

解説

拡散モデルを経路積分で考える

広野 雄士

筑波大学 システム情報系*

Generative Diffusion Models through the Lens of Path Integrals

Yuji Hirono

Institute of Systems and Information Engineering, University of Tsukuba*

概要

本稿では、機械学習における生成モデルの新たな潮流である拡散モデルと物理学の概念との接点について論じる。拡散モデルは画像・音声・動画生成において顕著な成果を上げており、その仕組みを紐解くと物理学の考え方と自然につながる点が多い。まず、拡散モデルの基本となる順過程・逆過程と Langevin 方程式の関連について述べ、次に量子力学の経路積分を応用した拡散モデルの新たな定式化について紹介する。この定式化では、確率的・決定論的サンプリング間を補間するパラメータが量子力学の Planck 定数に似た役割を果たす。また、この経路積分表示に基づいて WKB 近似を応用し、異なるサンプリング手法の性能差を解析した研究についても述べる。物理学の視点から拡散モデルを捉え直すことで、そのダイナミクスの本質への理解が深まり、新たな洞察が広がることが期待される。

1. はじめに

拡散モデル¹⁾は近年の深層生成モデル分野において高い性能を示す手法である。敵対的生成ネットワーク (Generative Adversarial Network, GAN)²⁾や変分オートエンコーダ (Variational Autoencoder, VAE)^{3,4)}などの他の生成モデルとは異なり、拡散モデルは、データに段階的にノイズを加えて完全なノイズ分布に変換する順過程と、学習したノイズ除去過程を通じて元の分布を復元する逆過程を組み合わせることで、サンプリングの安定性と多様性を両立させる特徴を持つ。この漸進的なノイズ除去アプローチにより、GAN が抱えるモード崩壊 (学習が一部のモードに集中し、多様性が失われる現象) や訓練の不安定性、VAE が示す質の限界といった問題を拡散モデルは克服している。これにより、画像、音声、動画など多種多様な高次元データに対して、非常に高品質なサンプルを得ることが可能となった。

拡散モデルの応用は連続値を持つデータだけでなく、離散的なデータにも広がっている。特に自然言語処理の分野では、離散拡散モデルが言語モデルに応用

され、テキスト生成タスクで顕著な成果を上げている⁵⁻⁷⁾。離散拡散モデルでは、連続的なノイズ付加の代わりにマスキングやトークン置換などの操作を用いて順過程を実現し、逆過程でテキストを生成する。このアプローチは、大規模言語モデル (Large Language Model, LLM) の分野においても新たな可能性を開いており、従来の言語モデルが文章を左から右へ一単語ずつ逐次的に生成するのとは異なり、複数の単語を同時に生成・修正できる並列処理が可能になる。これにより、生成の効率性が向上するだけでなく、文章全体の一貫性を考慮した段階的な推論過程を明示的にモデル化することで、より柔軟で制御可能なテキスト生成を実現している。

本稿は、拡散モデルの仕組みを物理学的視点から解釈することを目的とする。拡散モデルの訓練・生成過程で用いられる確率微分方程式は、Langevin 方程式として物理学で長らく用いられてきたものである。さらに、量子力学の定式化に使われる、全ての可能な経路の寄与を重ね合わせる経路積分の方法により、拡散モデルを記述することができる⁸⁾。この経路積分による定式化を通じて、拡散モデルと量子力学など物理学の基本的概念との自然な対応関係が浮かび上がる。量子力学に従う粒子の運動は経路積分によって記述さ

* 〒305-8577 茨城県つくば市天王台 1-1-1

DOI: 10.3902/jnns.32.81

れ、Planck 定数 \hbar がゼロの極限としての古典的軌道から、そこからの補正として量子的揺らぎを取り入れることができる。同様に、拡散モデルにおけるサンプリング過程も、決定論的な軌道を中心とした確率的揺らぎとして理解することができる。

