Unifying Divergences Across Domains: From Physics to Learning and Finance

Olivier Croissant

January 16, 2025

Abstract

This paper explores the profound connections between the laws of nature, learning in artificial intelligence, and the dynamic hedging of financial derivatives. By leveraging the mathematical framework of divergences (or relative entropy), we unify these domains under a common perspective. We illustrate the power of divergences in renormalization group flows, high-dimensional generative learning tasks, and hedging strategies for derivative portfolios under realistic constraints.

Contents

1 Introduction

2 The Power of Divergences in the Laws of Nature

At the heart of Cotler and Rezchikov's work [1] lies a reformulation of Polchinski's equation as a gradient flow in the space of probability measures:

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

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

This equation demonstrates that the evolution of a probability distribution $P_\Lambda[\phi]$ under the renormalization group flow is governed by a decrease in relative entropy. Importantly, it reveals that quantum field theory achieves its extraordinary precision by iteratively refining the order of perturbation theory used to construct $P_\Lambda[\phi]$.

Furthermore, divergences possess a rich mathematical structure, exemplified by a **Pythagorean theorem** that allows for their decomposition into orthogonal components [4]. This structure generalizes into a "Taylor-like" expansion, mirroring perturbative expansions in quantum field theory. This reformulation provides a powerful framework for understanding complex dynamic systems.

3 Divergences in High-Dimensional Learning

In generative tasks like image and text generation, divergences play a central role. Diffusion models, for example, use stochastic differential equations (SDEs) to progressively add random noise to data until the original information is erased. The reverse process reconstructs meaningful outputs by minimizing divergences between intermediate and target distributions.

The forward and reverse processes are governed by coupled SDEs:

$$\begin{aligned} dx_t &= f(x_t, t)\, dt + g(t)\, dW_t, && \text{(Forward process)} \\ dx_t &= \tilde{f}(x_t, t)\, dt + \tilde{g}(t)\, dW_t, && \text{(Reverse process)}, \end{aligned}$$

where f and $\tilde{f}$ encode the dynamics of noise addition and removal, respectively. A comprehensive overview of time reversal in stochastic processes can be found in [3]. Additionally, an exploration of image generation using diffusion models is available in [5].

This framework highlights an essential insight: the time variable in these equations is analogous to the precision parameter in renormalization group flows. Both represent a progression from coarse to fine resolution, unifying the role of divergences in learning and physics.

4 Divergences in Dynamic Hedging of Derivatives

Dynamic hedging under transaction costs and regulatory constraints can be reinterpreted using divergences. The hedging problem can be formulated as:

$$\text{Minimize: } S(\text{Hedging Cost}\|\text{Portfolio Payoff}),$$

where $S(\cdot\|\cdot)$ is a divergence measure, such as the Kullback-Leibler divergence.

Using the "Taylor-like" approach from Chapter 1, we can decompose this divergence into perturbative terms, iteratively refining the hedge strategy. Similarly, the iterative learning framework from Chapter 2 can be applied to reconstruct the optimal hedge by training a neural network to minimize divergences.

This methodology aligns with the approach described in [2]. By combining perturbative expansions with machine learning, we can account for transaction costs, stochastic dynamics, and regulatory constraints in a unified framework.

5 Conclusion

A Taylor-Like Decomposition of Divergences

This appendix details the "Taylor-like" expansion for divergences, demonstrating how they can be systematically decomposed into perturbative terms. Starting from the general divergence framework $S(P\|Q)$, we illustrate how higher-order corrections provide finer approximations to complex distributions. Examples from information geometry are provided to support this approach.

Divergences, such as the Kullback-Leibler (KL) divergence or more general Bregman divergences, exhibit a rich mathematical structure that allows for systematic decomposition. One fundamental property underpinning this decomposition is the **Pythagorean-like theorem**, which applies to certain classes of divergences in information geometry.

Let P and Q be probability distributions, and let $\mathcal{M}$ represent a family of probability distributions (e.g., an exponential family). Suppose P^* is the projection of P onto $\mathcal{M}$ under a divergence measure $S(P\|Q)$. Then, the Pythagorean-like theorem states:

$$S(P\|Q) = S(P\|P^*) + S(P^*\|Q),$$

where:

- $S(P\|Q)$: The total divergence between P and Q.
- $S(P\|P^*)$: The divergence between P and its projection P^*.
- $S(P^*\|Q)$: The divergence between the projection P^* and Q.

This decomposition reflects the orthogonality of P to $\mathcal{M}$ with respect to the divergence. It is a powerful result that enables the iterative refinement of approximations, as seen in applications such as optimization in information geometry and perturbative expansions.

A.1 Machine Learning and Information Geometry

- Some machine learning algorithms can be viewed abstractly as the optimization problem of finding a parametric probability distribution that minimizes the mismatch with the data.
- Divergence (or contrast) functions are asymmetric analogues of distance functions that provide a unified way of measuring information, energy, and entropy.
- Divergence functions induce metrics and connections on the information manifold, and these differential geometric structures yield more natural optimization algorithms in the sense of optimization on manifolds than working in the parametric representation.

A.2 Divergence Between Two Points

Let us consider two points P and Q in a manifold M, whose coordinates are $\boldsymbol{\xi}_P$ and $\boldsymbol{\xi}_Q$, respectively. A divergence $D[P:Q]$ is a function of $\boldsymbol{\xi}_P, \boldsymbol{\xi}_Q$ which satisfies certain criteria, and often we will write, $D[P:Q]$ as $D[\boldsymbol{\xi}_P : \boldsymbol{\xi}_Q]$.

Definition: $D[P:Q]$ is a divergence when it satisfies the following:

1. $D[P : Q] \geq 0$.
2. $D[P : Q] = 0$ if and only if $P = Q$.
3. When P, Q are sufficiently close, by denoting their coordinates by $\boldsymbol{\xi}_P, \boldsymbol{\xi}_Q = \boldsymbol{\xi}_P + d\boldsymbol{\xi}$, the Taylor expansion of D is given by:

$$D[\boldsymbol{\xi}_P : \boldsymbol{\xi}_P + d\boldsymbol{\xi}] = \frac{1}{2} \sum_{ij} g_{ij}(\boldsymbol{\xi}_P) d\xi^i d\xi^j + \mathcal{O}(|d\boldsymbol{\xi}|^2), \tag{1}$$

where $G = (g_{ij})$ is a positive-definite matrix depending on $\boldsymbol{\xi}_P$.

A.3 Examples of Divergence

K-L divergence for positive measures: The Kullback-Leibler (K-L) divergence for two positive measures m_1 and m_2 is given by

$$D_{KL}[m_1 : m_2] = \sum_i m_{1i} \log \frac{m_{1i}}{m_{2i}} - \sum_i m_{1i} + \sum_i m_{2i}, \tag{2}$$

where when the total mass of the two measures m_1, m_2 equals 1, they are probability distributions, and this definition reduces to the definition of K-L for probability distributions.

Divergence for positive-definite matrices: The divergence for positive-definite matrices P and Q is

$$D[P : Q] = \mathrm{tr}(P \log P - P \log Q - P + Q), \tag{3}$$

which is related to the von Neumann entropy of quantum mechanics. Furthermore, the divergence of two Gaussian distributions P and Q is

$$D[P : Q] = \mathrm{tr}(PQ^{-1} - \log PQ^{-1} - n), \tag{4}$$

where n is the dimensionality, and this is due to the K-L divergence of multivariate Gaussian distributions.

A.4 Properties of Divergence

A divergence represents a degree of separation between two points, but neither it nor its square root is a distance. It does not necessarily satisfy the symmetry condition, so in general

$$D[P : Q] \neq D[Q : P]. \tag{5}$$

It also does not (in general) satisfy the triangle inequality. When we write $D[P : Q]$, we refer to the divergence from P to Q, and it has the dimension of the square of a distance.

One can symmetrize a divergence using the formula:

$$D_S[P : Q] = \frac{1}{2} \left(D[P : Q] + D[Q : P] \right), \tag{6}$$

but asymmetry is an important property and plays an role as we will see later.

