A first view at Renormalized Quantum information Geometry

Olivier Croissant

July 29, 2024

Contents

Abstract

This paper explores several key topics in theoretical physics and information geometry, emphasizing their interconnections and applications. We begin with an investigation into the application of the Exact Renormalization Group (ERG) within scalar field theory, presenting it as an optimal transport gradient flow. The derivation of optimal transport equations for renormalization flows is revisited, highlighting the geometric and probabilistic aspects.

The discussion then extends to the interpretation of path integrals as Bregman divergences, leveraging the Pythagorean relationship for Bregman divergences to develop a new perturbative approach. This method decomposes the path integral into geometrically consistent steps, potentially offering a more effective means of handling divergences than traditional Feynman diagram techniques.

Next, we examine the triangle relationship for Bregman divergences using dual connections in information geometry. This relationship provides a deeper understanding of the structure and properties of Bregman divergences, illustrating their consistency with the underlying geometric framework. The connection between the dynamics of the renormalized action in quantum field theory and the derivative with respect to the Wasserstein distance in probabilistic divergences is also explored, underscoring the unifying themes across different domains.

Finally, we delve into the challenges of one-loop renormalization in quantum gravity, outlining the techniques used to handle divergences such as dimensional regularization and the running of

coupling constants. This is followed by a discussion on the non-renormalizability of quantum gravity at the two-loop level, where the need for an infinite number of counterterms signifies the limitations of perturbative approaches and suggests the necessity for alternative theories or effective field theory approaches.

By bridging concepts from quantum field theory, quantum gravity, and information geometry, this work offers new perspectives and potential avenues for further research in both theoretical physics and applied mathematics.

1 Introduction

This document provides a detailed comparison and discussion of the Exact Renormalization Group (ERG), the Polchinski equation, and the Wetterich equation. It includes mathematical formulations, key features, and their interconnections. Then We establish the connection with the optimal transport through a divergence.The minimization of the action should works while the regularisation may project the divergence to infinity.

2 Exact Renormalization Group (ERG)

The ERG framework describes the flow of the effective action as a function of a scale parameter. It provides a non-perturbative approach to understanding how a physical system's description changes as high-energy degrees of freedom are integrated out.

2.1 Polchinski Equation

The Polchinski equation for the Wilsonian effective action $S_\Lambda[\phi]$ is given by:

$$\Lambda \frac{\partial S_\Lambda[\phi]}{\partial \Lambda} = \frac{1}{2} \int d^d x\, d^d y \, \frac{\delta S_\Lambda[\phi]}{\delta \phi(x)} \, P_\Lambda(x-y) \, \frac{\delta S_\Lambda[\phi]}{\delta \phi(y)} - \frac{1}{2} \int d^d x\, d^d y \, \frac{\delta^2 S_\Lambda[\phi]}{\delta \phi(x) \delta \phi(y)} P_\Lambda(x-y), \tag{1}$$

where $P_\Lambda(x-y)$ is a regulated propagator that depends on the cutoff scale Λ.

2.1.1 Key Features

- **Wilsonian Effective Action** $S_\Lambda[\phi]$: Describes the action with high-energy modes (momenta above Λ) integrated out.
- **Functional Derivatives**: Includes both the first and second functional derivatives, providing a detailed account of the field interactions.
- **Regulated Propagator**: $P_\Lambda(x-y)$ ensures that modes with momenta $p^2 > \Lambda^2$ are suppressed.

3 Wetterich Equation

The Wetterich equation describes the flow of the effective average action $\Gamma_k[\phi]$ as the momentum scale k changes. It is given by:

$$k \frac{\partial \Gamma_k[\phi]}{\partial k} = \frac{1}{2} \int \frac{d^d q}{(2\pi)^d} \frac{\partial R_k(q^2)}{\partial k} \left[\Gamma_k^{(2)}[\phi] + R_k(q^2) \right]^{-1}_{q,-q}, \tag{2}$$

where $R_k(q^2)$ is a regulator function.

3.0.1 Key Features

- **Effective Average Action** $\Gamma_k[\phi]$: Interpolates between the bare action at high scales and the full quantum effective action at low scales.
- **Second Functional Derivative** $\Gamma_k^{(2)}[\phi]$: Represents the inverse propagator.
- **Regulator Function** $R_k(q^2)$: Suppresses low-momentum modes, ensuring that only high-momentum modes are integrated out.

4 Comparison of Polchinski and Wetterich Equations

4.1 Structure

- **Wetterich Equation**: Focuses on the second functional derivative $\Gamma_k^{(2)}[\phi]$, simplifying the equation.
- **Polchinski Equation**: Includes both first and second functional derivatives, providing a more detailed account of the field interactions.

4.2 Flow and Scope

- **Wetterich Equation**: Provides a flow equation for the effective average action, smoothly interpolating between scales.
- **Polchinski Equation**: Gives a comprehensive flow equation for the Wilsonian effective action, capturing more details of the field interactions.

4.3 Regulation and Integration

- **Wetterich Equation**: Uses a regulator function $R_k(q^2)$ to smoothly suppress low-momentum modes.
- **Polchinski Equation**: Uses a regulated propagator $P_\Lambda(x-y)$ with a sharper cutoff.

4.4 Practical Use

- **Wetterich Equation**: Preferred for practical computations due to its simpler form.
- **Polchinski Equation**: Suitable for in-depth theoretical studies due to its detailed structure.

5 Regulator Function $R_k(q^2)$

The regulator function $R_k(q^2)$ is defined to control the inclusion of quantum fluctuations in the effective average action $\Gamma_k[\phi]$.

5.1 Definition and Properties

- **UV Behavior**: $R_k(p^2) \to 0$ for $p^2 \gg k^2$.
- **IR Behavior**: $R_k(p^2) \to \infty$ for $p^2 \ll k^2$.
- **Boundary Condition**: $R_{k=0}(p^2) = 0$.
- **Smoothness**: The regulator function should be smooth for practical computations.

5.2 Common Choices for Regulator Functions

- **Exponential Regulator**:
$$R_k(p^2) = p^2 \left(\frac{e^{p^2/k^2} - 1}{e^{p^2/k^2}} \right). \tag{3}$$
- **Sharp Cutoff** (illustrative, not typically used in functional RG):
$$R_k(p^2) = k^2 \theta(k^2 - p^2), \tag{4}$$
where θ is the Heaviside step function.
- **Litim Regulator**:
$$R_k(p^2) = (k^2 - p^2)\theta(k^2 - p^2). \tag{5}$$

6 Relationship Between the Polchinski Equation and the Probability Functional

The Polchinski equation is a functional differential equation that describes the flow of the Wilsonian effective action $S_\Lambda[\phi]$ as the cutoff scale Λ is varied. This effective action is related to the probability functional $P_\Lambda[\phi]$, which describes the probability distribution of the field configurations ϕ at the scale Λ. The probability functional is given by:

$$P_\Lambda[\phi] = \frac{e^{-S_\Lambda[\phi]}}{Z_\Lambda}, \tag{6}$$

where Z_Λ is the partition function ensuring normalization:

$$Z_\Lambda = \int [d\phi]\, e^{-S_\Lambda[\phi]}. \tag{7}$$

The Polchinski equation is expressed as:

$$\Lambda \frac{\partial S_\Lambda[\phi]}{\partial \Lambda} = \frac{1}{2} \int d^d x\, d^d y\, \frac{\delta S_\Lambda[\phi]}{\delta \phi(x)}\, P_\Lambda(x-y)\, \frac{\delta S_\Lambda[\phi]}{\delta \phi(y)} - \frac{1}{2} \int d^d x\, d^d y\, \frac{\delta^2 S_\Lambda[\phi]}{\delta \phi(x) \delta \phi(y)} P_\Lambda(x-y), \tag{8}$$

where $P_\Lambda(x-y)$ is a regulated propagator.

To understand the relationship between the Polchinski equation and the probability functional, consider the derivative of $P_\Lambda[\phi]$ with respect to Λ:

$$\Lambda \frac{\partial P_\Lambda[\phi]}{\partial \Lambda} = \Lambda \frac{\partial}{\partial \Lambda} \left(\frac{e^{-S_\Lambda[\phi]}}{Z_\Lambda} \right). \tag{9}$$

Applying the product rule, we get:

$$\Lambda \frac{\partial P_\Lambda[\phi]}{\partial \Lambda} = P_\Lambda[\phi] \left(-\Lambda \frac{\partial S_\Lambda[\phi]}{\partial \Lambda} - \left\langle -\Lambda \frac{\partial S_\Lambda[\phi]}{\partial \Lambda} \right\rangle_{P_\Lambda} \right), \tag{10}$$

where the expectation value is defined as:

$$\left\langle -\Lambda \frac{\partial S_\Lambda[\phi]}{\partial \Lambda} \right\rangle_{P_\Lambda} = \int [d\phi]\, P_\Lambda[\phi] \left(-\Lambda \frac{\partial S_\Lambda[\phi]}{\partial \Lambda} \right). \tag{11}$$

Substituting the Polchinski equation into this expression yields:

$$\begin{aligned} -\Lambda \frac{\partial P_\Lambda[\phi]}{\partial \Lambda} = P_\Lambda[\phi] \Bigg(& \frac{1}{2} \int d^d x\, d^d y\, \frac{\delta S_\Lambda[\phi]}{\delta \phi(x)}\, P_\Lambda(x-y)\, \frac{\delta S_\Lambda[\phi]}{\delta \phi(y)} \\ & - \frac{1}{2} \int d^d x\, d^d y\, \frac{\delta^2 S_\Lambda[\phi]}{\delta \phi(x) \delta \phi(y)} P_\Lambda(x-y) \Bigg) \\ & + P_\Lambda[\phi] \left\langle \Lambda \frac{\partial S_\Lambda[\phi]}{\partial \Lambda} \right\rangle_{P_\Lambda}. \end{aligned} \tag{12}$$

This demonstrates the connection between the evolution of the effective action $S_\Lambda[\phi]$ described by the Polchinski equation and the corresponding evolution of the probability functional $P_\Lambda[\phi]$. The Polchinski equation thus provides a detailed description of how quantum fluctuations and interactions are incorporated into the probability distribution of the field configurations as the scale Λ is varied.

7 Reformulation of Polchinski's Equation in Terms of Optimal Transport

One of our main results is that Polchinski's equation can be written as:

$$-\Lambda \frac{d}{d\Lambda} P_\Lambda[\phi] = -\nabla_{W_2} S(P_\Lambda[\phi] \| Q_\Lambda[\phi]), \tag{13}$$

where ∇_{W_2} is a gradient with respect to a functional generalization of the Wasserstein-2 metric, and $S(P\|Q)$ is a functional version of the relative entropy:

$$S(P\|Q) = \int [d\phi]\, P[\phi] \log\left(\frac{P[\phi]}{Q[\phi]}\right). \tag{14}$$

Here, $Q_\Lambda[\phi]$ is a background probability functional that essentially defines our RG scheme.

We emphasize that our formula has the flexibility to capture an enormous class of RG schemes. The ingredients of our formula require further explanation, which we will provide in detail later. Intuitively, Equation (1) tells us that the coarse-graining of our theory is generated by a decrease in relative entropy. We will later see that the relative entropy is in fact an RG monotone; although this may seem clear from the form of Equation (1), a more detailed analysis is required which involves unpacking the definition of the gradient.

The remainder of the paper is organized as follows. In Section 2, we review the Exact Renormalization Group (ERG) with an emphasis on Polchinski's equation, as well as the theory of optimal transport. In Section 3, we establish Equation (1) and a generalization pertaining to a broader class of ERG equations. In Section 4, we prove that the relative entropy appearing in our flow equations is in fact a non-perturbative RG monotone. In Section 5, we compute some explicit examples of the RG monotone for both a free and interacting scalar field. In Section 6, we leverage dual formulations of optimal transport to develop a variational formula for RG flows, and then explain how it can be leveraged for new numerical methods. Finally, in Section 7, we conclude with a discussion.

In the spirit of Polchinski's analysis, we restrict ourselves to scalar field theory for simplicity. We note that Polchinski's equation can be generalized to fermionic theories [**?**, **?**, 2] and gauge theories [**?**, **?**, **?**].

Let us recapitulate a version of Polchinski's derivation from [2]. Consider a Euclidean scalar field theory with a source J. We will set $\hbar = 1$ throughout. The partition function is

$$Z_\Lambda[J] := \int [d\phi] e^{-\frac{1}{2}\int \frac{d^dp}{(2\pi)^d}(\phi(p)\phi(-p)(p^2+m^2)K_\Lambda^{-1}(p^2)+J(p)\phi(-p)) - S_{\text{int},\Lambda}[\phi]}, \tag{15}$$

where $S_{\text{int},\Lambda}[\phi]$ includes interaction terms (possibly including quadratic terms which contribute to the explicit kinetic term) and where $K_\Lambda(p^2)$ is a soft cutoff function, i.e. it is 1 for $p^2 \lesssim \Lambda^2$ and ≈ 0 for $p^2 \gtrsim \Lambda^2$, and $K_\Lambda^{-1}(p^2)$ denotes $1/K_\Lambda(p^2)$. This soft cutoff function ensures that correlation functions are regulated at high momentum. For our purposes, it will be convenient for $K_\Lambda(p^2)$ to never equal zero, even if it is extremely close to zero; this way $K_\Lambda^{-1}(p^2)$ is never strictly infinite. An example of a soft cutoff function is shown in Figure 1. Also note that the mass m appearing above in (2.1) is the bare mass, and the couplings implicit in $S_{\text{int},\Lambda}[\phi]$ are bare couplings.

