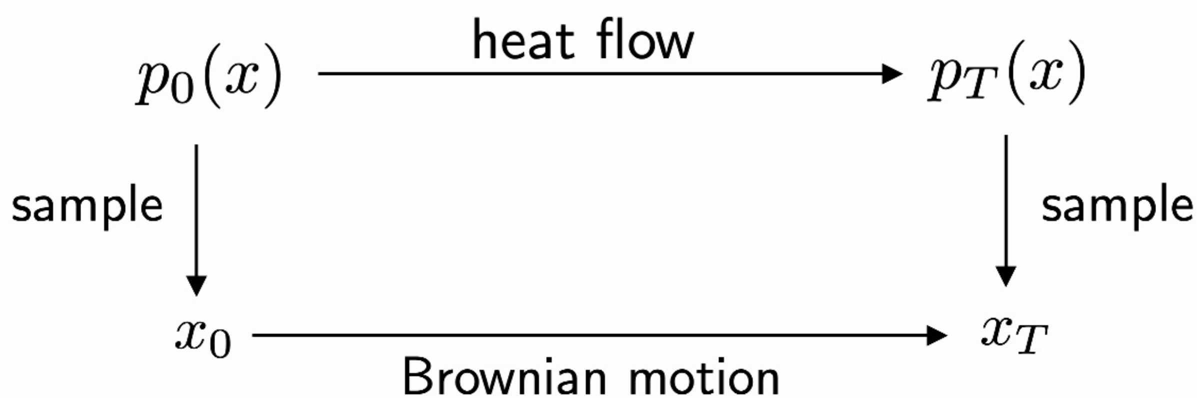
Want to parameterize p_θ so that it is efficient to sample from it

First tool: Fokker-Planck versus Langevin

Start with heat equation: $\frac{\partial}{\partial t} p_t(x) = \Delta p_t(x)$ **PDE**

Related to Brownian motion: $dx_t = \sqrt{2}\, dW_t$ **Stochastic ODE**

Question: Given some $p_0(x)$, how do we sample from $p_T(x)$?

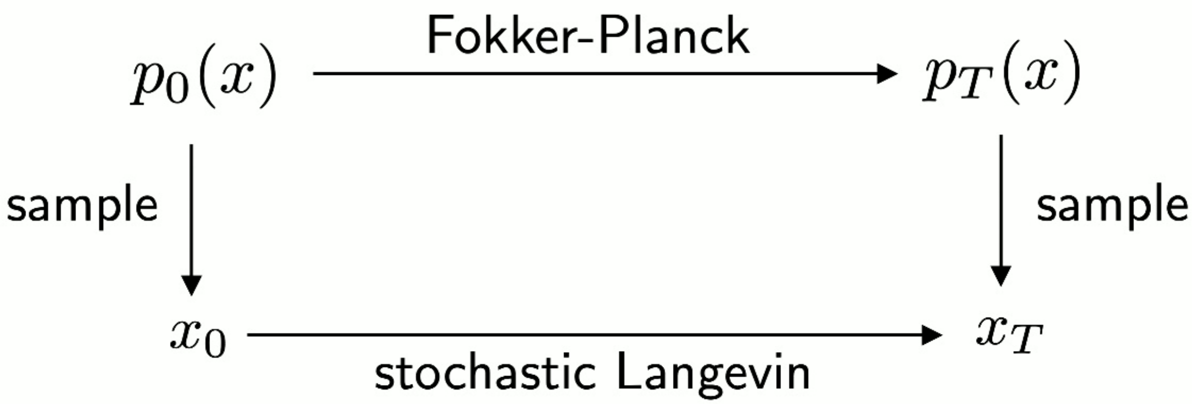


First tool: Fokker-Planck versus Langevin

Fokker-Planck: $\frac{\partial}{\partial t} p_t(x) = \partial_i(\partial^i V(x)\, p_t(x)) + \partial_i \partial^i p_t(x)$ **PDE**

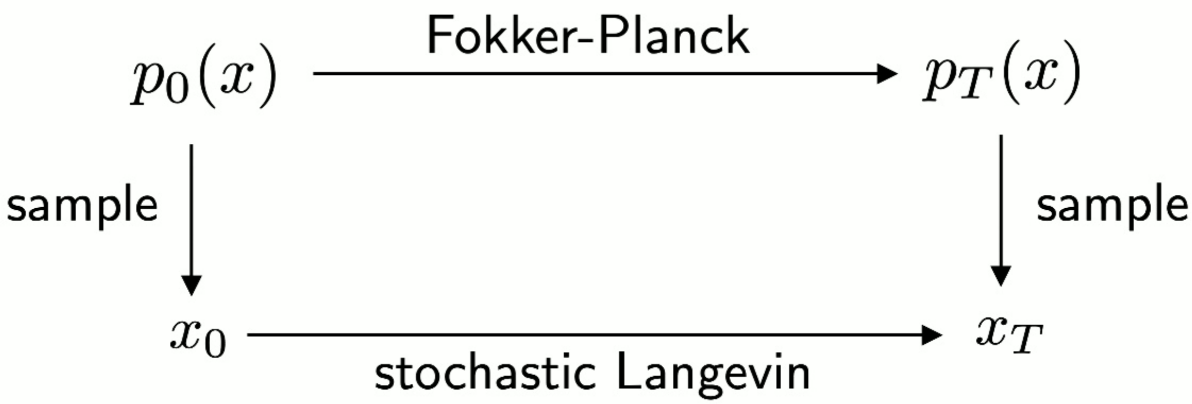
Stochastic Langevin: $dx_t = -\partial_i V(x_t)\, dt + \sqrt{2}\, dW_t$ **Stochastic ODE**

Question: Given some $p_0(x)$, how do we sample from $p_T(x)$?