このように拡散モデルと物理学の概念には多くの接点があり、本稿ではその深い関連性に焦点を当てる。学習済みのスコア関数を用いた逆過程では、ノイズから徐々に構造が現れる過程が、確率勾配に沿った制御された動力学系や最適輸送問題の枠組みとも関連している。拡散モデルを単なる実用的な生成技術としてではなく、物理学や統計力学の広範な理論と結びつけた現象として捉え直すことで、生成過程への理解がさらに深まり、生成モデルの理論的發展および応用に向けた新たな視点が得られることを期待する。

2. 拡散モデルの仕組み

本章では、拡散モデルの基本的な枠組みとその物理学的な背景について解説する。

2.1 拡散モデルの基本原理と物理学的背景

何らかの訓練データ（例えば画像の集合）が与えられたとしよう。生成モデルである拡散モデルの目的は、訓練データと似た画像を新たに作り出すことである。そのための基本的な仮定として、訓練データの画像はある確率分布 $p_{\text{data}}(\mathbf{x})$ からサンプルされたものである。とする。この仮定のもとで、我々の目的は、このデータ分布からのサンプリングを行うこととなる。

データ分布 $p_{\text{data}}(\mathbf{x})$ からのサンプリングを行うには、2つの課題が存在する。第一に、データ分布は未知であるため、与えられた訓練画像から推定する必要がある。第二に、仮にデータ分布を正確に知っていたとしても、サンプリング自体に困難が伴う。これは、データ分布が多くの場合、画像の次元数に比べて低次元に集中しており、また複数のピークを持つ多峰性の特徴があるためである。このような分布に対してマルコフ連鎖モンテカルロ法を適用すると、異なるピーク間の移動に多大な時間を要してしまう。拡散モデルはこれら2つの問題をうまく解決している。

ここでは、確率微分方程式 (Stochastic Differential Equation, SDE)⁹⁾を用いた拡散モデルの定式化¹⁰⁾を解説する。訓練データの集合を \mathcal{D} とし、その要素が D 次元ベクトル $\mathbf{x}_0 \in \mathbb{R}^D$ であるとしよう。例えば画像の場合は、ベクトルの要素が各画素の値だと思えることができる。 $\mathbf{x}_0 \in \mathcal{D}$ を一つ選び、このデータに徐々にノイズを加えていく過程を考えよう。具体的には、 $\mathbf{x}_{t=0} = \mathbf{x}_0$ という初期状態から出発し、次の SDE に

よる時間発展を考える：

$$d\mathbf{x}_t = \mathbf{f}(\mathbf{x}_t, t) dt + g(t) d\mathbf{w}_t \quad (1)$$

この式において、 $d\mathbf{x}_t = \mathbf{x}_{t+dt} - \mathbf{x}_t$ は微小時間 dt における状態変化を表し、右辺第1項 $\mathbf{f}(\mathbf{x}_t, t) dt$ は決定論的な変化を、右辺第2項 $g(t) d\mathbf{w}_t$ は確率的なノイズの寄与を表している。ここで \mathbf{w}_t は標準 Wiener 過程 (Brown 運動) であり、 $d\mathbf{w}_t$ はその微小時間における増分である。

この方程式に従い初期状態 \mathbf{x}_0 を $t=0$ から $t=T$ まで発展させると、ノイズの実現値 $\{\mathbf{w}_t\}_{t \in [0, T]}$ ごとに異なる軌跡 $\{\mathbf{x}_t\}_{t \in [0, T]}$ が生成される。同一の初期状態からスタートしても、ノイズの確率的性質により多様な経路が出現する。特に画像データの場合、このノイズによる擾乱は画像の構造を段階的に崩壊させ、十分な時間経過後には元の情報がほぼ失われた「砂嵐状態」へと至る。