We desire to consider some smaller scale Λ_R and integrate out all modes down to Λ_R. As such, we are only interested in computing correlation functions below the scale Λ_R, and so let us assume that our source satisfies $J(p) = 0$ for $p^2 > \Lambda_R^2 - \epsilon$ for some small $\epsilon > 0$. It is convenient to restrict $|m^2| \ll \Lambda_R$, i.e. we are not integrating out the mass scale.

Suppose that Λ_R is infinitesimally smaller than Λ. Then we would like for

$$-\Lambda\frac{d}{d\Lambda}Z_\Lambda[J] = C_\Lambda Z_\Lambda[J] \tag{16}$$

for some constant C_Λ only depending on Λ. This would mean that as we change the cutoff scale Λ, which both affects the kinetic term in the action in an explicit way and the interaction terms in a way to be determined, any correlation functions below the changed scale (i.e., generated by taking functional J derivatives) stay the same. Expanding out the left-hand side we find

$$-\Lambda\frac{d}{d\Lambda}Z_\Lambda[J] = \int [d\phi]\left(\frac{1}{2}\int\frac{d^dp}{(2\pi)^d}\phi(p)\phi(-p)(p^2+m^2)\Lambda\frac{\partial K_\Lambda^{-1}(p^2)}{\partial\Lambda} + \Lambda\frac{\partial S_{\text{int},\Lambda}[\phi]}{\partial\Lambda}\right)e^{-S_\Lambda[\phi,J]}. \tag{17}$$

If we want (2.2) to hold, then $\Lambda\frac{\partial S_{\text{int},\Lambda}[\phi]}{\partial\Lambda}$ must have an appropriate form to facilitate this. Remarkably, Polchinski found such a sufficient form which corresponds to a spatially local coarse-graining of $S_{\text{int},\Lambda}[\phi]$ upon Fourier-transforming to position space. In particular, we will demand that $S_{\text{int},\Lambda}[\phi]$ changes with respect to Λ via

$$-\Lambda\frac{\partial S_{\text{int},\Lambda}[\phi]}{\partial\Lambda} = \frac{1}{2}\int\frac{d^dp}{(2\pi)^d}(p^2+m^2)^{-1}\Lambda\frac{\partial K_\Lambda(p^2)}{\partial\Lambda}\left(\frac{\delta^2 S_{\text{int},\Lambda}}{\delta\phi(p)\delta\phi(-p)} - \frac{\delta S_{\text{int},\Lambda}}{\delta\phi(p)}\frac{\delta S_{\text{int},\Lambda}}{\delta\phi(-p)}\right). \tag{18}$$

This is what is known as Polchinski's equation, and it is sometimes written as

$$-\Lambda \frac{\partial}{\partial \Lambda} e^{-S_{\text{int},\Lambda}[\phi]} = \frac{1}{2} \int \frac{d^d p}{(2\pi)^d} (p^2 + m^2)^{-1} \Lambda \frac{\partial K_\Lambda(p^2)}{\partial \Lambda} \frac{\delta^2}{\delta\phi(p)\delta\phi(-p)} e^{-S_{\text{int},\Lambda}[\phi]} \tag{19}$$

in order to resemble a functional version of the heat equation. Note the appearance of $\Lambda \frac{\partial K_\Lambda(p^2)}{\partial \Lambda}$ in both (2.4) and (2.5); this is localized in momentum space around $p^2 = \Lambda^2$, corresponding to a smearing kernel with scale $\sim 1/\Lambda$ in position space. See Figure 2 for a depiction in momentum space. Plugging (2.4) into (2.3) and simplifying, we find

$$-\Lambda \frac{d}{d\Lambda} Z_\Lambda[J] = \left(\frac{1}{2} \int \frac{d^d p \Lambda \frac{\partial K_\Lambda(p^2)}{\partial \Lambda}}{(2\pi)^d} \delta^{(d)}(0) \right) Z_\Lambda[J], \tag{20}$$

which has the form of the desired transformation from (2.2).

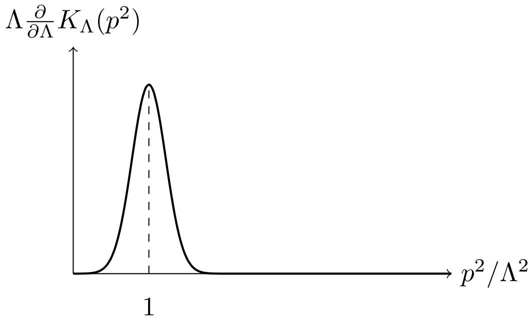

Figure 1: The derivative $\Lambda \frac{\partial}{\partial \Lambda} K_\Lambda(p^2)$ of the smooth cutoff function.

In the spirit of Polchinski's analysis, we restrict ourselves to scalar field theory for simplicity. We note that Polchinski's equation can be generalized to fermionic theories [?, ?, 2] and gauge theories [?, ?, ?].

Let us recapitulate a version of Polchinski's derivation from [2]. Consider a Euclidean scalar field theory with a source J. We will set $\hbar = 1$ throughout. The partition function is

$$Z_\Lambda[J] := \int [d\phi] e^{-\frac{1}{2} \int \frac{d^d p}{(2\pi)^d} (\phi(p)\phi(-p)(p^2+m^2)K_\Lambda^{-1}(p^2) + J(p)\phi(-p)) - S_{\text{int},\Lambda}[\phi]}, \tag{21}$$

where $S_{\text{int},\Lambda}[\phi]$ includes interaction terms (possibly including quadratic terms which contribute to the explicit kinetic term) and where $K_\Lambda(p^2)$ is a soft cutoff function, i.e. it is 1 for $p^2 \lesssim \Lambda^2$ and ≈ 0 for $p^2 \gtrsim \Lambda^2$, and $K_\Lambda^{-1}(p^2)$ denotes $1/K_\Lambda(p^2)$. This soft cutoff function ensures that correlation functions are regulated at high momentum. For our purposes, it will be convenient for $K_\Lambda(p^2)$ to never equal zero, even if it is extremely close to zero; this way $K_\Lambda^{-1}(p^2)$ is never strictly infinite. An example of a soft cutoff function is shown in Figure 1. Also note that the mass m appearing above in (2.1) is the bare mass, and the couplings implicit in $S_{\text{int},\Lambda}[\phi]$ are bare couplings.

We desire to consider some smaller scale Λ_R and integrate out all modes down to Λ_R. As such, we are only interested in computing correlation functions below the scale Λ_R, and so let us assume that our source satisfies $J(p) = 0$ for $p^2 > \Lambda_R^2 - \epsilon$ for some small $\epsilon > 0$. It is convenient to restrict $|m^2| \ll \Lambda_R$, i.e. we are not integrating out the mass scale.

Suppose that Λ_R is infinitesimally smaller than Λ. Then we would like for

$$-\Lambda \frac{d}{d\Lambda} Z_\Lambda[J] = C_\Lambda Z_\Lambda[J] \tag{22}$$

for some constant C_Λ only depending on Λ. This would mean that as we change the cutoff scale Λ, which both affects the kinetic term in the action in an explicit way and the interaction terms in a way to be determined, any correlation functions below the changed scale (i.e., generated by taking functional J derivatives) stay the same. Expanding out the left-hand side we find

$$-\Lambda \frac{d}{d\Lambda} Z_\Lambda[J] = \int [d\phi] \left(\frac{1}{2} \int \frac{d^d p}{(2\pi)^d} \phi(p)\phi(-p)(p^2+m^2) \Lambda \frac{\partial K_\Lambda^{-1}(p^2)}{\partial \Lambda} + \Lambda \frac{\partial S_{\text{int},\Lambda}[\phi]}{\partial \Lambda} \right) e^{-S_\Lambda[\phi, J]} \tag{23}$$

If we want (2.2) to hold, then $\Lambda \frac{\partial S_{\text{int},\Lambda}[\phi]}{\partial \Lambda}$ must have an appropriate form to facilitate this. Remarkably, Polchinski found such a sufficient form which corresponds to a spatially local coarse-graining of $S_{\text{int},\Lambda}[\phi]$ upon Fourier-transforming to position space. In particular, we will demand that $S_{\text{int},\Lambda}[\phi]$ changes with respect to Λ via

$$-\Lambda \frac{\partial S_{\text{int},\Lambda}[\phi]}{\partial \Lambda} = \frac{1}{2} \int \frac{d^d p}{(2\pi)^d} (p^2 + m^2)^{-1} \Lambda \frac{\partial K_\Lambda(p^2)}{\partial \Lambda} \left(\frac{\delta^2 S_{\text{int},\Lambda}}{\delta \phi(p) \delta \phi(-p)} - \frac{\delta S_{\text{int},\Lambda}}{\delta \phi(p)} \frac{\delta S_{\text{int},\Lambda}}{\delta \phi(-p)} \right) \tag{24}$$

While Polchinski's equation (2.4) is formulated in terms of a functional equation for $S_{\text{int},\Lambda}[\phi]$, it will be convenient for us to recast it in terms of a functional equation for the probability functional $P_\Lambda[\phi] = e^{-S_\Lambda[\phi]}/Z_\Lambda$. Reprocessing the above derivation we arrive at

$$-\Lambda \frac{d}{d\Lambda} P_\Lambda[\phi] = \frac{1}{2} \int \frac{d^d p}{(2\pi)^d} (p^2+m^2)^{-1} \Lambda \frac{\partial K_\Lambda(p^2)}{\partial \Lambda} \frac{\delta^2}{\delta\phi(p)\delta\phi(-p)} P_\Lambda[\phi] + \int d^d p \Lambda \frac{\partial \log K_\Lambda(p^2)}{\partial \Lambda} \frac{\delta}{\delta \phi(p)} (\phi(p) P_\Lambda[\phi]) \tag{25}$$

which has the form of a functional convection-diffusion equation. To see the connection more clearly, we rewrite the above as

$$-\Lambda \frac{d}{d\Lambda} P_\Lambda[\phi] = \frac{1}{2} \int \frac{d^d p}{(2\pi)^d} (p^2+m^2)^{-1} \Lambda \frac{\partial K_\Lambda(p^2)}{\partial \Lambda} \frac{\delta^2}{\delta\phi(p)\delta\phi(-p)} P_\Lambda[\phi] + \frac{1}{2} \int \frac{d^d p}{(2\pi)^d} (p^2+m^2)^{-1} \Lambda \frac{\partial K_\Lambda(p^2)}{\partial \Lambda} \frac{\delta}{\delta \phi(p)} \left(2\phi(p) + \frac{(2}{p^2} \right. \tag{26}$$

which formally takes the same form as the finite-dimensional convection-diffusion equation

$$\frac{d}{dt} p_t(x) = \partial_i \partial_i p_t(x) + \partial_i (v^i(x) p_t(x)) \tag{27}$$

where we identify $-\log \Lambda$ with t. An example of a solution to (2.7) (or equivalently (2.8)) is the free theory itself; that is, $P_\Lambda[\phi] = \frac{1}{Z_\Lambda} \exp\left(-\frac{1}{2} \int \frac{d^d p}{(2\pi)^d} \phi(p)\phi(-p)(p^2+m^2) K_\Lambda^{-1}(p^2) \right)$ solves Polchinski's equation.

Let us summarize the logic of Polchinski's derivation. We explicitly differentiated $Z_\Lambda[J]$ by $\Lambda \frac{d}{d\Lambda}$ and then found a choice of $\Lambda \frac{\partial S_{\text{int},\Lambda}[\phi]}{\partial \Lambda}$ such that $-\Lambda \frac{d}{d\Lambda} Z_\Lambda[J] = C_\Lambda Z_\Lambda$ is satisfied for a constant C_Λ. The suitable choice of $\Lambda \frac{\partial S_{\text{int},\Lambda}[\phi]}{\partial \Lambda}$, given by the functional differential equation in (2.4), corresponds to changing $S_{\text{int},\Lambda}$ in a manner which is localized in momentum space at scale Λ, and hence local in position space at scale $\sim 1/\Lambda$. While Polchinski's inspired ansatz (2.4) does the job, there are in fact an infinitude of other choices which have similar properties and also render $-\Lambda \frac{d}{d\Lambda} Z_\Lambda = C_\Lambda Z_\Lambda$. These other choices correspond to alternative RG schemes than the one proposed by Polchinski. We explore a large family of them via our discussion of the Wegner-Morris flow equation below.

7.1 Wegner-Morris flow equation

In the spirit of Polchinski's analysis, we restrict ourselves to scalar field theory for simplicity. We note that Polchinski's equation can be generalized to fermionic theories [**?**, **?**, 2] and gauge theories [**?**, **?**, **?**].