A.5 Infinitesimal Distance and Riemannian Structure

When points P and Q on a manifold M are sufficiently close, we can define the square of an infinitesimal distance ds,

$$ds^2 = 2D[\boldsymbol{\xi} : \boldsymbol{\xi} + d\boldsymbol{\xi}] = \sum_{i,j} g_{ij} d\xi^i d\xi^j, \tag{7}$$

where $D[\boldsymbol{\xi} : \boldsymbol{\xi} + d\boldsymbol{\xi}]$ is the divergence between the points P and Q expressed in coordinates $\boldsymbol{\xi}$ and $\boldsymbol{\xi} + d\boldsymbol{\xi}$, and g_{ij} are the components of the positive-definite metric tensor $G(\boldsymbol{\xi})$.

A manifold M is said to be Riemannian when a positive-definite matrix $G(\boldsymbol{\xi})$ is defined on M, and the square of the local distance between two nearby points is given by ds^2. Given a divergence, there is an induced Riemannian structure

A.6 Geometry Induced by Divergence Functions

A divergence function induces a Riemannian metric g and a pair of torsion-free dual connections Γ and Γ^*.

The induced metric is given by:

$$g_{ij}(x) = -D_{i,j}(x,y)\big|_{x=y}, \tag{8}$$

where $D_{i,j}$ represents the partial derivatives of the divergence function with respect to the coordinates i and j.

The torsion-free dual connections are defined as:

$$\Gamma_{ijk}(x) = -D_{ij,k}(x,y)\big|_{x=y}, \tag{9}$$

$$\Gamma^*_{ijk}(x) = -D_{k,ij}(x,y)\big|_{x=y}. \tag{10}$$

Here, Γ_{ijk} and Γ^*_{ijk} are the connection coefficients, and the semicolon in the subscript denotes the covariant derivative.

The connections Γ and Γ^* are torsion-free and are dual with respect to the induced metric g. This duality is particularly important as it relates the geometric structure with the underlying statistical model.

In the geometry induced by divergence functions, the connections Γ and Γ^* can be related to the Levi-Civita connection Γ^0 as follows:

$$\Gamma_{ijk} = \Gamma^0_{ijk} - \frac{1}{2}T_{ijk}, \tag{11}$$

where Γ^0_{ijk} is the Levi-Civita connection, and T_{ijk} is the skewness tensor.

Similarly, the dual connection Γ^* is given by:

$$\Gamma^*_{ijk} = \Gamma^0_{ijk} + \frac{1}{2}T_{ijk}. \tag{12}$$

The tensor T_{ijk} is the difference between the dual connections:

$$T_{ijk} = \Gamma^*_{ijk} - \Gamma_{ijk}. \tag{13}$$

Given a divergence function D, we induce a metric g and a pair of dual connections Γ, Γ^*. An interesting question arises:

Question: If we have a metric g and a pair of dual connections Γ, Γ^*, can we find a divergence function D that induces them?

Geometry can be induced by divergence functions. Specifically, one can construct a canonical divergence by applying Hamilton-Jacobi theory for the following Lagrangian:

$$L(q,v) = \frac{1}{2} g_{ij} v^i v^j + \frac{1}{12} T_{ijk}(q) v^i v^j v^k, \tag{14}$$

where g_{ij} is the Riemannian metric and T_{ijk} is the skewness tensor. This Lagrangian will then yield a generating function which is a canonical divergence.

Divergence functions can be seen as generating functions of symplectic maps. This has connections with discrete mechanics. If D is a generating function, we can define a map $\mathbf{D}$: $T^*M \to T^*M$, where $(q,p) \mapsto (q',p')$, and M is the manifold under consideration.

The transformation of coordinates in a symplectic map given a divergence function D can be described by:

$$p_i = -\frac{\partial D}{\partial q^i}(q,p'), \tag{15}$$

$$p'_i = \frac{\partial D}{\partial q'^i}(q,p'). \tag{16}$$

Given a divergence function D, not only is there an induced Riemannian metric g, but also considering $\frac{1}{2}D$ as a generating function of a symplectic map allows us to question the approximation quality of such a function. Specifically, we assess whether $\frac{1}{2}D$ is a good approximation of the time-h geodesic flow (cost $\frac{1}{2}D$). This approach is always consistent (first-order accurate), and if D is a Bregman divergence, then the approximation is second-order accurate. Such considerations are crucial for the analysis of the geometric and dynamical properties of the system.

A.7 Important examples of Bregman Divergences

Example: Generalized KL-Divergence

For the negative entropy, the function $\Phi(\xi)$ is defined as follows:

$$\Phi(\xi) = -\sum_i \xi_i \log \xi_i, \tag{17}$$

and the associated Bregman divergence is the KL divergence, given by:

$$D_\Phi[\xi : \xi'] = \sum_i \left(\xi_i \log \frac{\xi_i}{\xi'_i} - \xi_i + \xi'_i \right) \tag{18}$$

When $\sum_i \xi'_i = \sum_i \xi_i = 1$, this is the KL divergence for probability vectors. When these normalization do not hold, this a generalization that may be used to describe circonstances when the quantity of matter or number of particules is not conserved.

Example: Free Energy of the Exponential Family

In the exponential family of distributions, there is a normalizing factor $\Psi(\theta)$, defined by:

$$\Psi(\theta) = \log \int \exp\left(\sum_i \theta_i x_i + k(x) \right) dx. \tag{19}$$

We want to compute the associated Bregman divergence. The integral I is given by:

$$I = \int p(x, \theta)\, dx = \int \exp\left(\sum_i \theta_i x_i + k(x) - \Psi(\theta)\right) dx. \tag{20}$$

Differentiating with respect to θ_i, we have:

$$\int \left(x_i - \frac{\partial}{\partial \theta_i}\Psi(\theta)\right) \phi(x, \theta)\, dx = 0. \tag{21}$$

A.8 Legendre Transform

Given a convex function $\Psi(\xi)$, consider,

$$\xi^* = \nabla\Psi(\xi),$$

which is the normal vector to the supporting tangent hyperplane to Ψ at ξ. Observe that the map from ξ to ξ^* is one-to-one, and differentiable.

We can view this as a change of coordinates, and this transformation is referred to as the Legendre transform. The transformation $\xi \to \xi^*$ is given by

$$\xi^* = \nabla\Psi(\xi).$$

Question: Is there a Ψ^* such that

$$\xi = \nabla\Psi^*(\xi^*)?$$

Consider

$$\Psi^*(\xi^*) = \xi \cdot \xi^* - \Psi(\xi), \quad \text{with} \quad \xi^* = \nabla\Psi(\xi).$$

We differentiate Φ with respect to ξ^* and we have

$$\nabla\Phi(\xi^*) = \xi + \frac{\partial^2}{\partial {\xi^*}^2}\Phi(\xi^*) \cdot \xi^* - \nabla\Phi(\xi)\frac{\partial}{\partial \xi^*}\xi^*. \tag{22}$$

Since $\xi^* = \nabla\Phi(\xi)$, the last two terms cancel, and we are left with

$$\nabla\Phi^*(\xi^*) = \xi. \tag{23}$$

So, $\xi^* = \nabla\Phi(\xi)$, $\xi = \nabla\Phi^*(\xi^*)$, and thus, Φ^* is the Legendre dual of Φ.

Observe that

$$\Phi^*(\xi^*) = \mathrm{ext}_\xi \left\{\xi \cdot \xi^* - \Phi(\xi)\right\} = \xi \cdot \xi^* - \Phi(\xi)\Big|_{\xi = \nabla\Phi^*(\xi^*)}. \tag{24}$$

The use of the extremum operator is interesting when we apply the legendre transform to functions Φ which are not necessarely convex and therefore , it may exist several inverse to the equation $\xi^* = \nabla\Phi(\xi)$

We want to check that Φ^* is convex.