First tool: Fokker-Planck versus Langevin

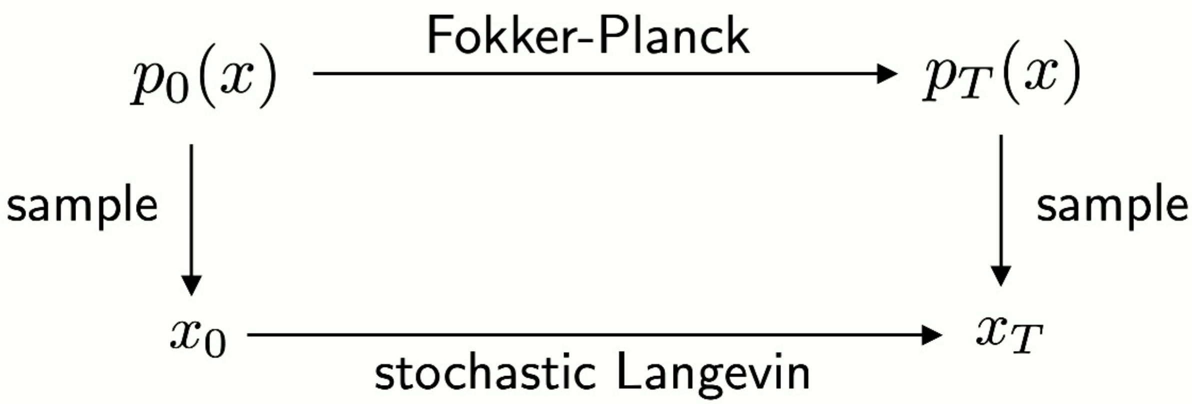
$$\frac{\partial}{\partial t} p_t(x) = \partial_i(\partial^i V(x)\, p_t(x)) + \partial_i \partial^i p_t(x)$$



First tool: Fokker-Planck versus Langevin

$$0 = \frac{\partial}{\partial t} p_t(x) = \partial_i (\partial^i V(x)\, p_t(x)) + \partial_i \partial^i p_t(x)$$

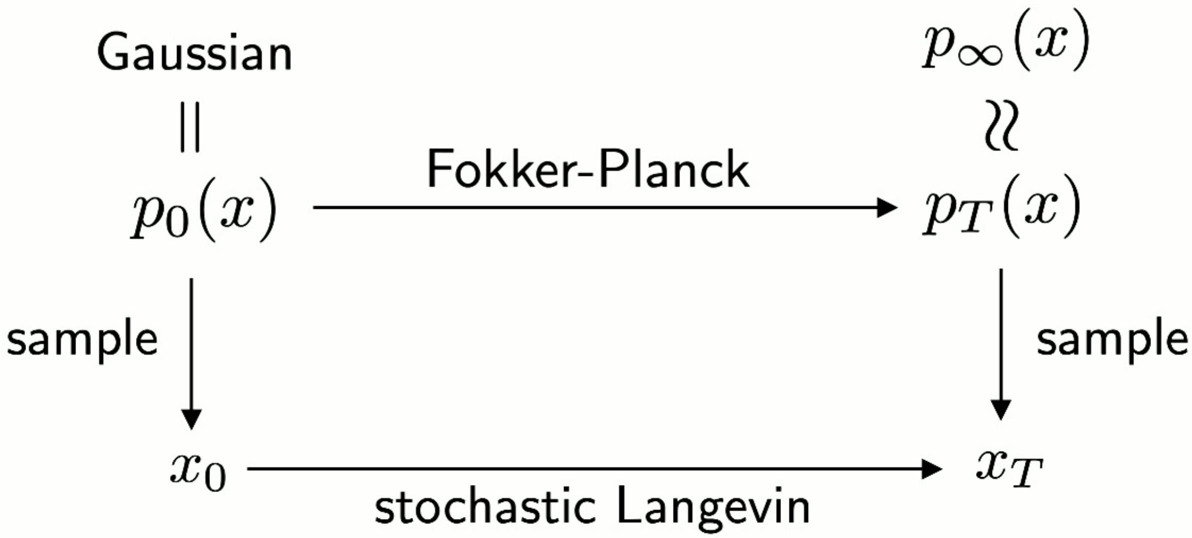
$$\Rightarrow \quad p_\infty(x) = \frac{1}{Z}\, e^{-V(x)}$$



First tool: Fokker-Planck versus Langevin

$$0 = \frac{\partial}{\partial t} p_t(x) = \partial_i(\partial^i V(x)\, p_t(x)) + \partial_i \partial^i p_t(x)$$

$$\Rightarrow \quad p_\infty(x) = \frac{1}{Z}\, e^{-V(x)}$$



Second tool: score-based sampling

Suppose we want to sample from $p_\theta(x) = \frac{1}{Z_\theta} e^{-V_\theta(x)}$

Define the *score* by $s_\theta(x) := \nabla \log p_\theta(x) = -\nabla V_\theta(x)$ } independent of Z_θ

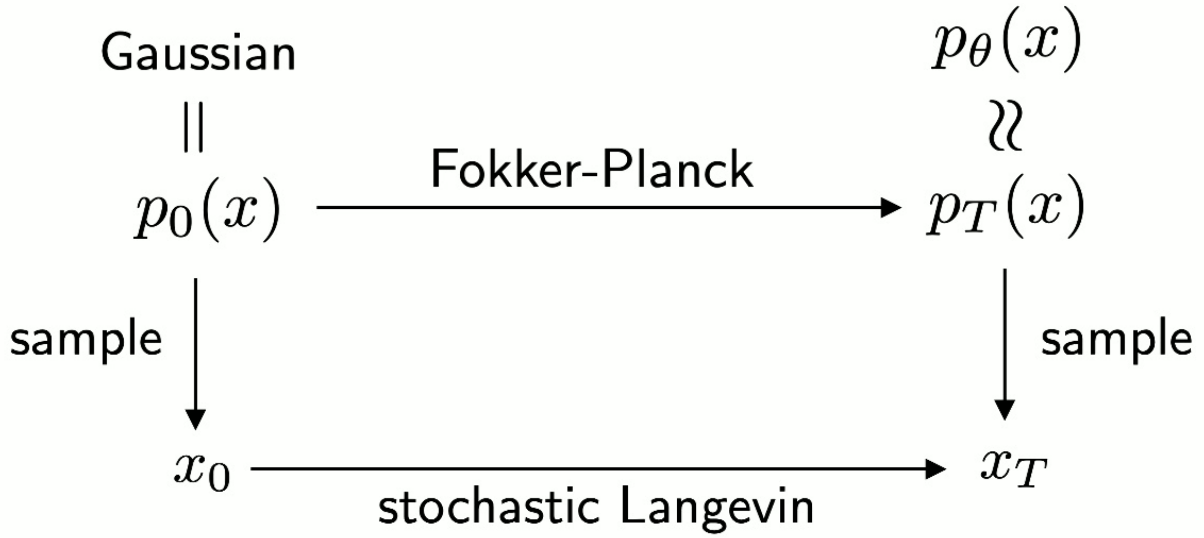
Fokker-Planck: $\frac{\partial}{\partial t} p_t(x) = -\nabla \cdot (s_\theta(x)\, p_t(x)) + \Delta p_t(x)$ **PDE**