Let us recapitulate a version of Polchinski's derivation from [2]. Consider a Euclidean scalar field theory with a source J. We will set $\hbar = 1$ throughout. The partition function is

$$Z_\Lambda[J] := \int [d\phi] e^{-\frac{1}{2} \int \frac{d^d p}{(2\pi)^d} (\phi(p)\phi(-p)(p^2+m^2) K_\Lambda^{-1}(p^2) + J(p)\phi(-p)) - S_{\text{int},\Lambda}[\phi]}, \tag{28}$$

where $S_{\text{int},\Lambda}[\phi]$ includes interaction terms (possibly including quadratic terms which contribute to the explicit kinetic term) and where $K_\Lambda(p^2)$ is a soft cutoff function, i.e. it is 1 for $p^2 \lesssim \Lambda^2$ and ≈ 0 for $p^2 \gtrsim \Lambda^2$, and $K_\Lambda^{-1}(p^2)$ denotes $1/K_\Lambda(p^2)$. This soft cutoff function ensures that correlation functions are regulated at high momentum. For our purposes, it will be convenient for $K_\Lambda(p^2)$ to never equal zero, even if it is extremely close to zero; this way $K_\Lambda^{-1}(p^2)$ is never strictly infinite. An example of a soft cutoff function is shown in Figure 1. Also note that the mass m appearing above in (2.1) is the bare mass, and the couplings implicit in $S_{\text{int},\Lambda}[\phi]$ are bare couplings.

We desire to consider some smaller scale Λ_R and integrate out all modes down to Λ_R. As such, we are only interested in computing correlation functions below the scale Λ_R, and so let us assume that our source satisfies $J(p) = 0$ for $p^2 > \Lambda_R^2 - \epsilon$ for some small $\epsilon > 0$. It is convenient to restrict $|m^2| \ll \Lambda_R$, i.e. we are not integrating out the mass scale.

Suppose that Λ_R is infinitesimally smaller than Λ. Then we would like for

$$-\Lambda \frac{d}{d\Lambda} Z_\Lambda[J] = C_\Lambda Z_\Lambda[J] \tag{29}$$

for some constant C_Λ only depending on Λ. This would mean that as we change the cutoff scale Λ, which both affects the kinetic term in the action in an explicit way and the interaction terms in a way to be determined, any correlation functions below the changed scale (i.e., generated by taking functional J derivatives) stay the same. Expanding out the left-hand side we find

$$-\Lambda\frac{d}{d\Lambda}Z_\Lambda[J] = \int [d\phi]\left(\frac{1}{2}\int\frac{d^dp}{(2\pi)^d}\phi(p)\phi(-p)(p^2+m^2)\Lambda\frac{\partial K_\Lambda^{-1}(p^2)}{\partial\Lambda} + \Lambda\frac{\partial S_{\mathrm{int},\Lambda}[\phi]}{\partial\Lambda}\right)e^{-S_\Lambda[\phi,J]} \tag{30}$$

If we want (2.2) to hold, then $\Lambda\frac{\partial S_{\mathrm{int},\Lambda}[\phi]}{\partial\Lambda}$ must have an appropriate form to facilitate this. Remarkably, Polchinski found such a sufficient form which corresponds to a spatially local coarse-graining of $S_{\mathrm{int},\Lambda}[\phi]$ upon Fourier-transforming to position space. In particular, we will demand that $S_{\mathrm{int},\Lambda}[\phi]$ changes with respect to Λ via

$$-\Lambda\frac{\partial S_{\mathrm{int},\Lambda}[\phi]}{\partial\Lambda} = \frac{1}{2}\int\frac{d^dp}{(2\pi)^d}(p^2+m^2)^{-1}\Lambda\frac{\partial K_\Lambda(p^2)}{\partial\Lambda}\left(\frac{\delta^2 S_{\mathrm{int},\Lambda}}{\delta\phi(p)\delta\phi(-p)} - \frac{\delta S_{\mathrm{int},\Lambda}}{\delta\phi(p)}\frac{\delta S_{\mathrm{int},\Lambda}}{\delta\phi(-p)}\right) \tag{31}$$

This is what is known as Polchinski's equation, and it is sometimes written as

$$-\Lambda\frac{\partial}{\partial\Lambda}e^{-S_{\mathrm{int},\Lambda}[\phi]} = \frac{1}{2}\int\frac{d^dp}{(2\pi)^d}(p^2+m^2)^{-1}\Lambda\frac{\partial K_\Lambda(p^2)}{\partial\Lambda}\frac{\delta^2}{\delta\phi(p)\delta\phi(-p)}e^{-S_{\mathrm{int},\Lambda}[\phi]} \tag{32}$$

in order to resemble a functional version of the heat equation. Note the appearance of $\Lambda\frac{\partial K_\Lambda(p^2)}{\partial\Lambda}$ in both (2.4) and (2.5); this is localized in momentum space around $p^2 = \Lambda^2$, corresponding to a smearing kernel with scale $\sim 1/\Lambda$ in position space. See Figure 2 for a depiction in momentum space. Plugging (2.4) into (2.3) and simplifying, we find

$$-\Lambda\frac{d}{d\Lambda}Z_\Lambda[J] = \left(\frac{1}{2}\int\frac{d^dp\Lambda\frac{\partial K_\Lambda(p^2)}{\partial\Lambda}}{(2\pi)^d}\delta^{(d)}(0)\right)Z_\Lambda[J] \tag{33}$$

which has the form of the desired transformation from (2.2).

While Polchinski's equation (2.4) is formulated in terms of a functional equation for $S_{\mathrm{int},\Lambda}[\phi]$, it will be convenient for us to recast it in terms of a functional equation for the probability functional $P_\Lambda[\phi] = e^{-S_\Lambda[\phi]}/Z_\Lambda$. Reprocessing the above derivation we arrive at

$$-\Lambda\frac{d}{d\Lambda}P_\Lambda[\phi] = \frac{1}{2}\int\frac{d^dp}{(2\pi)^d}(p^2+m^2)^{-1}\Lambda\frac{\partial K_\Lambda(p^2)}{\partial\Lambda}\frac{\delta^2}{\delta\phi(p)\delta\phi(-p)}P_\Lambda[\phi] + \int d^dp\Lambda\frac{\partial\log K_\Lambda(p^2)}{\partial\Lambda}\frac{\delta}{\delta\phi(p)}(\phi(p)P_\Lambda[\phi]) \tag{34}$$

which has the form of a functional convection-diffusion equation. To see the connection more clearly, we rewrite the above as

$$-\Lambda\frac{d}{d\Lambda}P_\Lambda[\phi] = \frac{1}{2}\int\frac{d^dp}{(2\pi)^d}(p^2+m^2)^{-1}\Lambda\frac{\partial K_\Lambda(p^2)}{\partial\Lambda}\frac{\delta^2}{\delta\phi(p)\delta\phi(-p)}P_\Lambda[\phi] + \frac{1}{2}\int\frac{d^dp}{(2\pi)^d}(p^2+m^2)^{-1}\Lambda\frac{\partial K_\Lambda(p^2)}{\partial\Lambda}\frac{\delta}{\delta\phi(p)}\left(2\phi(p) + \frac{(2\cdots}{p^2\cdots}\right. \tag{35}$$

which formally takes the same form as the finite-dimensional convection-diffusion equation

$$\frac{d}{dt}p_t(x) = \partial_i\partial_i p_t(x) + \partial_i(v^i(x)p_t(x)) \tag{36}$$

where we identify $-\log\Lambda$ with t. An example of a solution to (2.7) (or equivalently (2.8)) is the free theory itself; that is, $P_\Lambda[\phi] = \frac{1}{Z_\Lambda}\exp\left(-\frac{1}{2}\int\frac{d^dp}{(2\pi)^d}\phi(p)\phi(-p)(p^2+m^2)K_\Lambda^{-1}(p^2)\right)$ solves Polchinski's equation.

Let us summarize the logic of Polchinski's derivation. We explicitly differentiated $Z_\Lambda[J]$ by $\Lambda\frac{d}{d\Lambda}$ and then found a choice of $\Lambda\frac{\partial S_{\mathrm{int},\Lambda}[\phi]}{\partial\Lambda}$ such that $-\Lambda\frac{d}{d\Lambda}Z_\Lambda[J] = C_\Lambda Z_\Lambda$ is satisfied for a constant C_Λ. The suitable choice of $\Lambda\frac{\partial S_{\mathrm{int},\Lambda}[\phi]}{\partial\Lambda}$, given by the functional differential equation in (2.4), corresponds to changing $S_{\mathrm{int},\Lambda}$ in a manner which is localized in momentum space at scale Λ, and hence local in position space at scale $\sim 1/\Lambda$. While Polchinski's inspired ansatz (2.4) does the job, there are in fact an infinitude of other choices which have similar properties and also render $-\Lambda\frac{d}{d\Lambda}Z_\Lambda = C_\Lambda Z_\Lambda$. These other choices correspond to alternative RG schemes than the one proposed by Polchinski. We explore a large family of them via our discussion of the Wegner-Morris flow equation below.

Polchinski's equation is a special case of the Wegner-Morris flow equation [?, ?, 14]. The latter provides insights into the structure of RG flows which are obscured by Polchinski's formulation. The Wegner-Morris equation is

$$-\Lambda \frac{d}{d\Lambda} P_\Lambda[\phi] = \int d^d x \frac{\delta}{\delta\phi(x)} \left(\nabla\phi(x) P_\Lambda[\phi]\right) \tag{37}$$

and implements ERG for a scheme determined by $\Psi_\Lambda[\phi, x]$. We note that $\Psi_\Lambda[\phi, x]$ will depend on $P_\Lambda[\phi]$ in a non-trivial way, which we explain below. At first glance (2.10) does not appear to readily connect to RG flow, but its meaning will be clear shortly.

To gain some intuition for (2.10), it is useful to compare with a finite-dimensional analog. This would be the equation for p_t given by

$$\frac{d}{dt} p_t(x) + \partial_i (V^i(p_t, x) p_t) = 0. \tag{38}$$

In this equation the vector field V^i, the analog of Ψ_Λ in the Wegner-Morris flow, is chosen to depend not just on the coordinate position x but also on the entire probability distribution p_t. It is natural for V^i to satisfy $V^i(p_t, x) = \partial_i W(p_t, x)$, namely for V^i to have a potential W. This gives us

$$\frac{d}{dt} p_t(x) + \partial_i (\partial_i W(p_t, x) p_t) = 0. \tag{39}$$

We will find an analogue of this potential in the Wegner-Morris flow equation for many cases of interest.

Equation (2.10) has several features which illuminate its meaning. First, performing the functional integral of both sides of (2.10) with respect to $\phi(x)$ and noting that $\Psi_\Lambda[\phi, x] P_\Lambda[\phi]$ goes to zero for large $\phi(x)$, we immediately see that $-\Lambda \frac{d}{d\Lambda} \int [d\phi] P_\Lambda[\phi] = 0$ and so the flow equation preserves probability. More generally, the meaning of (2.10) is that as the scale Λ changes the flow induces the field reparameterization

$$\phi'(x) = \phi(x) + \frac{\delta\Lambda}{\Lambda} \Psi_\Lambda[\phi, x]. \tag{40}$$

This means that the probability functional is simply reparameterized by the flow, and so probability is clearly conserved and positivity of the probability density is maintained. As explained in [14], essentially all RG schemes (with a soft cutoff) can be cast into the form of the Wegner-Morris flow equation. In all schemes Ψ_Λ instantiates field redefinitions which are localized in momentum space near scale Λ, i.e. we are reparameterizing the field at or near the cutoff scale. We will henceforth refer to Ψ_Λ as the reparameterization kernel.

A common form of $\Psi_\Lambda[\phi, x]$ is given by [?, ?, 14]

$$\Psi_\Lambda[\phi, x] = -\int d^d y \frac{1}{2} C_\Lambda(x-y) \frac{\delta \Sigma_\Lambda[\phi]}{\delta\phi(y)}, \tag{41}$$

where $C_\Lambda(x-y)$ is called the ERG kernel which satisfies $C_\Lambda(x-y) \geq 0$, and

$$\Sigma_\Lambda[\phi] := S_\Lambda[\phi] - 2 S_\Lambda[\phi], \tag{42}$$

where $S_\Lambda[\phi]$ is the action appearing in $P_\Lambda[\phi] = e^{-S_\Lambda[\phi]}/Z_\Lambda$ and $\hat{S}_\Lambda[\phi]$ is another action called the 'seed action'. The multiplicative factor of 2 in front of the seed action is conventional. In $P_\Lambda[\phi] = e^{-S_\Lambda[\phi]}/Z_\Lambda$ and $\hat{S}_\Lambda[\phi]$ is another action called the 'seed action'. The multiplicative factor of 2 in front of the seed action is conventional. In its present form, the meaning of the seed action is physically obscure. Fortunately, our optimal transport analysis later on will elucidate its meaning. Notice that $\Psi_\Lambda[\phi, x]$ is a gradient of $\Sigma_\Lambda[\phi]$, where $\frac{1}{2} C_\Lambda(x-y)$ plays the role of an inverse metric, and so in this setting the Wegner-Morris flow equation (2.10) takes the form of the finite-dimensional equation (2.12).