このノイズを加えていく過程は、確率分布の連続的な変形として捉えることも可能である（流体力学におけるラグランジュ描像・オイラー描像の関係と類似している）。初期分布 $p_0(\mathbf{x})$ からサンプルされた多数の初期点が上記 SDE に従って時間発展するとき、任意の時刻 t におけるこれらの点の分布は、十分なサンプル数の極限において確率密度 $p_t(\mathbf{x})$ を形成する。すなわち、 $p_0(\mathbf{x})$ からスタートした分布が徐々に変形されて、最終的に $p_T(\mathbf{x})$ へと至るプロセスとして解釈できるのである。SDE(1) に対応する確率分布の時間発展は、以下の Fokker-Planck 方程式によって与えられる：

$$\frac{\partial}{\partial t} p_t(\mathbf{x}) = -\nabla \cdot \left[\mathbf{f}(\mathbf{x}, t) p_t(\mathbf{x}) - \frac{g(t)^2}{2} \nabla p_t(\mathbf{x}) \right] \quad (2)$$

この方程式は SDE の確率的な軌道を確率分布の時間発展として捉え直したものである。右辺第1項はドリフト項による分布の「流れ」を、第2項はノイズによる分布の「拡散」を表している。

拡散モデルの基本的なアイデアは以下の通りである。データ分布 $p_{t=0}(\mathbf{x}) = p_{\text{data}}(\mathbf{x})$ を出発点とし、これを徐々に変形していく過程（順過程）を考え、最終的に単純な分布 $p_{t=T}(\mathbf{x}) = \pi(\mathbf{x})$ が得られたとする。この変形プロセスを $t=T$ から時間を遡る形で進める（逆過程）ことで、分布 $\pi(\mathbf{x})$ から出発して、 $t=0$ においてデータ分布 $p_{\text{data}}(\mathbf{x})$ を復元することが可能となる。順過程は SDE (1) によって記述され、逆過程は後に述べるように SDE あるいは常微分方程式

(Ordinary Differential Equation, ODE) によって記述される。この逆過程を利用して、事前分布 $\pi(\mathbf{x})$ からサンプリングした \mathbf{x}_T を $t = 0$ まで逆時間発展させることにより、目標とするデータ分布 $p_{\text{data}}(\mathbf{x})$ からの標本生成が実現できる。

順過程の時間発展は $\mathbf{f}(\mathbf{x}, t)$ および $g(t)$ の選択に依存する。代表的な選択肢として、長時間の後にガウス分布へと収束する設定が広く採用されている。例として、Denoising Diffusion Probabilistic Models¹¹⁾ と呼ばれるモデル（の連続時間版）では、

$$\mathbf{f}(\mathbf{x}, t) = -\frac{1}{2}\beta(t)\mathbf{x}, g(t) = \sqrt{\beta(t)} \quad (3)$$

という形式が採用される。この時間依存パラメータの調整法はノイズスケジューリングと呼ばれ、典型的には時間 t の進行とともにノイズの強度が増大するように設計される。最適なスケジューリングは画像サイズや目的とするタスクに応じて個別に調整される。

式 (3) の特性として、確率流 $\mathbf{J}_t(\mathbf{x}) := \mathbf{f}(\mathbf{x}, t)p_t(\mathbf{x}) - \frac{g(t)^2}{2}\nabla p_t(\mathbf{x})$ が完全に消失する定常状態の存在が挙げられる。式 (3) を確率流の定義式に代入すると、 $\mathbf{J}_t(\mathbf{x}) = \mathbf{0}$ を満たす確率分布は、平均 $\mathbf{0}$ 、共分散行列が単位行列の多変量ガウス分布、

$$p_{\text{ss}}(\mathbf{x}) \propto e^{-\frac{\|\mathbf{x}\|^2}{2}}$$

となる。十分な時間経過後に、確率分布は定常状態 $p_{\text{ss}}(\mathbf{x})$ へと収束していく。

逆過程の時間発展の SDE 版がどのように与えられるのかを以下で述べる。Fokker-Planck 方程式は時間について逆向きに解くことは可能だが、対応する SDE を考える際に問題が生じる。Fokker-Planck 方程式は拡散項

$$\partial_t p_t(\mathbf{x}) = \frac{g(t)^2}{2}\nabla^2 p_t(\mathbf{x}) + \dots$$