We can check that the Hessian of Φ^*

$$G^*(\xi^*) = \nabla^2\Phi^*(\xi^*) = \frac{\partial^2}{\partial {\xi^*}^2}\Phi^*(\xi^*) \tag{25}$$

which is the Jacobian of the inverse transform from ξ^* to ξ, which is inverse of the Hessian of the convex function Φ. So, the Jacobian is also positive definite, since it is the inverse of a positive definite matrix. Hence, Φ^* is convex.

A.9 Relationship Between Duality of Convex Function and Duality of Bregman Divergence

Given ξ, ξ' and their dual coordinates ξ^*, $\xi^{*'}$, we define the Bregman divergence D_Φ and its dual divergence D_{Φ^*} as follows:

$$D_\Phi[\xi : \xi'] = ? \tag{26}$$

$$D_{\Phi^*}[\xi^* : \xi^{*'}] = ? \tag{27}$$

A new Bregman divergence can be derived from the dual convex function Φ^*:

$$D_{\Phi^*}[\xi^* : \xi^{*'}] = \Phi^*(\xi^*) - \Phi^*(\xi^{*'}) - \nabla\Phi^*(\xi^{*'})^T(\xi^* - \xi^{*'}), \tag{28}$$

which is the dual divergence:

$$D_{\Phi^*}[\xi^* : \xi^{*'}] = D_\Phi[\xi' : \xi] \tag{29}$$

We can check that this is true.

Theorem 1. *The Bregman divergence $D_\Phi[P : Q]$ is given by*

$$D_\Phi[P : Q] = \Phi(\xi_P) + \Phi^*(\xi_Q^*) - \langle \xi_P, \xi_Q^* \rangle, \tag{30}$$

where ξ_P is the coordinates of P in ξ coordinate system, and ξ_Q^* is the coordinates of Q in ξ^* coordinate system.

Proof. The convex conjugate Φ^* at ξ_Q^* is given by

$$\Phi^*(\xi_Q^*) = \langle \xi_Q, \xi_Q^* \rangle - \Phi(\xi_Q). \tag{31}$$

Substituting this into (1) and using the property $\nabla\Phi(\xi_Q) = \xi_Q^*$ gives the theorem. □

This theorem justifies the concept of Bregman divergence because it shows that thanks to the relationships between Bregman divergence and convex duality, we have an expression of the Bregman divergence which uses only the convex duality concept and avoid any reference to gradient and differentials. The symmetry of this expression with respect of the use of functions and its dual makes it also powerful, as we will see.

A.10 Affine and Dual Affine Coordinate Systems

Observe that when a function $\Phi(\Theta)$ is convex in a coordinate system Θ, the same function expressed in another coordinate system ζ,

$$\hat{\Phi}(\zeta) := \Phi(O(\zeta)) \tag{32}$$

is not necessarily convex as a function of ζ. The notion of convexity is invariant under affine transformations

$$O' := A \cdot O + b, \tag{33}$$

where A is a nonsingular constant matrix, b is a constant vector.

Now, we fix a coordinate system Θ in which $\Phi(\Theta)$ is convex, and we introduce geometric structures on M based on it. Consider Θ as an affine coordinate system, which endows M with an affine flat structure .

M is a flat manifold, and the coordinate axes are straight lines. Any curve $\Theta(t) \in M$ of the form,

$$\Theta(t) = at + b \tag{34}$$

is a straight line, where a, b are constant vectors.

We refer to such curves as geodesics of an affine manifold. In this context, geodesics do not mean the shortest path connecting two points, but one where the curve keeps pointing in the same direction.

The geodesics (of an affine manifold) is invariant under affine transformations.

As before, we can consider the dual coordinates Θ^*,

$$\Theta^* = \nabla \Phi(\Theta) \tag{35}$$

and consider this as another affine coordinate system,

So, this gives two notions of geodesics and flatness. We say that a manifold M with these dual structures are dually flat, and the two flat coordinates are related by the Legendre transform,

A.11 Tangent Space, Basis Vectors, and Riemannian Metric

When $d\Theta$ is an (infinitesimally) small line segment, the square of its length ds is given by,

$$ds^2 = 2D_\Phi[\Theta : \Theta + d\Theta] = \sum_{i,j} g_{ij} d\Theta^i d\Theta^j, \tag{36}$$

where the upper indices i, j represent components.

Then, we can deduce that the Riemannian metric g_{ij} is given by the Hessian of Φ,

$$g_{ij}(\Theta) = \frac{\partial^2 \Phi}{\partial \Theta^i \partial \Theta^j}(\Theta). \tag{37}$$

Remark: The tangent space is spanned by the tangent vectors to the coordinate axes of the chart.

$$\frac{\partial}{\partial q^i} = \mathbf{e}_i$$

The vector space spanned by $\{\partial/\partial\Theta^i\}$ is the tangent space of M at each point. Since Θ is an affine coordinate system, $\{\partial/\partial\Theta^i\}$ looks the same at every point.

We can write any tangent vector A as,

$$A = \sum_i A^i e_i \tag{38}$$

A small line segment $d\Theta$ can be expressed as,

$$d\Theta = \sum_i d\Theta^i e_i \tag{39}$$

We can construct $\{e^i\}$ which are the tangent vectors of the dual affine coordinate curves of Θ^*.

$$d\Theta^* = \sum_i d\Theta^i e^{*i} \tag{40}$$

We can express a vector A in terms of its components and basis vectors as:

$$A = \sum_i A^i e_i = \sum_i A_i e^{*i} \tag{41}$$

In general, $A^i \neq A_i$.

Einstein Summation Convention

When the same index appears twice, once as an upper index, and another as a lower index, then summation over that repeated index is assumed, even without the summation symbol. For example:

$$A = A^i e_i = A_i e^{*i} \tag{42}$$

Also, for a line element ds, we have:

$$ds^2 = \langle d\Theta, d\Theta \rangle = g_{ij} d\Theta^i d\Theta^j \tag{43}$$
$$= \langle d\Theta^i e_i, d\Theta^j e_j \rangle = \langle e_i, e_j \rangle d\Theta^i d\Theta^j \tag{44}$$
$$\Rightarrow g_{ij} = \langle e_i, e_j \rangle \tag{45}$$

n Euclidean space, with an orthonormal coordinate system, the metric tensor g_{ij} is defined as:

$$g_{ij} = \delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j, \end{cases} \tag{46}$$

where δ_{ij} is the Kronecker delta.

In general, a manifold induced by a convex function is non-Euclidean.

The Riemannian metric can be represented in the dual affine coordinate system Θ^*.

$$d\Theta^* = d\Theta^*_i e^{*i}$$
$$ds^2 = \langle d\Theta^*, d\Theta^* \rangle = g^{ij} d\Theta_i d\Theta_j$$

where

$$g^{ij} = \langle e^{*i}, e^{*j} \rangle$$

$$d\Theta^{*i} = g^{ij} d\Theta_j \quad \text{and} \quad d\Theta^j = g_{ij} d\Theta^{*i}$$

where $G = G^{-1}$, so the Riemannian metric tensors are mutual inverses.

$$e^{*i} = g^{ij} e_j \quad \text{and} \quad e_i = g_{ij} e^{*j}$$

$\{e_i\}$ and $\{e^{*i}\}$ are dual in the sense that

$$\langle e_i, e^{*j} \rangle = \delta_i^j$$

because $G = G^{-1}$. So, the bases are dual with respect to the Riemannian inner product.

$$A = A^i e_i = A_i e^{*i}$$
$$\langle A, e^{*j} \rangle = \left\langle A^i e_i, e^{*j} \right\rangle = A^i \langle e_i, e^{*j} \rangle = A^j$$
$$\langle A, e_j \rangle = \left\langle A_i e^{*i}, e_j \right\rangle = A_i \langle e^{*i}, e_j \rangle = A_j$$

A.12 Parallel Transport of a Vector

In the context of a dually flat manifold, this simple notion of parallel transport by keeping the coefficients fixed applies.

But, in a dually flat manifold, we have two sets of bases for the tangent space, $\{e_i\}$ and $\{e^{*i}\}$, so each of them induces a specific notion of parallel transport, which are not the same.