Importantly, we can reproduce the Polchinski's equation with the choices

$$C_\Lambda(p^2) = (2\pi)^d (p^2 + m^2)^{-1} \Lambda \frac{\partial K_\Lambda(p^2)}{\partial\Lambda} \tag{43}$$

$$\hat{S}_\Lambda = \frac{1}{2} \int \frac{d^d p}{(2\pi)^d} (p^2 + m^2) K_\Lambda^{-1}(p^2) \phi(p) \phi(-p) \tag{44}$$

here expressed in momentum space. Notice that $\hat{S}$ is just an action for a free massive scalar field with the same initial bare mass as our scalar field theory of interest.

An initially puzzling feature of Wegner-Morris flow is that (2.13) can be inverted if $\Psi_\Lambda[\phi, x]$ is well enough behaved. This would mean that the exact RG flow is invertible. However, we often think of RG as being non-invertible, perhaps the most famous example being Kadanoff's block spin decimation for spin systems (see e.g. [**?**, 19]). For continuum field theories, exact RG flows are typically invertible, although the inversion is ill-conditioned. As an example close in spirit to Kadanoff's block spin methods, suppose that our RG flow is prescribed by the coarse-graining $P_\Lambda[\phi] = \int [d\phi']\delta[\phi - b_\Lambda[\phi']] P_{\Lambda_0}[\phi']$ where Λ_0 is the initial RG scale and $\Lambda \leq \Lambda_0$. Here $b_\Lambda[\phi'](x) := \int d^d y f_\Lambda(x-y)\psi(y)$, where $f_\Lambda(x-y)$ is a smearing kernel with width $\sim 1/\Lambda$ in position space. Perhaps $f_\Lambda(x-y)$ is a Gaussian distribution, or a d-dimensional unit box function (which has compact support). So we are performing a continuum version of Kadanoff's procedure. However, a key difference is that the smearing $\int d^d y f_\Lambda(x-y)\psi(y)$ is invertible in the continuum as can be seen by transforming to Fourier space to get $f_\Lambda(p)\psi(p)$ and dividing by $f_\Lambda(p)$. Indeed, if $f_\Lambda(x-y)$ is a Gaussian, then so is its Fourier transform; dividing by a Gaussian is well-defined, albeit ill-conditioned since we are dividing by very small numbers in the tail regions. Likewise the Fourier transform of a box function is the product of sinc functions, and division by them is likewise ill-conditioned.

More broadly, even when we perform exact versions of the more standard Wilsonian RG, the flow is only in general invertible if we keep the infinitely many irrelevant terms in the action generated by the flow.

8 Optimal Transport

As discussed above, Polchinski's equation for P_Λ is an infinite-dimensional convection-diffusion equation, which can be thought of as a generalized form of heat flow. The RG monotones we present later on will be analogs of the entropy of a distribution. The fact that the entropy of a distribution is monotone along heat flows was already known to Gibbs. However, the understanding that the entropy functional generates heat flow under the Wasserstein metric required a synthesis [25] of ideas about optimal transport. This synthesis occurred relatively recently in the 90's, in the work of Otto, Benamou-Brenier, and many others. We will review some of these developments here.

At a high level, the problem of optimal transportation is to determine an optimal method for moving and rearranging a given mass distribution into a desired mass distribution, given a cost for moving mass across a specified distance. In the next three subsections, we will review the basic mathematical formalization, discuss fundamental results about this problem, and explain how it connects with heat flow. Beyond this connection, there is a rich theory connecting optimal transport with probability theory and mathematical physics, and we will provide a short guide to relevant literature for interested readers.

8.1 Monge and Kantorovich formulations

Given a space X and a pair of probability or mass distributions p, q on X, the Monge formulation of the optimal transport problem asks to find a (measurable) transport function $T : X \to X$ such that:

1. The pushforward of p under T is q, i.e. $T_\# p = q$; equivalently $\int_{T^{-1}(S)} dx\, p(x) = \int_S dx\, q(x)$ for every measurable set S; and
2. The transport function minimizes the total cost
$$\mathrm{M}[T] = \int_X dx\, p(x)\, c(x, T(x)) \tag{45}$$
for some cost function $c : X \times X \to \mathbb{R}$.

A natural choice for the cost function is $c(x,y) = d(x,y)^2$, where d is a distance function on X. A depiction of the mapping $T_\# p = q$ can be seen in Figure 3.

The constraint $T_\# p = q$ is highly nonlinear, making the existence of a solution non-obvious. For concreteness, suppose that $X = \mathbb{R}^n$ and T is a smooth function. We let T_j denote the jth coordinate output of T. Then the constraint can be written as
$$q(T(x))\, |\det(\partial_i T_j(x))| = p(x). \tag{46}$$

This nonlinear constraint above makes it difficult to establish the existence of solutions to the Monge problem via methods from the calculus of variations. Worse, solutions to the Monge problem no longer exist once the distributions are not smooth: if the distributions p, q are sums of delta functions, i.e. $p(x) = \sum_i p_i \delta(x - a_i)$ while $q(x) = \sum_j q_j \delta(x - b_i)$, then for generic choices of supports $\{a_i\}, \{b_j\}$, it is clear that no transport map T exists. For instance, if p is supported on one point and q is supported on two points, there is no transport map T such that $T_{\#} p = q$.

To better understand the Monge problem, it is convenient to first solve a relaxation known as the Kantorovich problem. In the Kantorovich problem, one searches for a positive measure π on $X \times X$ such that:

1. The pushforward of π to X is p, and the pushforward of π to Y is q (i.e. $\int_Y dx\, \pi(x, y) = p(x)$ and $\int_X dy\, \pi(x, y) = q(y)$); and
2. The measure π minimizes
$$K(\pi) = \int_{X \times X} dx\, dy\, \pi(x, y)\, c(x, y). \tag{47}$$

The interpretation of $dx\, dy\, \pi(x, y)$ is that it is the infinitesimal amount of mass at x which is transported to y. If we set $\pi_{xy} = p(x)\delta(y - T(x))$ then it is clear that $K(\pi) = \mathrm{M}[T]$. Thus, candidate solutions to the Monge problem give candidate solutions to the Kantorovich problem. However, the Kantorovich problem is much easier, as it is a problem in finite-dimensional convex optimization. Indeed, the function $K(\pi)$ is a linear function on the convex cone of positive measures on $X \times X$ and the constraints arising from p and q are also linear. Discretizing this optimization problem yields a familiar finite-dimensional linear program: if $p(x) = \sum_i p_i \delta(x - a_i)$, $q(y) = \sum_j q_j \delta(y - b_j)$, and $\pi(x, y) = \sum_{i,j} \pi_{ij} \delta(x - a_i)\delta(y - b_j)$, then the Kantorovich problem immediately reduces to

$$\text{Minimize} \quad \sum_{i,j} \pi_{ij} c(a_i, b_j) \quad \text{subject to} \quad \sum_j \pi_{ij} \geq 0, \quad \sum_j \pi_{ij} = p_i, \quad \sum_i \pi_{ij} = q_j. \tag{48}$$

Despite that fact that the Kantorovich problem is a relaxation of the Monge problem, in a large class of cases solutions to the Kantorovich problem actually arise from solutions to the Monge problem:

Theorem 2.1. *If $p(x)$ and $q(x)$ are smooth functions having support on all of $\mathbb{R}^n$, then the Monge problem with $c(x, y) = |x - y|^2$ has a smooth solution; indeed, we have*
$$T_i(x) = \partial_i f(x) \tag{49}$$
for some smooth function $f : \mathbb{R}^n \to \mathbb{R}$.

This result requires the development of a significant amount of mathematics: it follows from a combination of duality for the Kantorovich problem, Brenier's theorem [22], and Caffarelli's regularity theory [23] for solutions to the Monge-Ampère equation. To explain the proof of this theorem would take us too far afield, although [47] gives a good introduction. We only note that once the existence of a function satisfying (2.22) is established by the duality theory, one concludes by the constraint (2.19) that f satisfies the equation
$$\det(\text{Hess } f(x)) = \frac{p(x)}{q(\nabla f(x))} \tag{50}$$
which is the form of the Monge-Ampère equation that appears in this setting.

8.1.1 Wasserstein distance

Since we will be primarily interested in cases for which the space X is a metric space (and often a Riemannian metric space), we will henceforth denote the space by M. For the quadratic cost $c(x, y) = |x - y|^2$ where x, y are Cartesian coordinates on Euclidean space $M = \mathbb{R}^n$, the optimum value of $K(\pi)$ in the Kantorovich problem is called the Wasserstein-2 distance $W_2(p_1, p_2)$. (This is alternatively called the L^2-Wasserstein distance.) The distance can be written as

$$W_2(p_1, p_2) := \left(\inf_{\pi \in \Gamma(p_1, p_2)} \int_{M \times M} dx\, dy\, \pi(x, y)\, |x - y|^2 \right)^{1/2} \tag{51}$$

where $\Gamma(p_1, p_2)$ is the space of probability distributions $\pi(x, y)$ on $M \times M$ such that $\int_M dy\, \pi(x, y) = p(x)$ and $\int_M dx\, \pi(x, y) = q(y)$. This metric distance on the space of probability distributions, and in particular various path integral generalizations of it, will play a central role in our analyses.

8.1.2 Otto calculus

To explain the connection between heat flow and optimal transport, we first recall how to view heat flow as a gradient flow with respect to the usual L^2 metric.

Heat flow as gradient flow of Dirichlet energy. For a function $F : \mathcal{M} \to \mathbb{R}$ on a Riemannian manifold $\mathcal{M}$ with metric $\langle \cdot, \cdot \rangle$, the gradient of F at $x_0 \in \mathcal{M}$ is the vector $\nabla F(x_0)$ such that

$$\frac{d}{dt} F(x(t)) \bigg|_{t=0} = \left\langle \nabla F(x_0), \frac{\partial x(t)}{\partial t} \right\rangle \bigg|_{t=0} \tag{52}$$

for every curve $x(t) \in \mathcal{M}$ with $x(0) = x_0$.

For our purposes, we let $\mathcal{M} = \mathrm{dens}(M)$ be the space probability densities on a manifold M, where we suppose M is equipped with a volume form dV. That is, $\mathrm{dens}(M)$ is an infinite-dimensional manifold defined by

$$\mathrm{dens}(M) := \left\{ p \in C^\infty(M) \mid p \geq 0, \int dV \, p = 1 \right\}. \tag{53}$$

The tangent space at $p \in \mathrm{dens}(M)$ is

$$T_p \mathrm{dens}(M) = \left\{ \eta \in C^\infty(M) \middle| \int dV \, \eta = 0 \right\}. \tag{54}$$

We can equip each tangent space $T_p\mathrm{dens}(M)$ with a Riemannian metric

$$\langle \eta_1, \eta_2 \rangle_{L^2} = \int dV \, \eta_1 \eta_2. \tag{55}$$

This corresponds to the L^2 inner product on functions on M. Defining the Dirichlet energy functional as

$$\mathcal{E}[p] := \frac{1}{2} \int dV \, |\nabla p|^2, \tag{56}$$

we can compute its gradient with respect to the infinite-dimensional L^2 metric in (2.28) using (2.25). In particular, let $\rho(t)$ be a differentiable path through $\mathrm{dens}(M)$ such that $\rho(0) = p$. Then

$$\frac{d}{dt} \mathcal{E}[\rho(t)] = \int dV \, \nabla p \cdot \nabla \frac{\partial \rho}{\partial t} = - \int dV \, \Delta p \frac{\partial \rho}{\partial t} = \left\langle -\Delta p, \frac{\partial \rho}{\partial t} \right\rangle_{L^2}. \tag{57}$$

Evaluating the above at $t = 0$ and comparing with (2.25), we read off that

$$\nabla_{L^2} \mathcal{E}[p] = -\Delta p. \tag{58}$$

It follows that the heat equation $\frac{\partial p}{\partial t} = \Delta p$ is the negative gradient flow of the Dirichlet energy functional $\mathcal{E}$, namely

$$\frac{\partial p(x,t)}{\partial t} = -\nabla_{L^2} \mathcal{E}[p(x,t)]. \tag{59}$$

This in fact implies that the Dirichlet energy monotonically decreases along the heat flow.

8.1.3 Wasserstein distance and the gradient flow of entropy

Another monotone for the heat flow is given by the differential entropy

$$\mathcal{S}[p] := - \int dV \, p \log(p). \tag{60}$$

We will have more to say about this quantity in Subsection 3.2. By analogy with (2.32) above, we might ask if there is any Riemannian metric g on $\mathrm{dens}(M)$ such that the heat equation can be written as $\frac{\partial p}{\partial t} = \nabla_g \mathcal{S}[p]$? In other words, is there some (natural) metric on the space of probability distributions for which the heat equation is the gradient flow of the differential entropy?

Remarkably, the answer is yes—this was discovered by Otto [21] and widely exploited by subsequent researchers in partial differential equations and probability theory. In fact, there are a large collection of entropy-like monotones $\tilde{\mathcal{S}}$ which have associated metrics $\tilde{g}$ on $\mathrm{dens}(M)$ such that the heat equation can be written as $\frac{\partial p}{\partial t} = \nabla_{\tilde{g}} \tilde{\mathcal{S}}[p]$. All of these metrics have deep connections to optimal transport. Since we will be interested in the particular case of the differential entropy, we will not discuss these related entropic gradient flow formulations here.