Stochastic Langevin: $dx_t = s_\theta(x_t)\, dt + \sqrt{2}\, dW_t$ **Stochastic ODE**

Second tool: score-based sampling

Suppose we want to sample from $p_\theta(x) = \frac{1}{Z_\theta} e^{-V_\theta(x)}$

Define the *score* by $s_\theta(x) := \nabla \log p_\theta(x) = -\nabla V_\theta(x)$ } independent of Z_θ

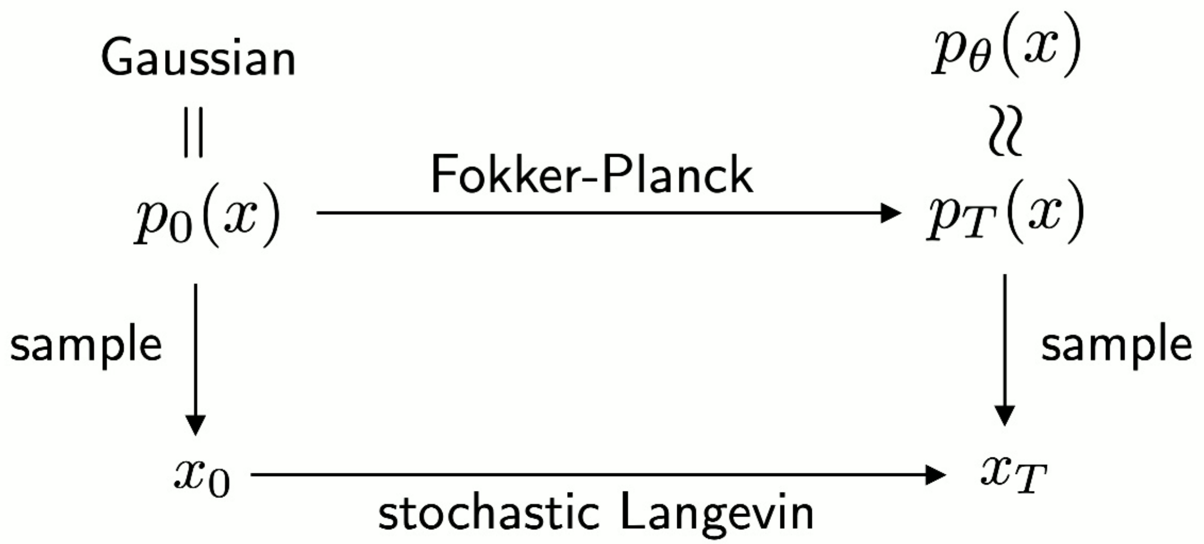


Only need the score!

Second tool: score-based sampling

Suppose we want to sample from $p_\theta(x) = \frac{1}{Z_\theta} e^{-V_\theta(x)}$

Define the *score* by $s_\theta(x) := \nabla \log p_\theta(x) = -\nabla V_\theta(x)$ } independent of Z_θ



Only need the score!

Score-based diffusion models

[Song et al. '19]
[Ho, Jain, Abbeel '20]
[Song et al. '21]

Idea: Consider flows

$$\frac{\partial}{\partial t} p_t(x) = \nabla \cdot (x\, p_t(x)) + \Delta p_t(x)$$

$$dx_t = -x_t\, dt + \sqrt{2}\, dW_t$$

converges to

$$p_\infty(x) = \frac{1}{(2\pi)^{d/2}} e^{-\frac{1}{2}|x|^2}$$

Let $p_0(x) = p^*(x)$

Flow $p_t(x)$ and learn score $s_t^*(x) = \nabla \log p_t^*(x)$ by optimizing a parameterized score $s_t^\theta(x)$ over θ

Score-based diffusion models

[Song et al. '19]
[Ho, Jain, Abbeel '20]
[Song et al. '21]

Idea: Consider flows

$$\frac{\partial}{\partial t} p_t(x) = \nabla \cdot (x\, p_t(x)) + \Delta p_t(x)$$
$$dx_t = -x_t\, dt + \sqrt{2}\, dW_t$$