To define parallel transport on a manifold, we need additional structure, in particular, an affine connection.

Since M is Riemannian and not Euclidean in general, even though we can define the parallel transport easily, the length of the parallel transported vector is not constant:

$$A = A^i e_i = A_i e^{*i} \tag{47}$$

$$|A|^2 = \langle A, A\rangle = g_{ij}(\Theta)A^i A^j = A^i A_j g^{ij}(\Theta) \tag{48}$$

This depends on Θ.

We define two parralel transports via the two basis we have at our disposal

$$\begin{aligned}
\langle \pi A, \pi^* B\rangle &= \left\langle A^i e_i(\Theta), B_j g^{jk}(\Theta) e_k(\Theta)\right\rangle \\
&= A^i B_j g^{jk}(\Theta)\langle e_i(\Theta), e_k(\Theta)\rangle \\
&= A^i g_{ik}(\Theta) B^k \\
&= A_k B^k
\end{aligned}$$

The last line simplifies the expression by recognizing that $\langle e_i(\Theta), e_k(\Theta)\rangle$ yields the Kronecker delta which further simplifies the summation. it shows that by transportting one vector with a connection and the other with its dual connection, we keep the scalar product, which is the essence of the geometry.

Dually flat manifolds play a crucial role in information geometry by providing a framework where the geometric structure significantly simplifies the complexity of statistical models. These manifolds are endowed with two affine connections, which are flat and dual to each other, facilitating a unique geometric interpretation of statistical parameters.

Definition and Properties

A dually flat manifold is defined by the presence of two connections: the Levi-Civita connection ∇ and a flat, torsion-free connection ∇^* that is dual to ∇. These connections satisfy:

- ∇ is metric-compatible and torsion-free.
- ∇^* is flat and metric-compatible, but not necessarily torsion-free.
- Both ∇ and ∇^* have vanishing curvature, making them flat.

The flatness implies that locally, the manifold behaves like Euclidean space, which simplifies the computation and understanding of geodesic paths.

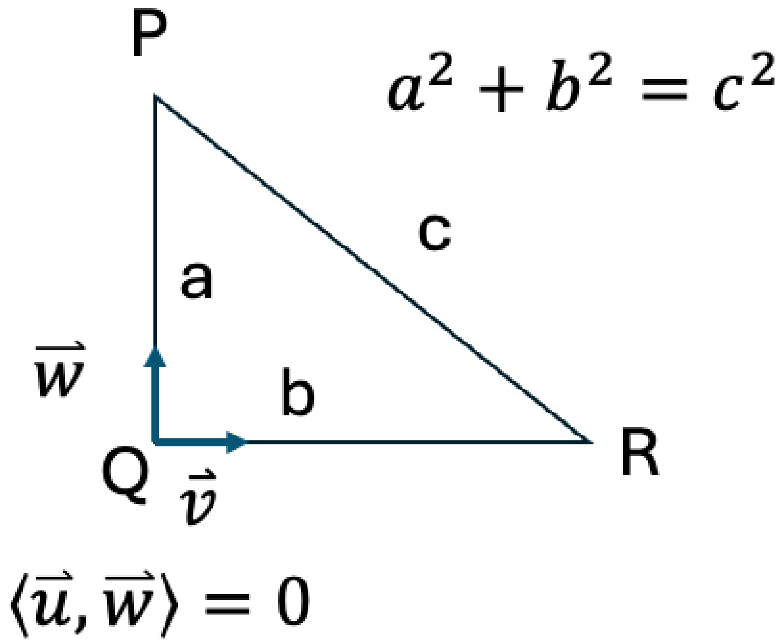

Figure 1: The pythagorean relationship when $\langle \vec{QP}, \vec{QR} \rangle = 0$.

A.13 Generalized Pythagorean Theorem

We are going to generalize the classical pythagoean theorem known for euclidian spaces:

Theorem 2 (Generalized Pythagorean Theorem)**.** *When triangle PQR is orthogonal such that the dual geodesic connecting P and Q is orthogonal to the geodesic connecting Q and R, then the generalized Pythagorean relationship holds:*

$$D_\Phi[P : R] = D_\Phi[P : Q] + D_\Phi[Q : R].$$

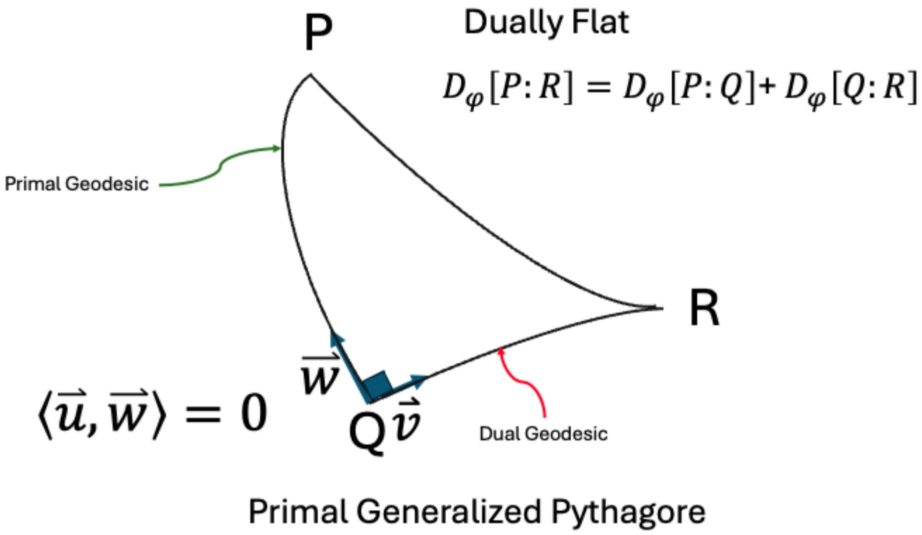

Figure 2: The pythagorean relationship when $\vec{QP}$ and $\vec{QR}$ are carried by dual vectors.

We are going to generalize the classical pythagoean theorem known for euclidian spaces:

Theorem 3 (Generalized Pythagorean Theorem)**.** *When triangle PQR is orthogonal such that the dual geodesic connecting P and Q is orthogonal to the geodesic connecting Q and R, then the generalized Pythagorean relationship holds:*

$$D_\Phi[P : R] = D_\Phi[P : Q] + D_\Phi[Q : R].$$

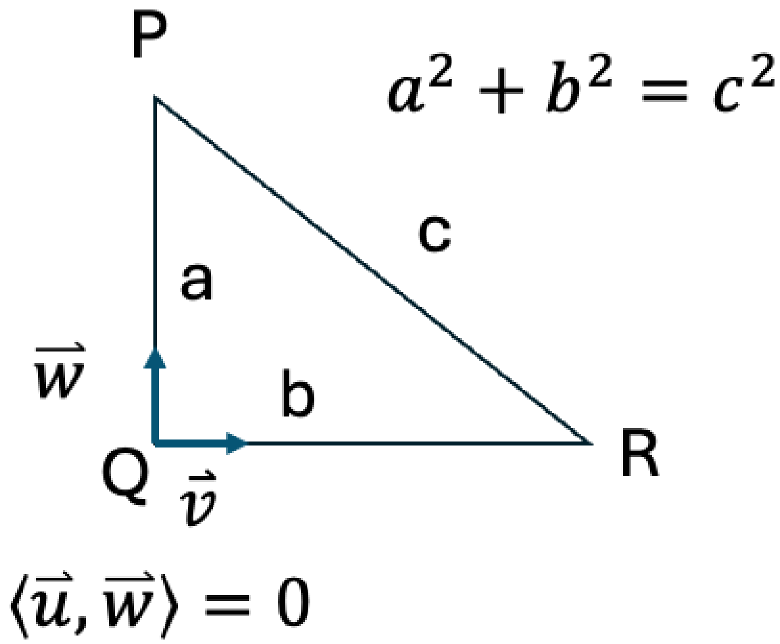

Figure 3: The pythagorean relationship when $\langle \vec{QP}, \vec{QR} \rangle = 0$.