We now turn to constructing the metric g on dens(M) such that $\frac{\partial p}{\partial t} = \nabla_g \mathcal{S}[p]$. To write the metric in the most transparent way, an isomorphism of the tangent space $T_p\text{dens}(M)$ is required. Given a tangent vector $\eta \in T_p\text{dens}(M)$, we can solve for a $\bar{\eta}$ satisfying

$$\nabla \cdot (p \nabla \bar{\eta}) = \eta. \tag{61}$$

The solution is unique up to addition of a constant, and so we get an identification $\eta \leftrightarrow \bar{\eta}$ which we notate by the isomorphism

$$T_p\text{dens}(M) \simeq \bar{T}_p\text{dens}(M) := \{\bar{\eta} \in C^\infty(M)/\{\text{constants}\}\}. \tag{62}$$

Using this identification we define the Riemannian metric

$$\langle \eta_1, \eta_2 \rangle_{W_2} := \int dV \, p \nabla \bar{\eta}_1 \cdot \nabla \bar{\eta}_2 = -\int dV \, \bar{\eta}_1 \Delta p \bar{\eta}_2 = \int dV \, \eta_1 \bar{\eta}_2 \tag{63}$$

where the last two equalities can be checked via integration by parts. This metric is in fact the infinitesimal form of the Wasserstein-2 distance W_2. A rigorous argument establishing this fact is given in [47]; we will explain the heuristic connection in Appendix A.

Now let us show that $\nabla_{W_2} \mathcal{S}[p] = \Delta p$. Let $\rho(t)$ be a path through dens(M) with $\rho(0) = p$, and define $\eta := \frac{d}{dt}\rho(t)\big|_{t=0}$ which is definitionally an element of $T_p\text{dens}(M)$. Let $\bar{\eta}$ be the corresponding solution to (2.34). Then we compute

$$\begin{aligned}
\frac{d}{dt}\mathcal{S}[\rho(t)]\bigg|_{t=0} &= -\int dV \, \eta(\log p + 1) \\
&= -\int dV \, \nabla \cdot (p \nabla \bar{\eta})(\log p + 1) \\
&= \int dV \, p \nabla \bar{\eta} \cdot \nabla p \\
&= -\int dV \, \bar{\eta} \cdot \Delta p \\
&= \langle \Delta p, \eta \rangle_{W_2} \\
&= \left\langle \Delta p, \frac{\partial \rho}{\partial t} \right\rangle_{W_2} \bigg|_{t=0}
\end{aligned} \tag{64}$$

and so comparing with (2.25) we indeed find

$$\nabla_{W_2} \mathcal{S}[p] = \Delta p. \tag{65}$$

Then the heat equation can be written as

$$\frac{\partial p(x,t)}{\partial t} = \nabla_{W_2} \mathcal{S}[p]. \tag{66}$$

Thus, the heat ow is the gradient ow of the differential entropy with respect to the Wasserstein-2 metric. While it was known to Gibbs that entropy is a heat ow monotone, the above equation clari

es that in fact heat ow is completely governed by the entropy, with optimal transport playing a central role in this formulation.

9 RG flow as an optimal transport gradient flow

9.1 Deriving the optimal transport gradient flow equation for RG

Since Polchinski's equation (2.7) is a special case of the Wegner-Morris flow equation (2.10), we find it prudent to derive our optimal transport equation for the latter. Suppose we intend to flow a Euclidean field theory with probability functional $P_\Lambda[\phi] = e^{-S_\Lambda[\phi]}/Z_{P,\Lambda}$. Recall the Wegner-Morris flow equation

$$-\Lambda \frac{d}{d\Lambda} P_\Lambda[\phi] = \int d^d x \frac{\delta}{\delta \phi(x)} \left(\Psi_\Lambda[\phi, x] P_\Lambda[\phi] \right) \tag{67}$$

where will adopt the functional forms in (2.14) and (2.15) for the reparameterization kernel Ψ_Λ, namely

$$\Psi_\Lambda[\phi, x] = -\int d^d y \frac{1}{2} C_\Lambda(x-y) \frac{\delta \Sigma_\Lambda[\phi]}{\delta \phi(y)}, \quad \Sigma_\Lambda[\phi] = S_\Lambda[\phi] - 2\hat{S}_\Lambda[\phi] \tag{68}$$

where $\hat{C}_\Lambda(x-y) \geq 0$. Now we define a Riemannian metric on the tangent space to the space of probability functionals which we will later explain is the infinitesimal version of a functional W_2 metric. We let

$$\langle \delta P_1[\phi], \delta P_2[\phi] \rangle_{W_2} = \frac{1}{2} \int [d\phi] P[\phi] \int d^x d^y C_\Lambda(x-y) \frac{\delta P_1[\phi]}{\delta \phi(x)} \frac{\delta P_2[\phi]}{\delta \phi(y)} \tag{69}$$

where we define Φ_i for $i = 1, 2$ via the functional differential equations

$$\delta P_i[\phi] - \frac{1}{2} \int d^x d^y C_\Lambda(x-y) \frac{\delta}{\delta \phi(x)} \left(P[\phi] \frac{\delta \Phi_i[\phi]}{\delta \phi(y)} \right) = 0. \tag{70}$$

Analogous to the finite-dimensional heat flow setting, the Φ_i's are only specified by the above equation up to additive functions not depending on ϕ. Note that since $C_\Lambda(x-y) \geq 0$, the norm induced by the metric is automatically greater than or equal to zero. Similar to Otto's calculation we can perform an integration by parts in (3.1) to obtain the more compact expressions

$$\langle \delta P_1[\phi], \delta P_2[\phi] \rangle_{W_2} = -\int [d\phi] \delta P_1[\phi] \Phi_1[\phi] \delta P_2[\phi]. \tag{71}$$

Co-opting the results of Otto and generalizing them appropriately to our setting, we have that our metric is the infinitesimal form of the distance

$$W_2(P_1, P_2) := \left(\inf_{\Pi \in \Gamma(P_1, P_2)} 2 \int [d\phi_1][d\phi_2] \Pi[\phi_1, \phi_2] \int d^x d^y \hat{C}_\Lambda^{-1}(x, y) (\phi_1(x) - \phi_2(x))(\phi_1(y) - \phi_2(y)) \right)^{1/2} \tag{72}$$

where $\Gamma(P_1, P_2)$ is the space of probability functionals $\Pi[\phi_1, \phi_2]$ such that $\int [d\phi_1] \Pi[\phi_1, \phi_2] = P[\phi_1]$ and $\int [d\phi_2] \Pi[\phi_1, \phi_2] = P[\phi_2]$. Above $\hat{C}_\Lambda^{-1}(x, y)$ is the inverse of the kernel $\hat{C}_\Lambda(x, y)$ in the sense that $\int dz \, \hat{C}_\Lambda^{-1}(x, z) \hat{C}_\Lambda(z, y) = \delta^d(x-y)$; since in our setting $\hat{C}_\Lambda(x, y) = \hat{C}_\Lambda(x-y)$, in momentum space the kernel $\hat{C}_\Lambda(p^2)$ has as its inverse $\hat{C}_\Lambda^{-1}(p^2) = 1/\hat{C}_\Lambda(p^2)$. The distance $W_2(P_1, P_2)$ represents the minimum cost of 'transporting' P_1 into P_2 (or vice-versa) where the cost is given by an L^2 penalty on rearranging field degrees of freedom away from the spatial scale $\ell \sim 1/\Lambda$.

We are now almost ready to state our main result, and then subsequently derive it. Define the probability functional

$$Q_\Lambda[\phi] := \frac{e^{-2\hat{S}_\Lambda[\phi]}}{Z_{Q,\Lambda}} \tag{73}$$

where $Z_{Q,\Lambda} = \int [d\phi] e^{-2\hat{S}_\Lambda[\phi]}$, and let the functional relative entropy be

$$\mathcal{S}[P[\phi] \| Q[\phi]] := \int [d\phi] P[\phi] \log \left(\frac{P[\phi]}{Q[\phi]} \right). \tag{74}$$

Then we have the remarkable formula

$$-\Lambda \frac{d}{d\Lambda} P_\Lambda[\phi] = -\nabla_{W_2} \mathcal{S}(P_\Lambda[\phi] \| Q_\Lambda[\phi]). \tag{75}$$

10 Non-Perturbative RG Monotone

In the context of renormalization group (RG) theory, a non-perturbative RG monotone is a quantity that consistently decreases (or increases) along the RG flow, providing insights into the behavior of a system across different scales. This concept is particularly important for understanding strongly interacting systems where perturbative methods fail.

Consider a functional version of the relative entropy (also known as Kullback-Leibler divergence) defined as:

$$S(P \| Q) = \int [d\phi] \, P[\phi] \log \left(\frac{P[\phi]}{Q[\phi]} \right), \tag{76}$$

where $P[\phi]$ and $Q[\phi]$ are probability functionals. In our framework, the relative entropy $S(P_\Lambda[\phi]\|Q_\Lambda[\phi])$ serves as an RG monotone. This means that as the cutoff scale Λ is varied, the relative entropy either consistently decreases or increases, indicating a directionality to the RG flow.

The Polchinski equation can be reformulated in terms of the Wasserstein-2 metric, leading to the expression:

$$-\Lambda \frac{d}{d\Lambda} P_\Lambda[\phi] = -\nabla_{W_2} S(P_\Lambda[\phi]\|Q_\Lambda[\phi]). \tag{77}$$

Here, ∇_{W_2} denotes the gradient with respect to the Wasserstein-2 metric. This formulation reveals that the coarse-graining process in the RG flow generates a decrease in the relative entropy, confirming that $S(P_\Lambda[\phi]\|Q_\Lambda[\phi])$ acts as an RG monotone.

The significance of having a non-perturbative RG monotone lies in its ability to provide a consistent measure of how the effective theory evolves under the RG flow, independent of the specific details of the perturbative expansions. This is particularly useful for analyzing systems with strong interactions, where traditional perturbative approaches are inadequate. The monotonic behavior of such a quantity can help identify fixed points and understand phase transitions in the theory.

In summary, a non-perturbative RG monotone like the relative entropy $S(P_\Lambda[\phi]\|Q_\Lambda[\phi])$ offers a powerful tool for studying the evolution of quantum field theories and statistical systems under RG transformations, providing a deeper understanding of their large-scale behavior.

11 Handling Divergences in Differential Entropy

In probabilistic and information theory, differential entropy is a measure of the uncertainty associated with a continuous random variable. Unlike discrete entropy, differential entropy can become infinite, leading to significant challenges. Various techniques are employed to manage these divergences and make the entropy finite and well-defined.

11.0.1 Differential Entropy and Divergences

The differential entropy $H(X)$ of a continuous random variable X with probability density function $p(x)$ is given by:

$$H(X) = -\int_{-\infty}^{\infty} p(x) \log p(x)\, dx$$

For certain distributions, this integral can diverge. For example, the differential entropy of the normal distribution $N(\mu, \sigma^2)$ is:

$$H(X) = \frac{1}{2} \log(2\pi e \sigma^2)$$

If $\sigma \to 0$, the entropy diverges to $-\infty$, reflecting the increasing precision (or decreasing uncertainty) of X.

11.0.2 Regularization Techniques

To handle these divergences, various regularization techniques are employed:

Cutoff Regularization In cutoff regularization, we introduce a cutoff scale Λ to limit the range of integration. For instance, the integral for differential entropy can be modified to:

$$H_\Lambda(X) = -\int_{-\Lambda}^{\Lambda} p(x) \log p(x)\, dx$$

This approach ensures the integral remains finite, but the choice of Λ can affect the result.

Renormalization Renormalization involves redefining the entropy by subtracting the divergent part. For example, if the differential entropy has a divergence that scales as $\log \epsilon$, where ϵ is a small parameter, we can define the renormalized entropy as:

$$H_{\text{ren}}(X) = H(X) + \log \epsilon$$

This subtraction removes the infinite part, yielding a finite entropy value.

Dimensional Regularization Dimensional regularization extends the dimension of the space to $d = 1 - \epsilon$, where ϵ is a small parameter, and then analytically continues back to $d = 1$. The differential entropy in d dimensions is:

$$H_d(X) = -\int p(x) \log p(x) \, d^d x$$

We then take the limit as $\epsilon \to 0$ to obtain a finite result.

11.0.3 Example: Entropy of a Uniform Distribution

Consider a uniform distribution $U(a, b)$ with density:

$$p(x) = \frac{1}{b-a} \quad \text{for } a \leq x \leq b$$

The differential entropy is:

$$H(U) = \log(b - a)$$

If $b - a \to 0$, the entropy diverges. Applying cutoff regularization, we limit the range:

$$H_\Lambda(U) = \log(\min(\Lambda, b - a))$$

Alternatively, using renormalization, we redefine the entropy to remove the divergence:

$$H_{\text{ren}}(U) = \log(b - a) + \log \epsilon$$

12 Handling Divergences and Regularization Techniques in QFT

At certain points in quantum field theory, divergences can become infinite, posing significant challenges to the formulation of a consistent theory. Various regularization techniques have been developed to manage these divergences, ensuring that physical quantities remain finite and well-defined. In this subsection, we provide an overview of some of the most commonly used regularization methods.