converges to

$$p_\infty(x) = \frac{1}{(2\pi)^{d/2}} e^{-\frac{1}{2}|x|^2}$$

Let $p_0(x) = p^*(x)$

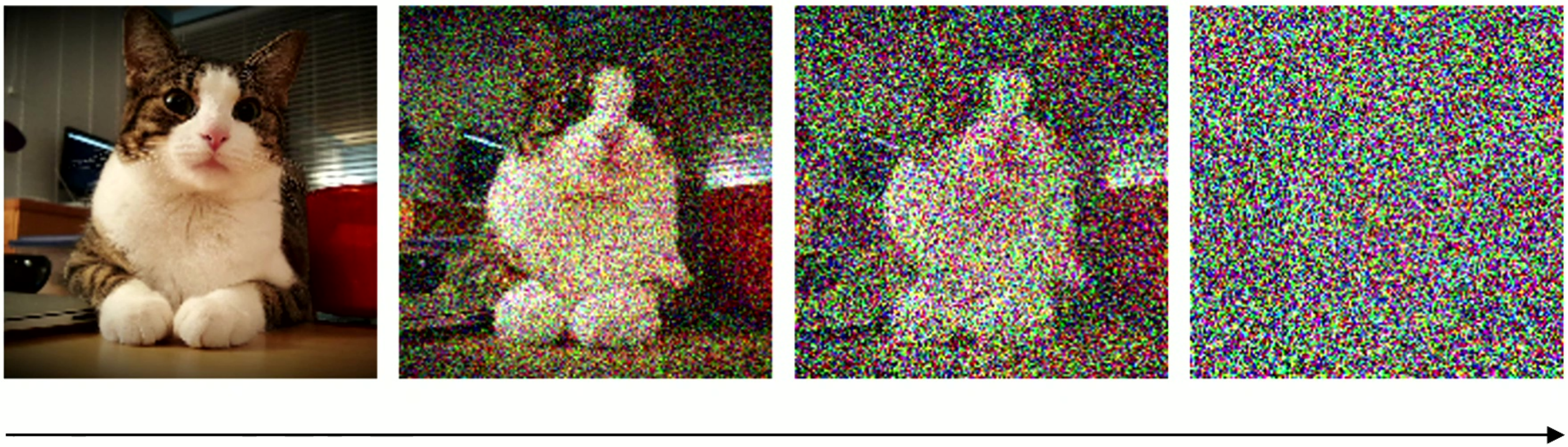
Flow $p_t(x)$ and learn score $s_t^*(x) = \nabla \log p_t^*(x)$ by optimizing a parameterized score $s_t^\theta(x)$ over θ

Construct flows

$$\frac{\partial}{\partial t} p_t(x) = -\nabla \cdot \left([x + 2\, s_{T-t}^\theta(x)]\, p_t(x)\right) + \Delta p_t(x)$$
$$dx_t = -[x_t + 2\, s_{T-t}^\theta(x_t)]\, dt + \sqrt{2}\, dW_t$$

Score-based diffusion models

[Song et al. '19]
[Ho, Jain, Abbeel '20]
[Song et al. '21]

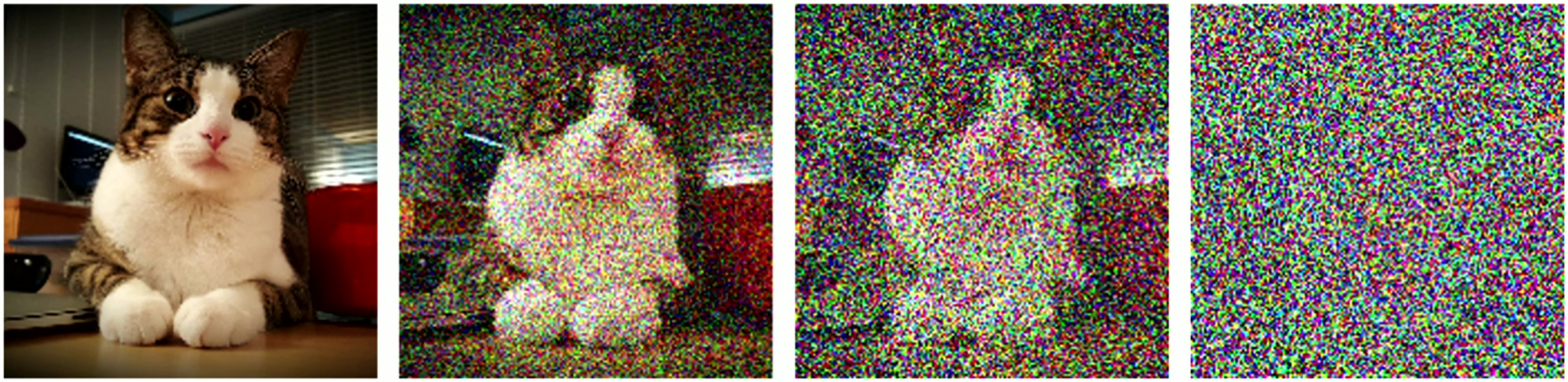


flow of true score $s_t^*(x)$

Source: NVIDIA

Score-based diffusion models

[Song et al. '19]
[Ho, Jain, Abbeel '20]
[Song et al. '21]



flow of true score $s_t^*(x)$

learn $s_t^\theta(x) \approx s_t^*(x)$

Source: NVIDIA

Score-based diffusion models

[Song et al. ‘19]
[Ho, Jain, Abbeel ‘20]
[Song et al. ‘21]

reverse flow of learned score $s_t^\theta(x)$

Source: NVIDIA

[Hyvärinen '05]

How to learn a score

Minimize the Fisher divergence

$$F[s_t^\theta(x)] = \int d^d x\, p_t^*(x) \left| s_t^\theta(x) - s_t^*(x) \right|^2$$

$$= \int d^d x\, p_t^*(x) \left[|s_t^\theta(x)|^2 - 2 s_t^\theta(x) \cdot s_t^*(x) \right] + C_t$$

$$= \int d^d x\, p_t^*(x) \left[|s_t^\theta(x)|^2 + 2\, \nabla \cdot s_t^\theta(x) \right] + C_t$$

$$\Longrightarrow \quad \mathbb{E}_{x \sim p_t^*(x)} \Big[|s_t^\theta(x)|^2 + 2\, \nabla \cdot s_t^\theta(x) \Big]$$

minimize over θ

Score-based diffusion models: putting it together

1. Sample N times from $p^*(x)$: $x^{*,1}, x^{*,2}, ..., x^{*,N}$

2. Sample N times from $[0, T]$: $t_1, t_2, ..., t_N$

3. Flow samples using Ornstein-Uhlenbeck as: $x^{*,1}_{t_1}, x^{*,2}_{t_2}, ..., x^{*,N}_{t_N}$

4. Minimize $\mathsf{L}(\theta) = \frac{1}{N}\sum_{i=1}^{N}\left[|s^{\theta}_{t_i}(x^{*,i}_{t_i})|^2 + 2\nabla \cdot s^{\theta}_{t_i}(x^{*,i}_{t_i})\right]$

➡ Use learned score to approximately sample from $p^*(x)$

State of the art

Me: Can you make an image of a scientist giving a talk about latent diffusion models and RG flow at Perimeter?