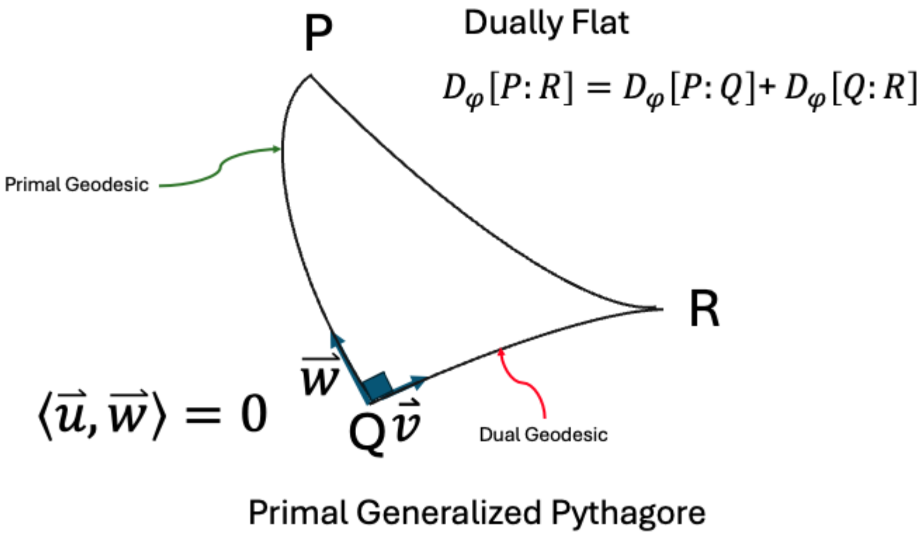

Figure 4: The pythagorean relationship when $\vec{QP}$ and $\vec{QR}$ are carried by dual vectors.

Proof of the Generalized Pythagorean Theorem

Using

$$D_\Phi[P:Q] = \Phi(\Theta_P) + \Phi^*(\Theta_Q^*) - \langle \Theta_P, \Theta_Q^* \rangle,$$

we show that

$$\begin{aligned} D_\Phi[P:Q] + D_\Phi[Q:R] - D_\Phi[P:R] &= (\Theta_P^* - \Theta_Q^*) \cdot (\Theta_Q - \Theta_R) \\ &= \text{the dual geodesic connecting } P, Q \text{ is} \\ \Theta_{PQ}^*(t) = (1-t)\Theta_P^* + t\Theta_Q^*, \\ \text{Tangent vector is } \Theta_Q^* - \Theta_P^*. \end{aligned}$$

Similarly, $\Theta_Q - \Theta_R$ is the tangent vector to the primal geodesic. But these two vectors are orthogonal, so $\star$ vanishes.

Theorem 4 (Dual Pythagorean Theorem)**.** *When triangle PQR is orthogonal such that the geodesic connecting P and Q is orthogonal to the dual geodesic connecting Q and R, the dual of the Generalized Pythagorean relation holds:*

$$D_{\Phi^*}[P:R] = D_\Phi[P:Q] + D_{\Phi^*}[Q:R].$$

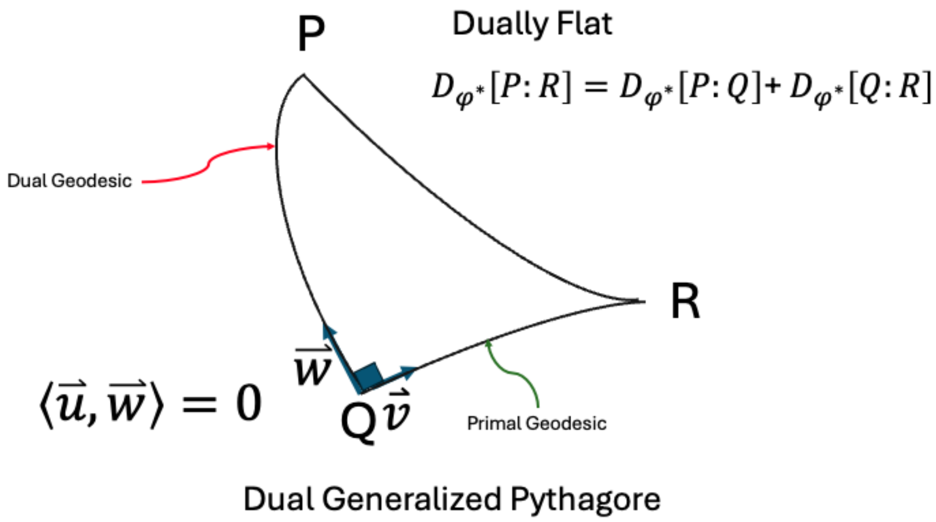

Figure 5: The dual pythagorean relationship when $\vec{QP}$ and $\vec{QR}$ are carried by dual vectors.

A.14 Projection Theorem

Geodesics for a connection ∇ are curves $\gamma(t)$ such that:

$$\nabla_{\dot\gamma(t)}\dot\gamma(t) = 0$$

where $\dot\gamma(t)$ is the tangent vector to γ at t.

Dual Geodesics for the connection ∇^* are curves $\gamma^*(t)$ that satisfy:

$$\nabla^*_{\dot\gamma^*(t)}\dot\gamma^*(t) = 0$$

Consider a point P and a smooth submanifold S in a dually flat manifold M. The divergence from a point P to the submanifold S is,

$$D_\Phi[P : S] = \inf_{R \in S} D_\Phi[P : R]. \tag{49}$$

It is natural to consider the question of finding a point in S that is closest to P in the sense of divergence. We can define the geodesic projection and dual geodesic projection of P to S in M. A curve $\Theta(t)$ is said to be orthogonal to S when its tangent vector $\dot{\Theta}(t)$ is orthogonal to any tangent vector to S at the intersection of $\Theta(t)$ with S.

Definitions : The point $\hat{P}_S$ is the geodesic projection of P to S when the geodesic connecting P to $\hat{P}_S$ is orthogonal to S.

Dual of this is that $\hat{P}_S^*$ is the dual geodesic projection of P to S, where the dual geodesic P to $\hat{P}_S^*$ is orthogonal to S.

Orthogonality at Intersection

When a geodesic γ intersects a submanifold S at a point $p = \gamma(t_0)$, it has the property that $\dot{\gamma}(t_0)$ is orthogonal to the tangent space T_pS at p. Mathematically,

$$g(\dot{\gamma}(t_0), v) = 0 \quad \forall v \in T_pS$$

where g is the metric tensor on the manifold. A similar property holds for dual geodesics γ^* intersecting S.

Geodesic vs. Dual Geodesic Projections

The projection of points via geodesics γ and dual geodesics γ^* onto S yields different intersection points (p and p^* respectively), generally with $p \neq p^*$. The tangent vectors at these points ($\dot{\gamma}(t_0)$ and $\dot{\gamma}^*(t_0^*)$) are orthogonal to T_pS and $T_{p^*}S$ respectively, but are not necessarily the same since they are determined by different connections and thus follow different paths.

A.15 Taylor-Like Expansion for Divergences

Building on this theorem, divergences can be expressed as a series expansion that mirrors a Taylor series in calculus. Starting from a base model Q_0, we iteratively refine the approximation Q_n by adding correction terms. Formally, for a divergence $S(P\|Q)$, we can write:

$$S(P\|Q_n) = \sum_{k=0}^{n} S_k,$$

where S_k represents the k-th order correction term, computed as the divergence between successive approximations. This process continues until convergence, yielding an accurate representation of the true distribution P.

A.16 Implications and Applications

The Pythagorean-like theorem and the Taylor-like expansion provide a framework for understanding divergences as dynamic, rather than static, quantities. By iteratively refining approximations, we bridge the gap between coarse-grained models and their fine-grained counterparts. This approach is fundamental to many areas, including:

- **Statistical Mechanics**: Where divergences quantify entropy changes during coarse-graining.
- **Machine Learning**: Where iterative optimization minimizes divergence-based loss functions.
- **Quantum Field Theory**: Where perturbative expansions are central to calculating physical quantities.