12.0.1 Cutoff Regularization

Cutoff regularization involves introducing a cutoff scale Λ beyond which the contributions to the integrals are suppressed. This approach modifies the propagator by introducing a smooth cutoff function $R_\Lambda(p^2)$ such that:

$$G(p) \to \frac{G(p)}{1 + R_\Lambda(p^2)}$$

where $R_\Lambda(p^2)$ is designed to approach zero as $p^2 \to 0$ and to grow rapidly for $p^2 \gg \Lambda^2$.

12.0.2 Dimensional Regularization

Dimensional regularization extends the number of spacetime dimensions from d to $d = 4 - \epsilon$, where ϵ is a small parameter. This technique modifies the integrals by analytically continuing them to non-integer dimensions. The integral of a function $f(p)$ in dimensional regularization is expressed as:

$$\int \frac{d^d p}{(2\pi)^d} f(p) \to \mu^{4-d} \int \frac{d^d p}{(2\pi)^d} f(p)$$

where μ is the renormalization scale introduced to keep the dimensions consistent.

12.0.3 Pauli-Villars Regularization

Pauli-Villars regularization introduces additional, fictitious particles with large masses to cancel the divergences. For a propagator $G(p)$, the regularized form is given by:

$$G_{\text{reg}}(p) = G(p) - \sum_i c_i \frac{1}{p^2 - M_i^2}$$

where M_i are the large masses of the fictitious particles and c_i are chosen coefficients ensuring the cancellation of divergences.

12.0.4 Examples and Applications

To illustrate these techniques, consider the one-loop correction to the propagator in a scalar field theory. The divergent integral is:

$$\int \frac{d^4k}{(2\pi)^4} \frac{1}{k^2 + m^2}$$

Using cutoff regularization, this integral becomes:

$$\int_0^\Lambda \frac{d^4k}{(2\pi)^4} \frac{1}{k^2 + m^2}$$

In dimensional regularization, the same integral is expressed as:

$$\mu^{4-d} \int \frac{d^dk}{(2\pi)^d} \frac{1}{k^2 + m^2} = \frac{\mu^{4-d}}{(4\pi)^{d/2}} \Gamma\left(1 - \frac{d}{2}\right) (m^2)^{\frac{d}{2}-1}$$

Regularization techniques are crucial for ensuring that quantum field theories provide finite and physically meaningful predictions. By applying these methods, we can systematically handle divergences and advance our understanding of fundamental interactions.

13 Link Between Handling Divergences in QFT and Probabilistic Divergences

In both Quantum Field Theory (QFT) and probabilistic settings, divergences present significant challenges. Techniques developed to manage these divergences in QFT can offer insights into handling similar issues in probabilistic contexts, particularly through the lens of optimal transport and Wasserstein distances.

13.0.1 Divergences in QFT and Regularization

In QFT, divergences occur due to the infinite number of degrees of freedom at high energies. Regularization techniques such as cutoff regularization, dimensional regularization, and Pauli-Villars regularization are used to manage these infinities. For instance, the integral for a loop correction in QFT might be expressed as:

$$\int \frac{d^4k}{(2\pi)^4} \frac{1}{k^2 + m^2}$$

Using dimensional regularization, this integral becomes:

$$\mu^{4-d} \int \frac{d^dk}{(2\pi)^d} \frac{1}{k^2 + m^2} = \frac{\mu^{4-d}}{(4\pi)^{d/2}} \Gamma\left(1 - \frac{d}{2}\right) (m^2)^{\frac{d}{2}-1}$$

13.0.2 Divergences in Probabilistic Settings

In probabilistic contexts, differential entropy can become infinite for certain distributions. Regularization techniques can similarly be applied to manage these divergences. For example, the differential entropy $H(X)$ for a continuous random variable X with density $p(x)$ is:

$$H(X) = -\int_{-\infty}^{\infty} p(x) \log p(x) \, dx$$

For certain distributions, this integral can diverge, requiring techniques like cutoff regularization or renormalization.

13.0.3 Connecting QFT and Probabilistic Divergences via Wasserstein Distance

Recent advances suggest a profound connection between the renormalization group (RG) flow in QFT and optimal transport theory. The RG flow can be interpreted as a gradient flow with respect to the Wasserstein distance in the space of probability measures. The action $S_\Lambda[\phi]$ in QFT, which evolves with the scale Λ, can be seen as analogous to the evolution of a probability distribution under the Wasserstein gradient flow.

Consider the probability functional $P_\Lambda[\phi]$:

$$-\Lambda \frac{d}{d\Lambda} P_\Lambda[\phi] = \int d^d x \frac{\delta}{\delta \phi(x)} \left(\Psi_\Lambda[\phi, x] P_\Lambda[\phi] \right)$$

This equation is structurally similar to the Fokker-Planck equation, which describes the gradient flow of a probability distribution with respect to the Wasserstein distance W_2. The Wasserstein gradient flow for a functional $\mathcal{S}[P]$ is given by:

$$\frac{\partial P}{\partial t} = \nabla_{W_2} \mathcal{S}[P]$$

In both settings, the divergence is handled by evolving the distribution or the action in a controlled manner, ensuring that the quantities remain finite and well-defined.

13.0.4 Example: Renormalization and Wasserstein Gradient Flow

To illustrate the connection, consider the renormalized action in QFT:

$$S_\Lambda[\phi] = S[\phi] + \delta S_\Lambda[\phi]$$

The change in the action with respect to Λ can be related to the change in the differential entropy with respect to the Wasserstein distance:

$$\frac{\partial S_\Lambda[\phi]}{\partial \Lambda} \sim \nabla_{W_2} H[P_\Lambda[\phi]]$$

Here, the differential entropy $H[P_\Lambda[\phi]]$ plays a role analogous to the action in QFT, and its gradient with respect to the Wasserstein distance drives the evolution of the probability functional.

14 One-Loop Renormalization in Quantum Gravity

Quantum gravity aims to describe the gravitational force within the framework of quantum mechanics. One of the significant challenges in quantum gravity is dealing with the divergences that arise in the perturbative expansion of the gravitational field. This section discusses the one-loop renormalization process in quantum gravity, highlighting the key concepts, techniques, and mathematical expressions involved.

14.1 Basics of Quantum Gravity

In classical general relativity, the gravitational field is described by the metric tensor $g_{\mu\nu}$. The action for gravity is given by the Einstein-Hilbert action:

$$S_{\text{EH}} = \frac{1}{16\pi G} \int d^4x \sqrt{-g} \, (R - 2\Lambda)$$

where G is Newton's gravitational constant, R is the Ricci scalar, and Λ is the cosmological constant. In the path integral formulation of quantum gravity, we aim to quantize the fluctuations of the metric tensor around a fixed background.

14.2 Perturbative Expansion and One-Loop Diagrams

To study quantum effects, we expand the metric tensor $g_{\mu\nu}$ around a background metric $\bar{g}_{\mu\nu}$:

$$g_{\mu\nu} = \bar{g}_{\mu\nu} + \kappa h_{\mu\nu}$$

where $\kappa = \sqrt{32\pi G}$ and $h_{\mu\nu}$ represents the quantum fluctuations. The Einstein-Hilbert action is then expanded in powers of $h_{\mu\nu}$. The leading-order quantum corrections are captured by the one-loop diagrams.

The one-loop effective action $\Gamma^{(1)}$ is given by the determinant of the differential operator arising from the quadratic terms in the action:

$$\Gamma^{(1)} = \frac{i}{2} \log \det (\Delta)$$

where Δ is the differential operator obtained from the quadratic expansion of the action.

14.3 Renormalization of One-Loop Diagrams

The one-loop effective action contains divergences that need to be regularized and renormalized. We employ dimensional regularization, where the spacetime dimension d is extended to $d = 4 - \epsilon$. The one-loop effective action in dimensional regularization is:

$$\Gamma^{(1)} = \frac{i}{2} \int d^dx \sqrt{-g} \left(\frac{1}{\epsilon} a_2(x) + \text{finite terms} \right)$$

where $a_2(x)$ is the heat kernel coefficient, which depends on the background geometry and the differential operator Δ.

The divergent part of the effective action is absorbed into the renormalization of the coupling constants. For gravity, the renormalized action at one-loop is:

$$S_{\text{ren}} = \frac{1}{16\pi G(\mu)} \int d^4x \sqrt{-g} \, (R - 2\Lambda(\mu)) + \frac{\beta}{32\pi G} \int d^4x \sqrt{-g} R^2$$

where $G(\mu)$ and $\Lambda(\mu)$ are the renormalized gravitational constant and cosmological constant at the scale μ, and β is a coefficient determined by the one-loop corrections.

14.4 Beta Functions and Running Couplings

The running of the coupling constants with the energy scale μ is governed by the beta functions. The beta function for Newton's constant G is:

$$\beta_G = \mu \frac{dG}{d\mu} = \frac{c}{\epsilon} G^2$$

where c is a constant determined by the one-loop calculation. Integrating this equation gives the running of G with the scale μ:

$$G(\mu) = \frac{G_0}{1 - \frac{c}{\epsilon} G_0 \log \left(\frac{\mu}{\mu_0} \right)}$$

Similarly, the running of the cosmological constant Λ is given by its beta function:

$$\beta_\Lambda = \mu \frac{d\Lambda}{d\mu} = \frac{d}{\epsilon} \Lambda G$$

where d is another constant from the one-loop calculation.

One-loop renormalization in quantum gravity involves dealing with divergences that arise in the perturbative expansion of the metric tensor. Dimensional regularization and the renormalization of coupling constants are essential techniques for managing these divergences. The running of the gravitational constant and the cosmological constant with the energy scale provides insights into the behavior of gravity at different scales. These techniques and results are crucial for developing a consistent quantum theory of gravity and understanding the quantum effects on spacetime.

15 Non-Renormalizability of Two-Loop Quantum Gravity

Quantum gravity aims to provide a quantum mechanical description of the gravitational field. While one-loop quantum gravity can be renormalized using standard techniques, at the two-loop level, the theory exhibits non-renormalizable divergences. This section explores the reasons behind the non-renormalizability of two-loop quantum gravity and the implications for the development of a consistent quantum theory of gravity.

15.1 Perturbative Expansion and Two-Loop Diagrams

In perturbative quantum gravity, the metric tensor $g_{\mu\nu}$ is expanded around a background metric $\bar{g}_{\mu\nu}$:

$$g_{\mu\nu} = \bar{g}_{\mu\nu} + \kappa h_{\mu\nu}$$

where $\kappa = \sqrt{32\pi G}$ and $h_{\mu\nu}$ represents the quantum fluctuations. The Einstein-Hilbert action is expanded in powers of $h_{\mu\nu}$, and the quantum corrections are calculated order by order in perturbation theory.

The two-loop effective action includes contributions from diagrams with two closed loops of graviton propagators. These contributions are significantly more complex than the one-loop diagrams and introduce additional divergences.

15.2 Structure of Two-Loop Divergences

The divergences in the two-loop effective action arise from the integration over high-momentum modes. The general form of the two-loop divergence is:

$$\Gamma^{(2)}_{\text{div}} = \frac{i}{(4\pi)^2} \frac{1}{\epsilon^2} \int d^4x \sqrt{-g} \left(\alpha R_{\mu\nu} R^{\mu\nu} + \beta R^2 + \gamma R_{\mu\nu\rho\sigma} R^{\mu\nu\rho\sigma} \right)$$

where $R_{\mu\nu}$, R, and $R_{\mu\nu\rho\sigma}$ are the Ricci tensor, Ricci scalar, and Riemann tensor, respectively, and α, β, and γ are coefficients that depend on the details of the two-loop calculations.

15.3 Non-Renormalizability of Quantum Gravity

Non-renormalizability means that the infinities that arise at the two-loop level cannot be absorbed by a finite number of counterterms that correspond to terms in the original action. In the case of two-loop quantum gravity, new types of divergences appear that are not present in the Einstein-Hilbert action, such as:

$$\int d^4x \sqrt{-g} R_{\mu\nu\rho\sigma} R^{\mu\nu\rho\sigma}$$

These terms require new counterterms that are not part of the original theory, indicating that an infinite number of parameters would be needed to renormalize the theory completely.

15.3.1 Dimensional Analysis and Power Counting

Dimensional analysis and power counting provide insights into the non-renormalizability of quantum gravity. The coupling constant κ in gravity has dimensions of inverse mass, $[\kappa] = \text{mass}^{-1}$. This implies that higher-order loop diagrams introduce more severe divergences. Specifically, the superficial degree of divergence D for a diagram with L loops, V vertices, and I internal lines is given by:

$$D = 4L - 2I + 2V$$

For two-loop diagrams ($L = 2$), the degree of divergence is:

$$D = 8 - 2I + 2V$$

Since I and V increase with the complexity of the diagrams, the divergences at two loops are more severe and involve higher powers of the curvature tensors, which are not present in the original action.