を含んでいるが、逆再生ダイナミクスを考えるとこの項の符号が反転し、「逆拡散方程式」となってしまう。このため、通常の SDE と Fokker-Planck 方程式の対応関係を直接適用することができない。この問題を解決するために、Fokker-Planck 方程式を以下のように変形する。

$$\partial_t p_t(\mathbf{x}) = -\nabla \cdot \left[\mathbf{f}(\mathbf{x}, t)p_t(\mathbf{x}) - \frac{g(t)^2}{2}\nabla p_t(\mathbf{x}) + \frac{g(t)^2}{2}\nabla p_t(\mathbf{x}) - \frac{g(t)^2}{2}\nabla p_t(\mathbf{x}) \right]$$

$$= -\nabla \cdot \left[\left(\mathbf{f}(\mathbf{x}, t) - g(t)^2\nabla \log p_t(\mathbf{x}) \right) p_t(\mathbf{x}) + \frac{g(t)^2}{2}\nabla p_t(\mathbf{x}) \right]$$

ここでは、下線を引いた部分（足すとゼロになる）を加え、また $\nabla \log p_t(\mathbf{x}) = \nabla p_t(\mathbf{x})/p_t(\mathbf{x})$ を用いた。すなわち、 $\mathbf{f}(\mathbf{x}, t) - g(t)^2\nabla \log p_t(\mathbf{x})$ の部分を改めて「力」と見做すことにより、拡散項の部分が

$$\partial_t p_t(\mathbf{x}) = -\frac{g(t)^2}{2}\nabla^2 p_t(\mathbf{x}) + \dots$$

と符号を変え、時間を逆再生した際に拡散項として正しい符号を持つようになる。Fokker-Planck 方程式と Langevin 方程式の対応関係を用いると、逆過程を表す SDE は以下のように与えられる：

$$d\mathbf{x}_t = \left[\mathbf{f}(\mathbf{x}_t, t) - g(t)^2\nabla \log p_t(\mathbf{x}_t) \right] dt + g(t)d\tilde{\mathbf{w}}_t \quad (4)$$

ここで $\tilde{\mathbf{w}}_t$ は時間逆向き Wiener 過程であり、 $\nabla \log p_t(\mathbf{x}_t)$ はスコア関数と呼ばれる。したがって、スコア関数が得られれば、この式を時間逆向きに解くことで拡散過程の逆再生、すなわち画像生成が可能となる。

生成過程の SDE はスコア関数 $\nabla \log p_t(\mathbf{x})$ に依存しているが、 $p_t(\mathbf{x}_t)$ は未知であるため、直接解くことはできない。そこで、ニューラルネットワークによる関数 $\mathbf{s}_\theta(\mathbf{x}, t)$ を導入し、 $\mathbf{s}_\theta(\mathbf{x}, t)$ が真のスコア $\nabla \log p_t(\mathbf{x})$ をよく近似するように、パラメータ θ を最適化する。これが拡散モデルにおける訓練過程である。

画像生成は、厳密な逆過程の SDE において、スコア関数の部分を学習された $\mathbf{s}_\theta(\mathbf{x}, t)$ で置き換えた SDE

$$d\mathbf{x}_t = \left[\mathbf{f}(\mathbf{x}_t, t) - g(t)^2\mathbf{s}_\theta(\mathbf{x}, t) \right] dt + g(t)d\tilde{\mathbf{w}}_t$$