GPT-4o + DALL-E 3:

How does this work? ✓

How can we leverage it for physics?

Basic setup

[Wilson, Kogut '74] [Polchinski '84]

Euclidean scalar field theory on $\mathbb{R}^d$, $\phi : \mathbb{R}^d \to \mathbb{R}$

Probability functional $P_\Lambda[\phi(x)] \propto e^{-S_\Lambda[\phi]}$

RG scale $\Lambda \sim 1/\ell$

Exact renormalization group (ERG) flow equation:

$$-\Lambda \frac{d}{d\Lambda} P_\Lambda[\phi] = \mathcal{F}\left[P_\Lambda[\phi], \frac{\delta P_\Lambda[\phi]}{\delta \phi}, \frac{\delta^2 P_\Lambda[\phi]}{\delta \phi \delta \phi}, \dots \right]$$

[Polchinski ‘84]

Exact Renormalization Group

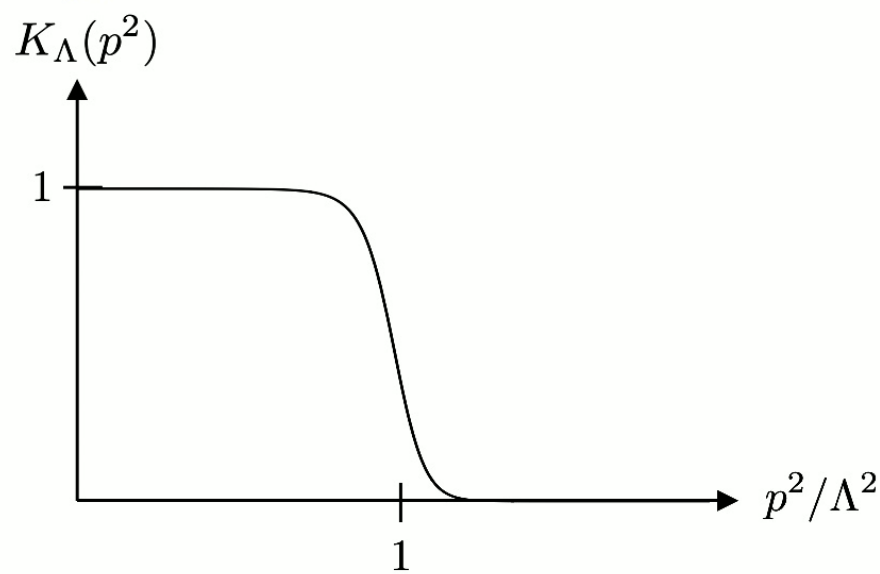
Assume the source vanishes above the cutoff scale

Euclidean scalar field with a source:

$$Z_\Lambda[J] := \int [d\phi]\, e^{-\frac{1}{2}\int \frac{d^d p}{(2\pi)^d}\left(\phi(p)\phi(-p)(p^2+m^2)K_\Lambda^{-1}(p^2)+J(p)\phi(-p)\right)-S_{\mathrm{int},\Lambda}[\phi]}$$

Physics below the cutoff scale is preserved under RG flow:

$$-\Lambda \frac{d}{d\Lambda} Z_\Lambda[J] = C_\Lambda\, Z_\Lambda[J]$$



[Polchinski '84]

Exact Renormalization Group

$$-\Lambda \frac{d}{d\Lambda} Z_\Lambda[J] = \int [d\phi] \left(\frac{1}{2} \int \frac{d^d p}{(2\pi)^d} \, \phi(p)\phi(-p)(p^2+m^2) \, \Lambda \frac{\partial K_\Lambda^{-1}(p^2)}{\partial \Lambda} + \Lambda \frac{\partial S_{\text{int},\Lambda}[\phi]}{\partial \Lambda} \right) e^{-S_\Lambda[\phi,J]}$$

★ $$-\Lambda \frac{d}{d\Lambda} Z_\Lambda[J] = C_\Lambda \, Z_\Lambda[J]$$

Jeopardy question: Is there a functional differential equation for this quantity such that ★ is satisfied?

Polchinski equation as a Fokker-Planck equation

Polchinski equation

$$-\Lambda \frac{d}{d\Lambda} P_\Lambda[\phi] = \int d^d p \, \Lambda \frac{\partial C_\Lambda(p^2)}{\partial \Lambda} \frac{\delta^2}{\delta\phi(p)\delta\phi(-p)} P_\Lambda[\phi] + \int d^d p \, \Lambda \frac{\partial C_\Lambda(p^2)}{\partial \Lambda} \frac{\delta}{\delta\phi(p)} \left(C_\Lambda^{-1}(p^2)\,\phi(p)\,P_\Lambda[\phi]\right)$$

Fokker-Planck equation

$$\frac{\partial}{\partial t} p_t(x) = \partial_i \partial^i p_t(x) + \partial_i(\partial^i V(x)\, p_t(x))$$

Wegner-Morris equation as a Fokker-Planck equation

Wegner-Morris equation [Wegner '74] [Morris '95]

$$-\Lambda \frac{d}{d\Lambda} P_\Lambda[\phi] = \frac{1}{2} \int d^d p \, B_\Lambda(|p|) \left(\frac{\delta^2 P_\Lambda[\phi]}{\delta\phi(p)\delta\phi(-p)} + 2 \frac{\delta}{\delta\phi(p)} \left(\frac{\delta \widehat{S}_\Lambda[\phi]}{\delta\phi(-p)} P_\Lambda[\phi] \right) \right)$$