15.4 Implications for Quantum Gravity

The non-renormalizability of two-loop quantum gravity suggests that a different approach is necessary to formulate a consistent quantum theory of gravity. Possible approaches include:

Effective Field Theory Quantum gravity can be treated as an effective field theory, valid at low energies. In this framework, the theory includes an infinite series of higher-dimensional operators suppressed by powers of the Planck scale:

$$S_{\text{eff}} = \int d^4x\sqrt{-g}\left(\frac{1}{16\pi G}R + c_1 R^2 + c_2 R_{\mu\nu}R^{\mu\nu} + \sum_i \frac{c_i}{M_{\text{Pl}}^{2i}} O_i\right)$$

where O_i are higher-dimensional operators and M_{Pl} is the Planck mass.

Alternative Theories Various alternative theories, such as string theory and loop quantum gravity, aim to provide a finite and consistent description of quantum gravity. These theories introduce new degrees of freedom and principles that can potentially resolve the issue of non-renormalizability.

The two-loop level non-renormalizability of quantum gravity indicates that the standard perturbative approach is insufficient to formulate a consistent quantum theory of gravity. While one-loop divergences can be managed, the two-loop divergences introduce new terms that require an infinite number of counterterms. Effective field theory and alternative approaches like string theory offer potential solutions, paving the way for further research in the quest for a complete theory of quantum gravity.

16 Triangle Relationship for Bregman Divergences Using Dual Connections

Bregman divergences are a family of measures that generalize many well-known divergences in information theory, such as the Kullback-Leibler divergence. They are closely related to convex functions and have important applications in optimization and information geometry. This section explores the triangle relationship that exists for Bregman divergences using dual connections in the context of information geometry.

16.1 Bregman Divergences

A Bregman divergence $D_\phi(p\|q)$ is defined for a strictly convex function ϕ as:

$$D_\phi(p\|q) = \phi(p) - \phi(q) - \langle \nabla\phi(q), p - q\rangle$$

where p and q are points in a convex set, and $\nabla\phi(q)$ denotes the gradient of ϕ at q. Bregman divergences capture the difference between the value of ϕ at p and the first-order Taylor expansion of ϕ around q.

16.2 Dual Connections in Information Geometry

In information geometry, a dual pair of affine connections, ∇ and ∇^*, is often considered. These connections are related to a potential function ϕ and its Legendre dual ϕ^*. The primal and dual connections ∇ and ∇^* are defined such that:

$$\nabla\nabla^*\phi = \phi$$

The connections ∇ and ∇^* correspond to the geometry induced by the potential function ϕ and its dual ϕ^*, respectively.

16.3 Triangle Relationship for Bregman Divergences

A significant property of Bregman divergences is the triangle relationship that can be understood using the dual connections in information geometry. Consider three points p, q, and r in a convex set. The Bregman divergence satisfies the following relationship:

$$D_\phi(p\|r) = D_\phi(p\|q) + D_\phi(q\|r) + \langle \nabla\phi(q) - \nabla\phi(r), p - q\rangle$$

This relationship can be interpreted geometrically using the primal and dual connections.

16.3.1 Geometric Interpretation Using Dual Connections

In the context of dual connections, the triangle relationship reflects the consistency of the Bregman divergence with the underlying geometry of the space. The Bregman divergence $D_\phi(p\|q)$ can be seen as the difference in the potential function ϕ evaluated at p and q, corrected by the linear term $\langle \nabla\phi(q), p-q\rangle$.

Using the dual connections, the Bregman divergence can be decomposed as:

$$D_\phi(p\|r) = D_\phi(p\|q) + D_\phi(q\|r) + \langle \nabla^*\phi(p) - \nabla^*\phi(q), q - r\rangle$$

Here, $\nabla^*\phi$ represents the gradient with respect to the dual connection. The term $\langle \nabla^*\phi(p) - \nabla^*\phi(q), q - r\rangle$ captures the adjustment needed to account for the change in geometry from q to r.

16.3.2 Example: Kullback-Leibler Divergence

The Kullback-Leibler (KL) divergence is a special case of a Bregman divergence with the convex function $\phi(p) = \sum_i p_i \log p_i$. For probability distributions p, q, and r, the triangle relationship for the KL divergence can be written as:

$$D_{\mathrm{KL}}(p\|r) = D_{\mathrm{KL}}(p\|q) + D_{\mathrm{KL}}(q\|r) + \sum_i (p_i - q_i) \log \frac{q_i}{r_i}$$

This demonstrates how the divergence between p and r can be decomposed into the divergence between p and q, the divergence between q and r, and a correction term involving the logarithmic ratios.

The triangle relationship for Bregman divergences provides valuable insights into the structure of these divergences and their geometric interpretation using dual connections. This relationship highlights the consistency of Bregman divergences with the underlying geometry of the space and illustrates the deep connections between convex analysis, optimization, and information geometry. Understanding this relationship is crucial for applications in various fields, including machine learning, signal processing, and statistical inference.

17 Path Integrals as Bregman Divergences

Path integrals are a fundamental tool in quantum mechanics and quantum field theory, providing a means to calculate quantum amplitudes by summing over all possible paths. In this section, we explore the interpretation of path integrals as Bregman divergences and discuss how the Pythagorean relationship for Bregman divergences can offer a new perturbative approach, potentially more effective than Feynman diagrams, particularly for handling quantum gravity.

17.1 Path Integrals and Bregman Divergences

The path integral formulation, introduced by Richard Feynman, expresses the quantum amplitude for a system transitioning from an initial state to a final state as an integral over all possible paths:

$$\mathcal{Z} = \int \mathcal{D}[\phi] \, e^{iS[\phi]}$$

where $\mathcal{D}[\phi]$ represents the measure over all paths (field configurations) ϕ, and $S[\phi]$ is the action functional. This formulation can be related to Bregman divergences by considering the action $S[\phi]$ as a convex functional.

A Bregman divergence $D_\phi(p\|q)$ for a strictly convex function ϕ is given by:

$$D_\phi(p\|q) = \phi(p) - \phi(q) - \langle \nabla\phi(q), p - q \rangle$$

In the context of path integrals, we can interpret the action $S[\phi]$ as the convex function, and the difference in action between different paths can be seen as a Bregman divergence.

17.2 Pythagorean Relationship for Bregman Divergences

The Pythagorean relationship for Bregman divergences states that for three points p, q, and r in a convex set, the following holds:

$$D_\phi(p\|r) = D_\phi(p\|q) + D_\phi(q\|r) + \langle \nabla\phi(q) - \nabla\phi(r), p - q \rangle$$

This relationship is useful for decomposing the divergence between two points into intermediate steps, providing a geometric interpretation of the divergence.

17.3 Developing a New Perturbative Approach

The traditional perturbative approach in quantum field theory involves expanding the path integral in terms of Feynman diagrams, which correspond to different interaction processes. However, this method can become cumbersome, especially in theories with strong interactions or in quantum gravity, where the non-renormalizability at higher loops poses significant challenges.

By leveraging the Pythagorean relationship for Bregman divergences, we can develop a new perturbative approach that decomposes the path integral in a geometrically consistent manner. This approach involves the following steps:

1. **Decompose the Path Integral:** Using the Pythagorean relationship, decompose the action $S[\phi]$ into intermediate steps, each representing a Bregman divergence.

2. **Hierarchical Expansion:** Construct a hierarchical expansion of the path integral, where each level of the hierarchy corresponds to a different scale of interaction. This expansion can be more effective in capturing the contributions from different energy scales.

3. **Geometric Interpretation:** Interpret the terms in the expansion geometrically, using the properties of Bregman divergences and the underlying geometry of the space of field configurations.

4. **Improved Convergence:** Use the geometric structure to improve the convergence properties of the perturbative expansion, reducing the need for regularization and renormalization.

17.4 Application to Quantum Gravity

In quantum gravity, where the perturbative expansion using Feynman diagrams faces significant difficulties, the Bregman divergence approach offers a promising alternative. The hierarchical expansion and geometric interpretation can help manage the non-renormalizable divergences that arise at higher loops.

17.4.1 Example: Gravitational Action

Consider the gravitational action in the Einstein-Hilbert form:

$$S[g_{\mu\nu}] = \frac{1}{16\pi G} \int d^4x \sqrt{-g}\,(R - 2\Lambda)$$

Using the Bregman divergence approach, we decompose the action into intermediate steps, each capturing different curvature contributions. The Pythagorean relationship helps ensure that the divergences are handled consistently at each step, potentially leading to a more manageable perturbative series.

17.4.2 Advantages and Challenges

The primary advantage of this approach is its potential to provide a more efficient and convergent perturbative expansion, especially in non-renormalizable theories like quantum gravity. However, challenges remain in formulating the precise rules for the decomposition and ensuring that the resulting series captures all relevant physical effects.

18 Fourth-Derivative Gravity

The next model of interest in the context of quantum gravity is the fourth-derivative action. Unlike Einstein's classical gravitational theory, which is strongly motivated by semiclassical theory, the fourth-derivative model incorporates additional terms that make it potentially renormalizable. This section discusses the formulation, implications, and alternative forms of the fourth-derivative gravity model.

18.1 Classical Action and Motivation

The classical action for fourth-derivative gravity includes higher-order curvature terms, extending the Einstein-Hilbert action. The action is given by:

$$S = -\int d^4x \sqrt{-g} \left[\frac{1}{2\lambda} C^2 - \frac{1}{\rho} E + \frac{1}{\xi} R^2 + \tau \Box R + \frac{1}{\kappa^2}(R - 2\lambda) \right]$$

where $C^2 = C_{\mu\nu\alpha\beta} C^{\mu\nu\alpha\beta}$ is the square of the Weyl tensor, E is the integrand of the Gauss-Bonnet topological term, and λ, ρ, ξ, and τ are independent parameters. The sign of the coupling λ is chosen to ensure the positivity of the energy for the free tensor modes at low energy.

18.2 Parameter Definitions and Coupling Ratios

The coefficients λ, ρ, ξ, and τ have specific roles in the fourth-derivative gravity model. They can be related to other parameters as:

$$\theta = \frac{\lambda}{\rho} \quad \text{and} \quad \omega = -\frac{3\lambda}{\xi}$$

These ratios help to understand the interaction dynamics, particularly the exchange of massive scalar modes and tensor modes.

18.3 Observations on Classical Solutions

The inclusion of fourth-derivative terms implies that the Einstein-Hilbert term becomes the second derivative of the metric, multiplied by a factor related to the Planck mass M_P:

$$\frac{1}{\kappa^2} = \frac{M_P^2}{16\pi}$$

The fourth-derivative terms, however, do not inherently have dimensional parameters as their influence is regulated by the number of derivatives of the metric. In classical solutions, where $\partial_\alpha \partial_\beta g \ll M_P$, the effect of fourth-derivative terms is small. This is known as "Planck suppression," where these terms are negligible in typical laboratory experiments or astrophysical phenomena, given that λ is of order unity.

Nevertheless, the presence of higher derivatives can alter the stability properties of solutions, and it remains unclear if Planck suppression applies to perturbations.

18.4 Alternative Action Formulation

An alternative form of the fourth-derivative action emphasizes the contributions of different curvature invariants explicitly:

$$S=\int d^4x\sqrt{-g}\left(xR^2_{\mu\nu\alpha\beta}+yR^2_{\mu\nu}+zR^2+\tau\Box R-\frac{1}{\kappa^2}(R+2\lambda)\right)$$

In this formulation, the parameters x, y, z, and τ correspond to the coefficients of the respective curvature invariants, providing a more detailed view of the contributions from each term.

18.5 Implications and Challenges

The fourth-derivative gravity model offers a framework that potentially addresses some of the renormalizability issues in quantum gravity. However, the inclusion of higher-order terms introduces complexity in the analysis of classical solutions and their stability. Further studies are necessary to determine the viability of these terms in providing a consistent and predictive theory of quantum gravity.

Fourth-derivative gravity extends the Einstein-Hilbert action by incorporating higher-order curvature terms, with the aim of addressing the renormalizability issues in quantum gravity. The classical action, parameter definitions, and implications of these terms provide a foundation for further exploration. The alternative action formulation offers additional insights into the contributions of different curvature invariants. While the model presents promising avenues, significant challenges remain in understanding the full implications of higher-derivative terms and their impact on the stability of solutions.

19 Conclusion

Interpreting path integrals as Bregman divergences and leveraging the Pythagorean relationship offers a novel perturbative approach in quantum field theory and quantum gravity. This method provides a geometric framework that can improve the handling of divergences and enhance the convergence properties of the perturbative series. Further research is needed to fully develop and test this approach, but it holds promise for advancing our understanding of quantum gravity and other complex quantum systems.

References

[1] K. G. Wilson and J. Kogut, *The renormalization group and the ϵ expansion*, Physics Reports, 12 (1974), no. 2, 75–199.

[2] J. Polchinski, *Renormalization and effective Lagrangians*, Nuclear Physics B, 231 (1984), no. 2, 269–295.

[3] C. Bagnuls and C. Bervillier, *Exact renormalization group equations. An Introductory review*, Phys. Rept., 348 (2001) 91, [hep-th/0002034].