を $t = T$ から $t = 0$ に向かって解くことで実行される。

逆過程の初期時刻 ($t = T$) におけるモデル分布を $q_T(\mathbf{x}) = \pi(\mathbf{x})$ と設定する。ここで $\pi(\mathbf{x})$ は事前分布と呼ばれ、通常はガウス分布が採用される。 $q_T(\mathbf{x})$ からスタートして、訓練されたスコア関数 $\mathbf{s}_\theta(\mathbf{x}, t)$ を用いて時間を逆向きに発展させたときの各時刻における確率分布を $q_t(\mathbf{x})$ と表記する。スコア関数が $\mathbf{s}_\theta(\mathbf{x}, t) \simeq \nabla \log p_t(\mathbf{x})$ という十分良い近似になっていれば、 $q_{t=0}(\mathbf{x})$ が元のデータ分布 $p_{t=0}(\mathbf{x}) = p_{\text{data}}(\mathbf{x})$ を良く近似することが期待される。 $\mathbf{s}_\theta(\mathbf{x}, t)$ は以下の

ような損失関数を用いて最適化される：

$$\mathcal{L}(\theta) = \int_0^T \frac{g(t)^2}{2} \mathbb{E}_{p_t} [\|s_\theta(\mathbf{x}_t, t) - \nabla \log p_t(\mathbf{x}_t)\|^2]$$

この損失関数は、モデルの作る分布とデータ分布の Kullback-Leibler ダイバージェンスの上限を与える：

$$D_{\text{KL}}(p_0 \| q_0) \leq D_{\text{KL}}(p_T \| q_T) + \mathcal{L}(\theta)$$

データ分布の再現においてスコア関数が重要な役割を果たすことには直感的な解釈がある。例えば、人物画像に帽子を被せるといった修正を行った場合、確率分布の値が上昇するならば、そのような特徴を持つ人物が実際に存在する蓋然性が高いことを意味する。つまり、確率分布の勾配は画像の特定の特徴が訓練データに即しているかを示す指標となっている。

2.2 拡散モデルと Langevin 方程式

拡散モデルの基礎となる数学的枠組みとして中心的な役割を果たすのが確率微分方程式 (1) であるが、この方程式は、物理学において Brown 運動を記述する手法として 19 世紀末から発展してきた。Brown 運動とは、液体中に懸濁したコロイド粒子が、周囲の流体分子からの絶え間ない衝突によって引き起こされる不規則運動を指す。この微視的な分子衝突の集積効果が、巨視的には確率的な振る舞いとして観測されるのである。

物理学の標準的な文献では、SDE (1) に対応する表現として、以下のような Langevin 方程式が採用されている。

$$\dot{\mathbf{x}}_t = \mathbf{f}(\mathbf{x}_t, t) + \boldsymbol{\xi}_t \quad (5)$$

ここで確率的ノイズ項 $\boldsymbol{\xi}_t$ は、以下の統計的性質を満たすガウスの白色ノイズとして特徴づけられる。

$$\mathbb{E}[\boldsymbol{\xi}_t^i \boldsymbol{\xi}_{t'}^j] = g(t)^2 \delta^{ij} \delta(t - t')$$

この式は、ノイズの各成分が互いに独立であり、異なる時刻間での相関がないことを表現しており、 $\mathbb{E}[\cdot]$ は統計的平均操作を意味する。

式 (5) の物理的解釈をより深めるため、粘性流体中の微粒子運動の古典力学的描像から出発しよう。流体中の微粒子の運動量を \mathbf{p}_t 、位置座標を \mathbf{x}_t とすると、その時間発展は以下の方程式で記述される。

$$\dot{\mathbf{x}}_t = \frac{\mathbf{p}_t}{m}, \quad \dot{\mathbf{p}}_t = -\gamma \dot{\mathbf{x}}_t + \mathbf{F}(\mathbf{x}_t, t) + \boldsymbol{\xi}_t$$