Fokker-Planck equation

$$\frac{\partial}{\partial t} p_t(x) = \partial_i \partial^i p_t(x) + \partial_i (\partial^i V(x) \, p_t(x))$$

$\Rightarrow$ $t = -\log(\Lambda/\Lambda_0)$

[Cotler, Rezchikov '22]

Functional generalization of optimal transport

Initial and final distributions: $P_1[\phi], P_2[\phi]$

Transport kernel: $\Pi[\phi_1, \phi_2]$

$$\int [d\phi_2]\, \Pi[\phi_1, \phi_2] = P_1[\phi_1]\,, \quad \int [d\phi_1]\, \Pi[\phi_1, \phi_2] = P_2[\phi_2]$$

Cost: $\mathcal{C} : \mathscr{F} \times \mathscr{F} \to \mathbb{R}$

Minimize

$$\mathrm{K}[\Pi] = \int [d\phi_1]\, [d\phi_2]\, \Pi[\phi_1, \phi_2]\, \mathcal{C}[\phi_1, \phi_2]$$

Wasserstein-2 distance

$$\mathcal{W}_2(P_1, P_2) :=$$
$$\left(\inf_{\Pi \in \Gamma(P_1, P_2)} \int [d\phi_1]\, [d\phi_2]\, \Pi[\phi_1, \phi_2] \int d^d x\, d^d y\, B_\Lambda^{-1}(x - y)\, (\phi_1(x) - \phi_2(x))\, (\phi_1(y) - \phi_2(y)) \right)^{1/2}$$

[Cotler, Rezchikov '22]

RG flow equation

Distribution of interest: $P_\Lambda[\phi] = \frac{1}{Z_{P,\Lambda}} e^{-S_\Lambda[\phi]}$

"Background" distribution: $Q_\Lambda[\phi] = \frac{1}{Z_{Q,\Lambda}} e^{-2\hat{S}_\Lambda[\phi]}$

Relative entropy: $S(P\|Q) = \int [d\phi]\, P[\phi] \log(P[\phi]/Q[\phi])$

$$-\Lambda \frac{d}{d\Lambda} P_\Lambda[\phi] = -\nabla_{\mathcal{W}_2} S(P_\Lambda[\phi] \,\|\, Q_\Lambda[\phi])$$

[Cotler, Rezchikov '22]

Wegner-Morris equation as a Fokker-Planck equation

RG flow equation

Distribution of interest: $P_\Lambda[\phi] = \frac{1}{Z_{P,\Lambda}} e^{-S_\Lambda[\phi]}$

"Background" distribution: $Q_\Lambda[\phi] = \frac{1}{Z_{Q,\Lambda}} e^{-\hat{S}_\Lambda[\phi]}$

Relative entropy: $S(P\|Q) = \int [d\phi]\, P[\phi] \log(P[\phi]/Q[\phi])$

$$-\Lambda \frac{d}{d\Lambda} P_\Lambda[\phi] = -\nabla_{\mathcal{W}_2} S(P_\Lambda[\phi] \,\|\, Q_\Lambda[\phi])$$

Wegner-Morris equation

RG flow is a gradient flow with respect to the relative entropy!

Wegner-Morris equation [Wegner '74] [Morris '95]

$$-\Lambda \frac{d}{d\Lambda} P_\Lambda[\phi] = \frac{1}{2} \int d^d p\, B_\Lambda(|p|) \left(\frac{\delta^2 P_\Lambda[\phi]}{\delta\phi(p)\delta\phi(-p)} + 2\frac{\delta}{\delta\phi(p)} \left(\frac{\delta \widehat{S}_\Lambda[\phi]}{\delta\phi(-p)} P_\Lambda[\phi] \right) \right)$$

Fokker-Planck equation

$$\frac{\partial}{\partial t} p_t(x) = \partial_i \partial^i p_t(x) + \partial_i(\partial^i V(x)\, p_t(x))$$

→ $t = -\log(\Lambda/\Lambda_0)$

Wegner-Morris equation as a Fokker-Planck equation

Wegner-Morris equation [Wegner '74] [Morris '95]

$$-\Lambda \frac{d}{d\Lambda} P_\Lambda[\phi] = \frac{1}{2} \int d^d p \, B_\Lambda(|p|) \left(\frac{\delta^2 P_\Lambda[\phi]}{\delta\phi(p)\delta\phi(-p)} + 2 \frac{\delta}{\delta\phi(p)} \left(\frac{\delta \widehat{S}_\Lambda[\phi]}{\delta\phi(-p)} P_\Lambda[\phi] \right) \right)$$

Fokker-Planck equation

$$\frac{\partial}{\partial t} p_t(x) = \partial_i \partial^i p_t(x) + \partial_i (\partial^i V(x) \, p_t(x))$$

$$\Rightarrow \quad t = -\log(\Lambda/\Lambda_0)$$

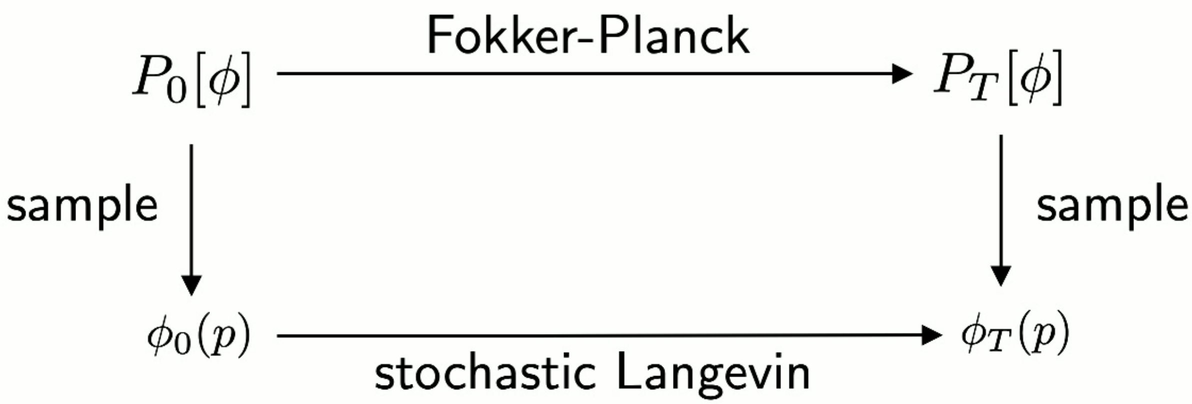
[Carosso '19]
[Cotler, Rezchikov '22]
[Cotler, Rezchikov '23]