B Chaos Decomposition for Divergences in Generative Systems

B.1 Chaos Expansion of a Stochastic Process

The chaos expansion, also known as the Wiener chaos decomposition, states that any square-integrable random variable F in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, adapted to a Gaussian process, can be uniquely represented as an infinite sum of orthogonal components. These components are constructed using multiple stochastic integrals with respect to the underlying Gaussian process.

Formally, for a random variable $F \in L^2(\Omega)$, the expansion is given by:

$$F = \sum_{n=0}^{\infty} I_n(f_n),$$

where:

- $I_0(f_0) = \mathbb{E}[F]$ is the mean of F,
- $I_n(f_n)$ is the n-th order multiple stochastic integral of a symmetric function $f_n \in L_s^2(\mathbb{R}^n)$ with respect to the Gaussian process,
- $\{I_n(f_n)\}_{n=0}^{\infty}$ are orthogonal in $L^2(\Omega)$, meaning:

  $$\mathbb{E}[I_n(f_n) I_m(f_m)] = \delta_{n,m} n! \|f_n\|^2,$$

  where $\delta_{n,m}$ is the Kronecker delta.

B.1.1 Properties of the Chaos Expansion

- **Orthogonality**: Each component $I_n(f_n)$ belongs to the n-th Wiener chaos space, which is orthogonal to components of other orders. - **Convergence**: The series converges in $L^2(\Omega)$, ensuring that F is well-represented by the expansion. - **Interpretation**: The n-th term $I_n(f_n)$ captures the n-th order dependence of F on the underlying Gaussian process.

B.1.2 Applications

Chaos expansions are widely used in: - **Stochastic Analysis**: For solving stochastic partial differential equations (SPDEs). - **Machine Learning**: For analyzing high-dimensional stochastic processes. - **Mathematical Finance**: For decomposing complex payoff functions in derivative pricing.

Here, we explore the chaos decomposition of divergences in the context of generative systems, such as diffusion models. By breaking down divergences into components corresponding to different levels of noise and refinement, this decomposition provides a structured way to analyze the learning process of generative models.

C Quantum Field Theory: A Perturbative Approach

Quantum field theory (QFT) provides a mathematical framework to describe the interactions of particles as fields. Central to QFT is the idea that particles are excitations of underlying quantum fields, with dynamics governed by the principles of quantum mechanics and special relativity.

C.1 The Scalar Field and the Action

The simplest QFT involves a scalar field $\phi(x)$, which represents spin-0 particles. The dynamics of $\phi(x)$ are encoded in the action S, which integrates the Lagrangian density $\mathcal{L}$ over spacetime:

$$S[\phi] = \int d^4x \, \mathcal{L}(\phi, \partial_\mu \phi),$$

where $\partial_\mu \phi = \frac{\partial \phi}{\partial x^\mu}$ denotes the spacetime derivatives of $\phi(x)$.

A common example is the Klein-Gordon field, whose Lagrangian is:

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \phi)(\partial^\mu \phi) - \frac{1}{2}m^2\phi^2 - V(\phi),$$

where m is the mass of the scalar field, and $V(\phi)$ is a potential term representing interactions, e.g., $V(\phi) = \frac{\lambda}{4!}\phi^4$.

C.2 Quantization and Path Integral Formalism

In QFT, the field $\phi(x)$ is quantized. Observables are computed as expectation values of operators, which can be expressed using the path integral formalism:

$$\langle \mathcal{O} \rangle = \frac{\int \mathcal{D}[\phi] \, \mathcal{O}[\phi] \, e^{iS[\phi]}}{\int \mathcal{D}[\phi] \, e^{iS[\phi]}},$$

where $\mathcal{D}[\phi]$ denotes integration over all possible field configurations.

C.3 Perturbative Approach

For interacting fields, exact solutions to the equations of motion are often intractable. Instead, a perturbative expansion is used around the free (non-interacting) theory.

1. **Free Theory**: In the absence of interactions ($V(\phi) = 0$), the Klein-Gordon equation:

$$(\Box + m^2)\phi(x) = 0,$$

describes a free scalar particle with mass m, where $\Box = \partial_\mu \partial^\mu$ is the d'Alembertian operator.

2. **Interacting Theory**: When interactions are introduced ($V(\phi) \neq 0$), the field equation becomes non-linear:

$$(\Box + m^2)\phi(x) + \frac{\delta V}{\delta \phi} = 0.$$

Solutions are approximated using a perturbative expansion in terms of a small coupling constant λ.

3. **Feynman Diagrams**: Perturbative corrections to observables, such as scattering amplitudes, are computed using Feynman diagrams. Each diagram corresponds to a term in the perturbative series and is evaluated using Feynman rules derived from the Lagrangian.

C.4 Limits of Validity of the Theory

A crucial realization in quantum field theory is that regularization and renormalization imply the theory does not aim to describe phenomena occurring at extremely small distances or very high energies, which are equivalent due to the duality between space and energy scales.

This limitation is encapsulated in the introduction of a cutoff scale Λ, which acts as a boundary separating the regime where the theory is valid from the regime where new physics is required. Mathematically, this can be expressed as:

$$\Lambda \sim \frac{1}{(\text{minimum distance})}.$$

The renormalized theory provides accurate predictions as long as the particle energies remain well below Λ.

This acknowledgment has profound implications:

- **High-energy physics is undefined**: Ultraviolet divergences signal that the theory cannot be extrapolated indefinitely to describe phenomena at arbitrarily high energies.
- **Renormalization encodes our ignorance**: Counterterms absorb infinities, but these corrections do not explicitly describe the underlying physics at small scales.
- **New theories are required**: At energies near the Planck scale ($\sim 10^{19}$ GeV), it is expected that phenomena related to quantum gravity or other fundamental structures will emerge.

C.5 Feynman Integrals and Regularization Techniques

In quantum field theory, matrix elements at order n are computed as sums of Feynman diagrams, with each diagram corresponding to an integral over the momenta of the virtual particles. For instance, at order n, the matrix element can be expressed as:

$$\langle f|\mathcal{T}e^{-i\int d^4x\,\mathcal{L}_{\text{int}}(x)}|i\rangle = \sum_{n=0}^{\infty} \frac{(-i)^n}{n!} \int \prod_{j=1}^{n} d^4x_j \, \langle f|\mathcal{T}\left[\mathcal{L}_{\text{int}}(x_1)\cdots\mathcal{L}_{\text{int}}(x_n)\right]|i\rangle,$$

where $\mathcal{L}_{\text{int}}(x)$ is the interaction term in the Lagrangian density.

For a specific Feynman diagram D with E external lines and I internal lines, the integral is given by:

$$\mathcal{A}_D = \int \prod_{k=1}^{I} \frac{d^4 p_k}{(2\pi)^4} \prod_{\text{internal lines}} \frac{i}{p_k^2 - m^2 + i\epsilon} \prod_{\text{vertices}} \mathcal{V}(x_j),$$

where p_k is the momentum associated with an internal line, and $\mathcal{V}(x_j)$ encodes the vertex factors according to Feynman rules.

However, these integrals often diverge due to contributions from very large momenta (ultraviolet divergences). To handle these divergences, regularization techniques are applied.

C.5.1 Regularization Techniques

1. **Cutoff Regularization**: A cutoff Λ is introduced to limit the integration over large momenta:

$$\int d^4 k \to \int_{|k| \leq \Lambda} d^4 k.$$

This physically corresponds to ignoring contributions from energy scales above Λ.

2. **Dimensional Regularization**: The number of spacetime dimensions is analytically continued to $d = 4 - \epsilon$:

$$\int d^4 k\, f(k) \to \mu^{2\epsilon} \int d^d k\, f(k),$$

where μ is a renormalization scale.