第一式は運動量と速度の基本的関係を表し、第二式は粒子に作用する力の総和を示している。ここで $-\gamma \dot{\mathbf{x}}_t$ は粘性媒質による速度比例型の抵抗力 (Stokes 抵抗)

を表現しており、 γ は摩擦係数である。また $\boldsymbol{\xi}_t$ は熱的ゆらぎに由来する確率的力、 $\mathbf{F}(\mathbf{x}_t, t)$ は保存力などの決定論的な外力成分を表している。この 2 つの方程式から運動量を消去すると、位置の時間発展に関する二階の微分方程式が得られる。

$$m\ddot{\mathbf{x}}_t + \gamma\dot{\mathbf{x}}_t = \mathbf{F}(\mathbf{x}_t, t) + \boldsymbol{\xi}_t$$

慣性項 $m\ddot{\mathbf{x}}_t$ を含むこの方程式は、アンダーダンブ系の Langevin 方程式と呼ばれる。粒子の質量が軽い場合、粘性効果が慣性効果を圧倒的に上回るため、緩和時間のスケールを超えた観測では慣性項の寄与が無視できる。このとき、 $\gamma\dot{\mathbf{x}}_t = \mathbf{F}(\mathbf{x}_t, t) + \boldsymbol{\xi}_t$ という一階の常微分方程式が得られるが、これが通常拡散モデルで用いられる SDE に対応する。またアンダーダンブ系の Langevin 方程式を用いた拡散モデルの拡張¹²⁾もある。

2.3 確率フロー ODE

上述では拡散モデルの生成過程を SDE として定式化した。実際には生成過程におけるノイズレベルは調節することが可能で、決定論的に生成することも可能である。ここではその定式化について述べる。

前節と同様の議論に基づき、Fokker-Planck 方程式を以下のように変形する。

$$\begin{aligned} \partial_t p_t(\mathbf{x}) &= -\nabla \cdot \left[\mathbf{f}(\mathbf{x}, t) p_t(\mathbf{x}) - \frac{g(t)^2}{2} \nabla p_t(\mathbf{x}) \right. \\ &\quad \left. + \mathfrak{h} \frac{g(t)^2}{2} \nabla p_t(\mathbf{x}) - \mathfrak{h} \frac{g(t)^2}{2} \nabla p_t(\mathbf{x}) \right] \\ &= -\nabla \cdot \left[\left(\mathbf{f}(\mathbf{x}, t) - \frac{1+\mathfrak{h}}{2} g(t)^2 \nabla \log p_t(\mathbf{x}) \right) \right. \\ &\quad \left. p_t(\mathbf{x}) + \frac{\mathfrak{h}}{2} g(t)^2 \nabla p_t(\mathbf{x}) \right] \end{aligned}$$

ここで $\mathfrak{h} < \mathbb{R}_{\geq 0}$ は任意のパラメータである。前に行った変形との違いは、このパラメータ \mathfrak{h} の導入にある。ここで

$$\mathbf{F}_{\mathfrak{h}}(\mathbf{x}, t) := \mathbf{f}(\mathbf{x}, t) - \frac{1+\mathfrak{h}}{2} g(t)^2 \nabla \log p_t(\mathbf{x})$$

と定義すると、上記のように変形した Fokker-Planck 方程式に対応する SDE は以下のように表される。

$$d\mathbf{x}_t = \mathbf{F}_{\mathfrak{h}}(\mathbf{x}_t, t) dt + \sqrt{\mathfrak{h}} g(t) d\tilde{\mathbf{w}}_t$$

この式で $\mathfrak{h} = 1$ とすると、式 (4) が再現される。 $\mathfrak{h} = 0$ と選ぶことで、ノイズ項を完全に除去することもできる。

$$\begin{aligned} d\mathbf{x}_t &= \left[\mathbf{f}(\mathbf{x}_t, t) - \frac{1}{2} g(t)^2 \nabla \log p_t(\mathbf{x}) \right] dt \\ &=: \tilde{\mathbf{f}}^{\text{PF}}(\mathbf{x}_t, t) dt \end{aligned} \quad (6)$$