Reprise: Fokker-Planck versus Langevin

Fokker-Planck: $\frac{\partial}{\partial t} P_t[\phi] = \frac{1}{2} \int d^d p \, B_t(|p|) \left(\frac{\delta^2 P_t[\phi]}{\delta\phi(p)\delta\phi(-p)} + 2 \frac{\delta}{\delta\phi(p)} \left(\frac{\delta \widehat{S}_t[\phi]}{\delta\phi(-p)} P_t[\phi] \right) \right)$ **FPDE**

Stochastic Langevin: $d\phi_t(p) = -B_t(|p|) \frac{\delta \widehat{S}_t[\phi]}{\delta\phi(-p)} \, dt + \sqrt{B_t(|p|)} \, d\eta_t(p)$ **Stochastic PDE**

Question: Given some $P_0[\phi]$, how do we sample from $P_T[\phi]$?



Towards Renormalizing Diffusion Models

[Carosso '19]
[Cotler, Rezchikov '22]
[Cotler, Rezchikov '23]

Sampling from field theories is an essential operation in statistical and quantum field theory

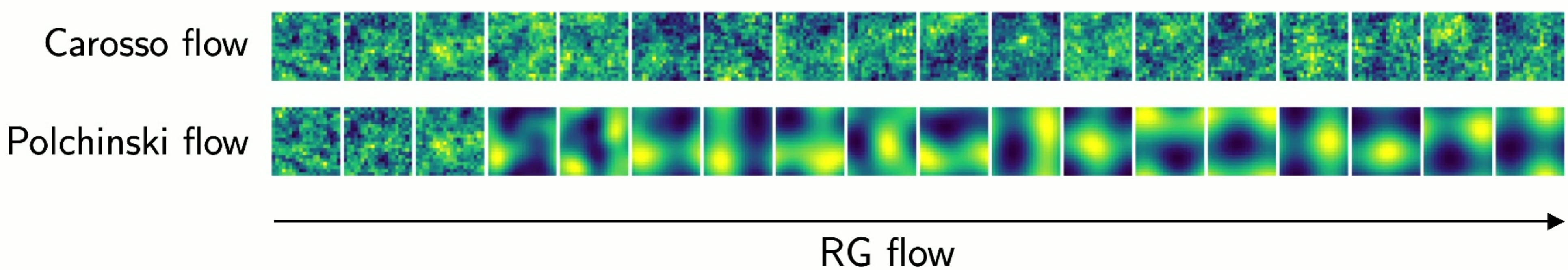
Field theories come equipped with canonical Fokker-Planck / Langevin flows, namely their RG flows

We can leverage score-based diffusion models (and similar methods) to build novel algorithms for sampling from field theories, exploiting their RG flow structure

Overview

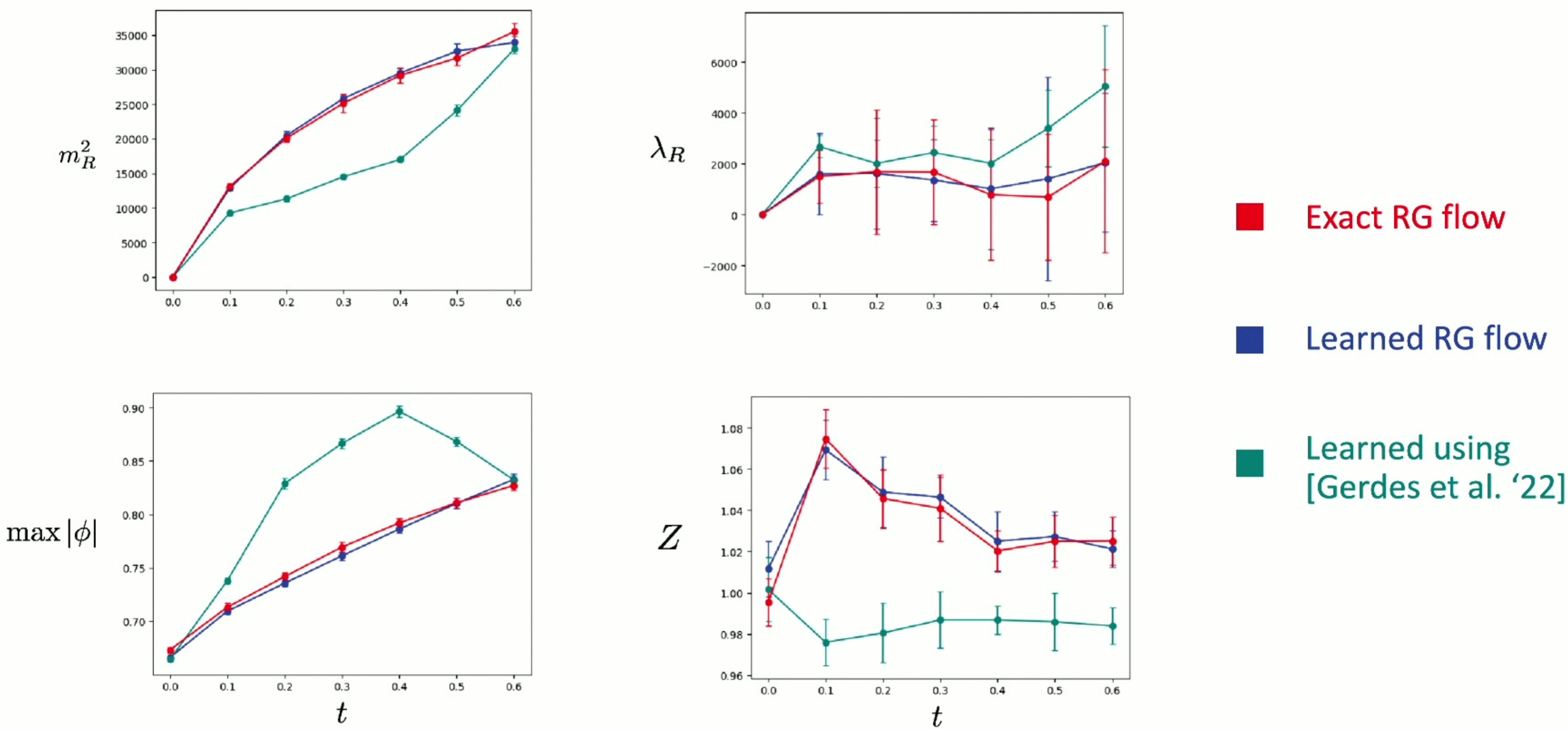
In [Cotler, Rezchikov '23] we developed detailed algorithms leveraging insights from RG to build tailored flow-based and diffusion model algorithms for field theories