3. **Pauli-Villars Regularization**: Propagators are modified to suppress ultraviolet contributions:

$$\frac{1}{k^2 - m^2 + i\epsilon} \to \frac{1}{k^2 - m^2 + i\epsilon} - \frac{1}{k^2 - \Lambda^2 + i\epsilon}.$$

C.5.2 Renormalization

After regularization, divergences are absorbed into redefined ("renormalized") parameters: - **Mass Renormalization**:

$$m_{\text{phys}}^2 = m_0^2 + \delta m^2,$$

where m_{phys} is the physical mass, m_0 the bare mass, and δm^2 a counterterm.

- **Coupling Renormalization**:

$$\lambda_0 = \lambda_R + \delta\lambda,$$

where λ_R is the renormalized coupling constant.

- **Field Renormalization**:

$$\phi_0 = Z_\phi^{1/2} \phi_R,$$

where Z_ϕ is the wavefunction renormalization constant.

C.6 A Theory of Effective Scales

Quantum field theory is, therefore, an **effective theory**, designed to provide accurate predictions within a limited energy range. It does not claim to describe the ultimate laws of nature but instead serves as an approximation valid at intermediate scales where effects of quantum gravity or other fundamental interactions can be neglected.

This conceptual framework has allowed for remarkably precise predictions, such as the anomalous magnetic moment of the electron, while acknowledging that physics at extremely high energies remains beyond reach.

C.7 Renormalization and Divergences

Interactions in QFT lead to divergences in perturbative calculations. Renormalization addresses these divergences by absorbing infinities into redefined parameters (e.g., mass m, coupling λ).

Key steps include: - Regularization: Introduce a cutoff or dimensional regularization to control divergences. - Renormalization: Redefine physical parameters to ensure finite, measurable predictions.

C.8 Applications and Simplifications

The scalar field theory provides a simplified setting to understand concepts in QFT without the additional complexity of spinors or gauge fields. Despite its simplicity, it shares many features with more complex theories, such as: - **Harmonic Oscillator Analogy**: The field modes behave like an infinite collection of coupled harmonic oscillators. - **Perturbative Insights**: Scalar QFT serves as a pedagogical model for understanding perturbative expansions in more intricate theories.

This approach provides a foundation for more advanced topics, including gauge theories and quantum electrodynamics, by focusing on the interplay between fields, interactions, and symmetries.

D Cotler-Rezchikov Equation Revisited

In this appendix, we revisit the Cotler-Rezchikov equation, analyzing its implications for optimal transport and its reinterpretation as a gradient flow. Connections to the Taylor-like decomposition of divergences and chaos decomposition are discussed, showing how these concepts unify under a common framework.

The renormalization group (RG) is a central concept in quantum field theory (QFT) and statistical field theory. It provides a framework to understand how the effective description of a physical system evolves as we adjust the precision of our measurement apparatus. Polchinski's equation offers a functional differential equation describing RG flow in terms of relative entropy.

D.1 Renormalization Group Flow

For a scalar field theory, the RG flow is represented as:

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

where: - $P_\Lambda[\phi] \propto e^{-S_\Lambda[\phi]}$ is the probability functional corresponding to the effective description of the system at the cutoff scale Λ. - $\mathcal{F}$ encodes the specifics of the RG scheme.

The left-hand side represents the coarse-graining of the momentum space as we lower the cutoff Λ. The functional $\mathcal{F}$ on the right-hand side depends on the interaction terms, which evolve as modes above Λ are integrated out.

D.2 Matrix Element Expansion

The matrix element expansion provides the probability amplitude for a transition between an initial state $|i\rangle$ and a final state $|f\rangle$. This is expressed as:

$$\langle f|Te^{-i\int d^4x\,\mathcal{L}_{\text{int}}(x)}|i\rangle = \sum_{n=0}^{\infty} \frac{(-i)^n}{n!} \int \prod_{j=1}^{n} d^4x_j \,\langle f|T\big[\mathcal{L}_{\text{int}}(x_1)\cdots\mathcal{L}_{\text{int}}(x_n)\big]|i\rangle.$$

Each term in this expansion corresponds to an n-point correlation function, represented by Feynman diagrams at order n.

D.2.1 Polchinski's Equation and the Partition Function

Polchinski's equation governs the evolution of the partition function $Z_\Lambda[J]$ under changes in the momentum cutoff Λ. The partition function is:

$$Z_\Lambda[J] = \int \mathcal{D}[\phi]\, e^{-\frac{1}{2}\int \frac{d^4p}{(2\pi)^4}\left[\phi(p)(p^2+m^2)K_\Lambda^{-1}(p^2)\phi(-p)+J(p)\phi(-p)\right]-S_{\text{int},\Lambda}[\phi]},$$

where: - $K_\Lambda(p^2)$ is a soft cutoff function that suppresses high-momentum modes. - $S_{\text{int},\Lambda}[\phi]$ encodes the effects of interactions and bare couplings.

Polchinski's equation describes how $Z_\Lambda[J]$ evolves as Λ changes:

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

D.2.2 Partition Function and Generating Functional

The partition function with a source $J(x)$ is given by:

$$Z_\Lambda[J] = \int \mathcal{D}[\phi]\, e^{-S_\Lambda[\phi]+\int d^4x\, J(x)\phi(x)},$$

where: - $S_\Lambda[\phi]$ is the effective action at the cutoff scale Λ. - $J(x)$ is an external source that allows us to probe the theory.

The generating functional $W_\Lambda[J]$ is the logarithm of the partition function:

$$W_\Lambda[J] = \ln Z_\Lambda[J].$$

From this, the effective action can be derived:

$$\Gamma_\Lambda[\phi] = W_\Lambda[J] - \int d^4x\, J(x)\phi(x),$$

where:

$$\phi(x) = \frac{\delta W_\Lambda[J]}{\delta J(x)}.$$

D.2.3 Correlation Functions from Functional Differentiation

The n-point correlation function is obtained by taking functional derivatives of $Z_\Lambda[J]$ with respect to $J(x)$:

$$G_\Lambda^{(n)}(x_1, x_2, \ldots, x_n) = \left. \frac{\delta^n Z_\Lambda[J]}{\delta J(x_1)\delta J(x_2)\cdots\delta J(x_n)} \right|_{J=0}.$$

For example: - The 2-point function (propagator):

$$G_\Lambda^{(2)}(x_1, x_2) = \left. \frac{\delta^2 Z_\Lambda[J]}{\delta J(x_1)\delta J(x_2)} \right|_{J=0}.$$

- The 4-point function (vertex interaction):

$$G_\Lambda^{(4)}(x_1, x_2, x_3, x_4) = \left. \frac{\delta^4 Z_\Lambda[J]}{\delta J(x_1)\delta J(x_2)\delta J(x_3)\delta J(x_4)} \right|_{J=0}.$$

These correlation functions are directly related to the amplitudes computed using Feynman diagrams.

D.2.4 Matrix Elements from Correlation Functions

The n-point correlation functions are connected to the S-matrix elements via the LSZ reduction formula. The LSZ formula relates the amputated correlation functions (without external propagators) to the physical scattering amplitudes.

For example, the matrix element for an interaction is:

$$\langle f|Te^{-i\int d^4x\,\mathcal{L}_{\text{int}}(x)}|i\rangle = \sum_{n=0}^{\infty} \frac{(-i)^n}{n!} \int \prod_{j=1}^{n} d^4x_j\, \langle f|T\left[\mathcal{L}_{\text{int}}(x_1)\cdots\mathcal{L}_{\text{int}}(x_n)\right]|i\rangle.$$

Each term in this expansion corresponds to an n-point correlation function and is computed from $Z_\Lambda[J]$ through functional derivatives.

D.2.5 Connection Between Polchinski's Equation and Matrix Elements

Polchinski's equation modifies $Z_\Lambda[J]$ as the cutoff scale Λ is changed. This, in turn, affects the correlation functions and the matrix elements derived from $Z_\Lambda[J]$. Specifically:

- Polchinski's equation governs how the interaction terms in the Lagrangian evolve with Λ, ensuring that $S_\Lambda[\phi]$ is finite.
- The modified interaction terms affect the propagators and vertices in Feynman diagrams, thus altering the matrix elements.

This demonstrates that the matrix element expansion is a perturbative realization of the effective action evolution described by Polchinski's equation.