[4] J. Berges, N. Tetradis, and C. Wetterich, *Nonperturbative renormalization flow in quantum field theory and statistical physics*, Phys. Rept., 363 (2002) 223–386, [hep-ph/0005122].

[5] O. J. Rosten, *Fundamentals of the exact renormalization group*, Physics Reports, 511 (2012), no. 4, 177–272, [1003.1366].

[6] C. Villani, *Optimal transport: old and new*, vol. 338. Springer, 2009.

[7] S. Arnone, A. Gatti, T. R. Morris, and O. J. Rosten, *Exact scheme independence at two loops*, Phys. Rev. D, 69 (2004), 065009, [hep-th/0309242].

[8] T. R. Morris and O. J. Rosten, *A Manifestly gauge invariant, continuum calculation of the SU(N) Yang-Mills two-loop beta function*, Phys. Rev. D, 73 (2006), 065003, [hep-th/0508026].

[9] T. R. Morris and O. J. Rosten, *Manifestly gauge invariant QCD*, J. Phys. A, 39 (2006), 11657–11681, [hep-th/0606189].

[10] J. M. Pawlowski, *Aspects of the functional renormalisation group*, Annals Phys., 322 (2007), 2831–2915, [hep-th/0512261].

[11] M. Salmhofer and C. Honerkamp, *Fermionic renormalization group flows: Technique and theory*, Prog. Theor. Phys., 105 (2001), 1–35.

[12] M. Reuter and C. Wetterich, *Effective average action for gauge theories and exact evolution equations*, Nucl. Phys. B, 417 (1994), 181–214.

[13] H. Gies, *Introduction to the functional RG and applications to gauge theories*, Lect. Notes Phys., 852 (2012), 287–348, [hep-ph/0611146].

[14] F. Wegner, *Some invariance properties of the renormalization group*, Journal of Physics C: Solid State Physics, 7 (1974), no. 12, 2098.

[15] T. R. Morris, *A Gauge invariant exact renormalization group. 1.*, Nucl. Phys. B, 573 (2000), 97–126, [hep-th/9910058].

[16] J. I. Latorre and T. R. Morris, *Exact scheme independence*, Journal of High Energy Physics, 2000 (2000), no. 11, 004, [hep-th/0008123].

[17] S. Arnone, A. Gatti, and T. R. Morris, *A Proposal for a manifestly gauge invariant and universal calculus in Yang-Mills theory*, Phys. Rev. D, 67 (2003), 085003, [hep-th/0209162].

[18] S. Arnone, T. R. Morris, and O. J. Rosten, *A Generalised manifestly gauge invariant exact renormalisation group for SU(N) Yang-Mills*, Eur. Phys. J. C, 50 (2007), 467–504, [hep-th/0507154].

[19] L. P. Kadanoff, *Scaling laws for Ising models near Tc*, Physics Physique Fizika, 2 (1966), no. 6, 263.

[20] L. P. Kadanoff, *Statistical physics: statics, dynamics and renormalization*, World Scientific, 2000.

[21] F. Otto, *The geometry of dissipative evolution equations: the porous medium equation*, Communications in Partial Differential Equations, 26 (2001), no. 1-2, 101–174, [https://doi.org/10.1081/PDE-100002243].

[22] Y. Brenier, *Polar factorization and monotone rearrangement of vector-valued functions*, Communications on Pure and Applied Mathematics, 44 (1991), no. 4, 375–417, [https://onlinelibrary.wiley.com/doi/pdf/10.1002/cpa.3160440402].

[23] L. A. Caffarelli, *The regularity of mappings with a convex potential*, Journal of the American Mathematical Society, 5 (1992), no. 1, 99–104.

[24] F. Otto and M. Westdickenberg, *Eulerian calculus for the contraction in the Wasserstein distance*, SIAM Journal on Mathematical Analysis, 37 (2005), no. 4, 1227–1255.

[25] J.-D. Benamou and Y. Brenier, *A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem*, Numerische Mathematik, 84 (2000), 375–393.

[26] N. Guillen and R. McCann, *Five lectures on optimal transportation: Geometry, regularity and applications*, 2010.

[27] L. Gross, *Logarithmic Sobolev Inequalities*, American Journal of Mathematics, 97 (1975), no. 4, 1061–1083.

[28] F. Otto and C. Villani, *Generalization of an Inequality by Talagrand and Links with the Logarithmic Sobolev Inequality*, Journal of Functional Analysis, 173 (2000), no. 2, 361–400.

[29] R. Bauerschmidt and T. Bodineau, *Log-Sobolev Inequality for the Continuum Sine-Gordon Model*, Communications on Pure and Applied Mathematics, 74 (2021), no. 10, 2064–2113, [https://onlinelibrary.wiley.com/doi/pdf/10.1002/cpa.21926].

[30] G. Perelman, *The entropy formula for the Ricci flow and its geometric applications*, 2002.

[31] R. J. McCann and P. M. Topping, *Ricci flow, entropy and optimal transportation*, American Journal of Mathematics, 132 (2010), no. 3, 711–730.

[32] T. M. Cover, *Elements of information theory*, John Wiley & Sons, 1999.

[33] E. Witten, *A Mini-Introduction To Information Theory*, Riv. Nuovo Cim., 43 (2020), no. 4, 187–227, [1805.11965].

[34] C. Bény and T. J. Osborne, *Information-geometric approach to the renormalization group*, Phys. Rev. A, 92 (2015), no. 2, 022330, [1206.7004].

[35] C. Bény and T. J. Osborne, *The renormalization group via statistical inference*, New J. Phys., 17 (2015), no. 8, 083005, [1402.4949].

[36] V. Balasubramanian, J. J. Heckman, and A. Maloney, *Relative Entropy and Proximity of Quantum Field Theories*, JHEP, 05 (2015), 104, [1410.6809].

[37] N. Lashkari, *Entanglement at a scale and renormalization monotones*, Journal of High Energy Physics, 2019 (2019), no. 1, 1–15, [1704.05077].

[38] K. Furuya, N. Lashkari, and S. Ouseph, *Real-space RG, error correction and Petz map*, JHEP, 01 (2022), 170, [2012.14001].

[39] R. Fowler and J. J. Heckman, *Misanthropic Entropy and Renormalization as a Communication Channel*, 2021, [2108.02772].

[40] A. Koenigstein, M. J. Steil, N. Wink, E. Grossi, and J. Braun, *Numerical fluid dynamics for FRG flow equations: Zero-dimensional QFTs as numerical test cases - Part II: Entropy production and irreversibility of RG flows*, 2021, [2108.10085].

[41] J. Erdmenger, K. T. Grosvenor, and R. Jefferson, *Towards quantifying information flows: relative entropy in deep neural networks and the renormalization group*, 2021, [2107.06898].

[42] G. Zumbach, *Almost second order phase transitions*, Phys. Rev. Lett., 71 (1993), 2421–2424.

[43] G. Zumbach, *The Local potential approximation of the renormalization group and its applications*, Phys. Lett. A, 190 (1994), 225–230.

[44] G. Zumbach, *The Renormalization group in the local potential approximation and its applications to the O(n) model*, Nucl. Phys. B, 413 (1994), 754–770.

[45] J. Generowicz, C. Harvey-Fros, and T. R. Morris, *C function representation of the local potential approximation*, Phys. Lett. B, 407 (1997), 27–32, [hep-th/9705088].

[46] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows: in metric spaces and in the space of probability measures*, Springer Science & Business Media, 2005.

[47] C. Villani, *Topics in optimal transportation*, American Mathematical Society, Providence, RI, 2003.

[48] R. Jordan, D. Kinderlehrer, and F. Otto, *The variational formulation of the Fokker-Planck equation*, SIAM Journal on Mathematical Analysis, 29 (1998), 1–17.

[49] M. Cuturi, *Sinkhorn distances: Lightspeed computation of optimal transport*, Advances in Neural Information Processing Systems, 26 (2013), 2292–2300.

[50] G. Peyré, *Entropic Approximation of Wasserstein Gradient Flows*, SIAM Journal on Imaging Sciences, 8 (2015), no. 4, 2323–2351.

[51] J.-D. Benamou, G. Carlier, Q. Mérigot, and É. Oudet, *Discretization of functionals involving the Monge-Ampère operator*, Numerische Mathematik, 134 (2015), 611–636.

[52] G. Carlier, V. Duval, G. Peyré, and B. Schmitzer, *Convergence of entropic schemes for optimal transport and gradient flows*, SIAM Journal on Mathematical Analysis, 49 (2017), no. 2, 1385–1418.

[53] W. Li, J. Lu, and L. Wang, *Fisher information regularization schemes for Wasserstein gradient flows*, Journal of Computational Physics, 416 (2020), 109449.

[54] J. A. Carrillo, K. Craig, L. Wang, and C. Wei, *Primal Dual Methods for Wasserstein Gradient Flows*, Foundations of Computational Mathematics, 2021.

[55] M. Jacobs, W. Lee, and F. Léger, *The back-and-forth method for Wasserstein gradient flows*, ESAIM: Control, Optimisation and Calculus of Variations, 27 (2021), 28.

[56] P. Mokrov, A. Korotin, L. Li, A. Genevay, J. Solomon, and E. Burnaev, *Large-Scale Wasserstein Gradient Flows*, 2021.

[57] C. Bonet, N. Courty, F. Septier, and L. Drumetz, *Sliced-Wasserstein Gradient Flows*, 2021.

[58] D. Alvarez-Melis, Y. Schiff, and Y. Mroueh, *Optimizing functionals on the space of probabilities with input convex neural networks*, 2021.

[59] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016, `http://www.deeplearningbook.org`.

[60] A. Krizhevsky, I. Sutskever, and G. E. Hinton, *Imagenet classification with deep convolutional neural networks*, Advances in Neural Information Processing Systems, 25 (2012), 1097–1105.

[61] K. He, X. Zhang, S. Ren, and J. Sun, *Deep residual learning for image recognition*, 2015.

[62] M. Karami, D. Schuurmans, J. Sohl-Dickstein, L. Dinh, and D. Duckworth, *Invertible convolutional flow*, Advances in Neural Information Processing Systems, 32 (2019).

[63] C. Bény, *Deep learning and the renormalization group*, 2013.

[64] P. Mehta and D. J. Schwab, *An exact mapping between the variational renormalization group and deep learning*, 2014.

[65] H. W. Lin, M. Tegmark, and D. Rolnick, *Why does deep and cheap learning work so well?*, 2016, [1608.08225].

[66] M. Koch-Janusz and Z. Ringel, *Mutual information, neural networks and the renormalization group*, 2017, [1704.06279].

[67] S. S. Funai and D. Giataganas, *Thermodynamics and feature extraction by machine learning*, 2018, [1810.08177].

[68] S. Iso, S. Shiba, and S. Yokoo, *Scale-invariant feature extraction of neural network and renormalization group flow*, 2018, [1801.07172].

[69] E. de Mello Koch, R. de Mello Koch, and L. Cheng, *Is deep learning a renormalization group flow?*, 2019, [1906.05214].

[70] S. S. Schoenholz, J. Gilmer, S. Ganguli, and J. Sohl-Dickstein, *Deep information propagation*, 2016.

[71] L. Xiao, Y. Bahri, J. Sohl-Dickstein, S. S. Schoenholz, and J. Pennington, *Dynamical isometry and a mean field theory of CNNs: How to train 10,000-layer vanilla convolutional neural networks*, 2018.

[72] J. Ho, A. Jain, and P. Abbeel, *Denoising diffusion probabilistic models*, Advances in Neural Information Processing Systems, 33 (2020), 6840–6851.

[73] J. Haegeman, T. J. Osborne, H. Verschelde, and F. Verstraete, *Entanglement renormalization for quantum fields in real space*, Physical Review Letters, 110 (2013), no. 10, 100402, [1102.5524].

[74] J. Cotler, M. R. M. Mozaffar, A. Mollabashi, and A. Naseh, *Renormalization group circuits for weakly interacting continuum field theories*, Fortschritte der Physik, 67 (2019), no. 10, 1900038, [1806.02835].

[75] J. S. Cotler, M. R. M. Mozaffar, A. Mollabashi, and A. Naseh, *Entanglement renormalization for weakly interacting fields*, Physical Review D, 99 (2019), no. 8, 085005, [1806.02835].

[76] Q. Hu, A. Franco-Rubio, and G. Vidal, *Continuous tensor network renormalization for quantum fields*, 2018, [1809.05176].

[77] J. R. Fliss, R. G. Leigh, and O. Parrikar, *Unitary Networks from the Exact Renormalization of Wave Functionals*, Physical Review D, 95 (2017), no. 12, 126001, [1609.03493].

[78] E. A. Carlen and J. Maas, *An analog of the 2-Wasserstein metric in non-commutative probability under which the fermionic Fokker-Planck equation is gradient flow for the entropy*, Communications in Mathematical Physics, 331 (2014), no. 3, 887–926.

[79] E. A. Carlen and J. Maas, *Gradient flow and entropy inequalities for quantum Markov semigroups with detailed balance*, Journal of Functional Analysis, 273 (2017), no. 5, 1810–1869.