Here let us show some results from our learned models in the context of Euclidean scalar ϕ^4 theory in 2 dimensions (20 x 20 grid)



[Cotler, Rezchikov ‘23]

Learning RG flows



[Albergo, Cotler WIP]

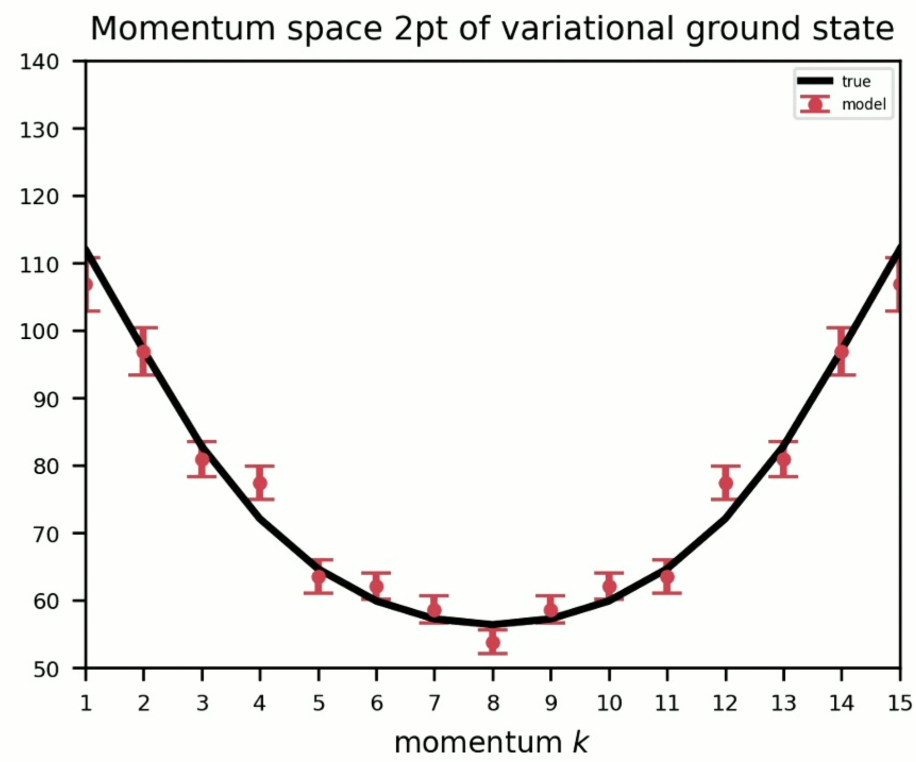
Learning quantum ground states

Variationally learn ground state wavefunctionals of e.g. 2+1 scalar φ^4

Use a MERA-inspired ansatz

Currently working with 128 bosonic lattice sites (!)

Should scale to at least 1024 bosonic lattice sites



Connections to other stories

Mathematical physics (**[Bauerschmidt-Bodineau], [Bauerschmidt '23]**)

Mixing of Markov chains in high-temperature ϕ^4, sine-Gordon

Optimal transport (**above, also [Cotler, Rezchikov '22]**)

Bakry-Emery, log-Sobolev inequalities, connection to correlation decay

Bayesian inference **[Berman et al.]**

Stochastic localization (**[Montanari '23], [Eldan-Koehler-Zeitouni '21], many others**)

Many powerful results about spin glasses, convex geometry, mixing of Markov chains

Wilson/gradient flow (**[Lüscher], [Carosso], many others**)

Wilson flow is an interesting smoothing process (Yang-Mills gradient flow)

Motivation for work on normalizing flows

Discussion

Many rich connections between field theory and latent diffusion models

New algorithms for efficient sampling of field theories, sampling along RG flows, and learning ground states of quantum field theories

Insights from RG flows may also help to improve latent diffusion models for image generation

Much more to understand and explore

[Carosso '19]
[Cotler, Rezchikov '22]
[Cotler, Rezchikov '23]

Reprise: Fokker-Planck versus Langevin

Fokker-Planck: $$\frac{\partial}{\partial t} P_t[\phi] = \frac{1}{2} \int d^d p \, B_t(|p|) \left(\frac{\delta^2 P_t[\phi]}{\delta\phi(p)\delta\phi(-p)} + 2 \frac{\delta}{\delta\phi(p)} \left(\frac{\delta \widehat{S}_t[\phi]}{\delta\phi(-p)} P_t[\phi] \right) \right)$$ **FPDE**

Stochastic Langevin: $$d\phi_t(p) = -B_t(|p|) \frac{\delta \widehat{S}_t[\phi]}{\delta\phi(-p)} \, dt + \sqrt{B_t(|p|)} \, d\eta_t(p)$$ **Stochastic PDE**

Question: Given some $P_0[\phi]$, how do we sample from $P_T[\phi]$?