D.3 The Role of the Cutoff Function in Effective Field Theory

One of the most remarkable aspects of renormalized quantum field theory is the behavior of the cutoff function $K_\Lambda(p^2)$. This function allows us to clearly distinguish the **meso-physics**, which the theory aims to describe, from the **micro-physics**, where we acknowledge our ignorance. The cutoff function achieves this through its algebraic and functional properties, which regulate the contributions of different energy scales to the effective action $S_\Lambda[\phi]$.

D.3.1 The Cutoff Function and Scale Separation

The cutoff function $K_\Lambda(p^2)$ plays a crucial role in separating energy scales:

$$K_\Lambda(p^2) \sim \begin{cases} 1, & \text{for } p^2 \ll \Lambda^2 \quad \text{(low-energy modes contribute fully)}, \\ 0, & \text{for } p^2 \gg \Lambda^2 \quad \text{(high-energy modes are suppressed)}. \end{cases}$$

This separation ensures that the effective action $S_\Lambda[\phi]$ retains only the relevant degrees of freedom at the scale Λ, while the effects of high-energy modes ($p^2 > \Lambda^2$) are integrated out.

The partition function incorporating the cutoff is:

$$Z_\Lambda[J] = \int \mathcal{D}[\phi]\, e^{-S_\Lambda[\phi] + \int d^4x\, J(x)\phi(x)},$$

where the cutoff function modifies the propagators and interaction terms in $S_\Lambda[\phi]$.

D.3.2 Polchinski's Equation and the Diffusion Along Λ

The evolution of $S_\Lambda[\phi]$ with the cutoff scale Λ is governed by Polchinski's equation:

$$-\Lambda \frac{d}{d\Lambda} S_\Lambda[\phi] = \int \frac{d^4p}{(2\pi)^4} \left[K'_\Lambda(p^2) \frac{\delta S_\Lambda[\phi]}{\delta \phi(p)} \frac{\delta S_\Lambda[\phi]}{\delta \phi(-p)} - K_\Lambda(p^2) \frac{\delta^2 S_\Lambda[\phi]}{\delta \phi(p) \delta \phi(-p)} \right],$$

where:

- The term $K'_\Lambda(p^2)$ controls the "flow" of high-energy modes being integrated out.
- The term $K_\Lambda(p^2)$ acts as a smoothing or diffusive operator, spreading contributions across scales.

This equation reveals a **diffusion-like process** along the dimension Λ, where Λ acts as a "scale-time" parameter.

D.3.3 The Functional Dependence of $S_\Lambda[\phi]$

The functional dependence of $S_\Lambda[\phi]$ on the cutoff $K_\Lambda(p^2)$ is key to understanding the separation of meso- and micro-physics. Specifically:

$$\frac{\delta S_\Lambda[\phi]}{\delta \phi(p)} \quad \text{and} \quad \frac{\delta^2 S_\Lambda[\phi]}{\delta \phi(p) \delta \phi(-p)}$$

encode how the effective action evolves as modes are integrated out. The explicit dependence on $K_\Lambda(p^2)$ ensures that:

- Low-energy modes ($p^2 \ll \Lambda^2$) contribute fully to the effective dynamics.
- High-energy modes ($p^2 \gg \Lambda^2$) influence the effective action only indirectly, through renormalization of the couplings.

The interplay between these terms leads to a controlled evolution of $S_\Lambda[\phi]$, where the function $K_\Lambda(p^2)$ acts as a filter between scales.

D.3.4 Separation of Meso-Physics and Micro-Physics

The cutoff function allows for a clear distinction between:

- **Meso-Physics**: The phenomena at energy scales well below Λ, which are fully described by the effective action $S_\Lambda[\phi]$.
- **Micro-Physics**: The phenomena at very high energy scales ($p^2 \gg \Lambda^2$), which are not explicitly modeled but whose effects are encoded in the renormalized parameters.

This separation is achieved dynamically through the diffusion-like behavior of Polchinski's equation, as the high-energy contributions are progressively absorbed into effective couplings.

D.3.5 A Reflection of Natural Laws

The behavior of $K_\Lambda(p^2)$ mirrors natural processes, such as:

- **Diffusion**: Contributions from high-energy modes spread out and are redistributed as the cutoff Λ decreases.
- **Hierarchy of Scales**: The introduction of a cutoff naturally reflects the hierarchical structure of physical laws, where macroscopic phenomena emerge from microscopic interactions without requiring detailed knowledge of the latter.

D.3.6 Conclusion

The algebraic and functional properties of the cutoff function $K_\Lambda(p^2)$ form the cornerstone of renormalized quantum field theory. They allow us to:

- Distinguish between meso- and micro-physics, acknowledging our ignorance of phenomena at very small scales.
- Manage divergences through renormalization while retaining predictive power at accessible energy scales.
- Recognize the effective nature of quantum field theory as a tool for describing physics within a given range of scales.

This approach highlights the elegance of quantum field theory as both a practical predictive framework and a philosophical statement about the limits of human knowledge in physics.

D.4 Polchinski's Equation

Polchinski's equation is derived for a Euclidean scalar field theory with a source J. The partition function is:

$$Z_\Lambda[J] := \int \mathcal{D}[\phi]\, e^{-\frac{1}{2}\int \frac{d^4p}{(2\pi)^4}\left[\phi(p)(p^2+m^2)K_\Lambda^{-1}(p^2)\phi(-p)+J(p)\phi(-p)\right]-S_{\mathrm{int},\Lambda}[\phi]},$$

where: - $K_\Lambda(p^2)$ is a soft cutoff function ensuring that high-momentum correlations are suppressed. - $S_{\mathrm{int},\Lambda}[\phi]$ includes interaction terms and bare couplings.

The RG flow integrates out modes above a smaller scale $\Lambda_R < \Lambda$, which modifies the effective action and parameters.

Polchinski's equation for the scale Λ is given by:

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

where C_Λ depends only on the cutoff scale Λ. This equation reflects how the cutoff scale Λ modifies both the kinetic term and the interaction terms in the action.

D.5 Effective Description and the Cutoff Scale

The RG analysis implies that the effective description of the system changes as the cutoff scale Λ is lowered. The coarse-graining process reduces the precision of the theory by integrating out high-momentum modes, leading to a simpler description at larger scales.

Key features of Polchinski's equation:

- It describes how the effective theory evolves under changes in the cutoff Λ.
- It highlights the interplay between the kinetic term and the interaction terms in the action.
- It provides a non-perturbative framework for studying RG flow.

References

[1] Cotler, Jordan, and Semon Rezchikov. *Renormalization Group Flow as Optimal Transport.* arXiv preprint arXiv:2202.11737, 2022. https://arxiv.org/abs/2202.11737

[2] Buehler, Hans, Peter Murray, Mikko S. Pakkanen, and Ben Wood. *Deep Hedging: Learning to Remove the Drift under Trading Frictions with Minimal Equivalent Near-Martingale Measures.* arXiv preprint arXiv:1904.05234, 2019. https://arxiv.org/abs/1904.05234

[3] Cattiaux, Patrick, Arnaud Guillin, and Pierre Monmarché. *Time Reversal of Diffusion Processes Under a Finite Entropy Condition.* arXiv preprint arXiv:2104.07708, 2021. https://arxiv.org/abs/2104.07708

[4] Zhang, Xinhua. *Bregman Divergence and Mirror Descent.* Lecture Notes, University of Illinois at Chicago. https://www2.cs.uic.edu/~zhangx/teaching/bregman.pdf

[5] Ho, Jonathan, Ajay Jain, and Pieter Abbeel. *Denoising Diffusion Probabilistic Models.* arXiv preprint arXiv:2006.11239, 2020. https://arxiv.org/abs/2006.11239

[6] Nualart, David. *The Malliavin Calculus and Related Topics.* Springer Science & Business Media, 2006. ISBN: 978-3-540-28329-6. This book provides a comprehensive treatment of the chaos expansion of stochastic processes, particularly in the context of Malliavin calculus.