この場合、生成過程は SDE ではなく常微分方程式 (ODE) となる。確率フロー ODE¹⁰⁾と呼ばれるこの手法では、上式においてスコア関数を $s_\theta(\mathbf{x}, t)$ に置き換えた ODE を解くことで画像生成を行う。このように、生成過程におけるノイズレベルは \mathfrak{h} を通じて調整することができる。

どの \mathfrak{h} の値を選んでも、Fokker-Planck 方程式としては等価であることから、この方程式を解いて得られる確率分布 $p_t(\mathbf{x})$ は \mathfrak{h} に依存しない。したがって、上記の SDE 式によって実現される確率分布も \mathfrak{h} に依存しないことに注意が必要である。しかし、学習されたスコア関数を用いる場合、 $s_\theta(\mathbf{x}, t)$ と $\nabla \log p_t(\mathbf{x}_t)$ は厳密には一致しないため、モデルによって生成される分布は \mathfrak{h} に依存する。経験的には、確率フロー ODE による決定論的生成よりも、SDE によるノイズを含む生成の方が最終的な画像の質が高くなることが報告されている¹³⁾。一方で、確率フロー ODE では時間発展が決定論的であるため、データ分布の変数と事前分布の変数（潜在変数）の間に 1 対 1 対応を確立できるという利点がある。また、確率フロー ODE を用いると対数尤度 $\log p_0(\mathbf{x}_0)$ を直接計算することができる。具体的には、 \mathbf{x}_0 の実現確率（の対数）を求めるには、 \mathbf{x}_0 を初期条件として確率フロー ODE を $t = T$ まで解いて軌道 $\{\mathbf{x}_t\}_{t \in [0, T]}$ を得た後、以下の「変数変換公式」により対数尤度を評価することができる：

$$\log p_0(\mathbf{x}_0) = \log p_T(\mathbf{x}_T) + \int_0^T \nabla \cdot \tilde{\mathbf{f}}^{\text{PF}}(\mathbf{x}_t, t) dt$$

3. 経路積分による拡散モデルの定式化

経路積分は、R. P. Feynman によって 1940 年代に導入された理論的枠組みであり、量子力学および場の量子論を数学的に定式化する方法として広く用いられている。この強力な手法は高エネルギー物理学、凝縮系物理学、量子統計力学のみならず、確率過程理論を通じて金融数学や計算科学など多岐にわたる応用分野にも影響を与えている。

3.1 量子力学における経路積分

まずは量子力学の定式化としての経路積分法をレビューする。量子論の前提として、まず古典力学における変分原理から議論を始めよう。点粒子を考えると、古典力学においては、この粒子が実際に辿る軌道は以下で定義される「作用」と呼ばれる汎関数 \mathcal{A} を極小化する経路として特定される。

$$\mathcal{A}[\{\mathbf{x}_t\}_{t \in [0, T]}] = \int_0^T L(\mathbf{x}_t, \dot{\mathbf{x}}_t) dt$$

関数 $L(\mathbf{x}_t, \dot{\mathbf{x}}_t)$ は Lagrangian と呼ばれる。典型的な力学系では、Lagrangian は運動エネルギーと位置エネルギーの差として表現され、点粒子の場合は

$$L(\mathbf{x}_t, \dot{\mathbf{x}}_t) = \frac{m}{2} \|\dot{\mathbf{x}}_t\|^2 - V(\mathbf{x}_t)$$

という形式を取る。「最小作用の原理」として知られるこの変分原理は、実現される物理的軌道において作用 \mathcal{A} が停留値を取ることを要請する。すなわち、任意の軌道変分 $\mathbf{x}_t \mapsto \mathbf{x}_t + \delta \mathbf{x}_t$ に対して、作用の変化が一次のオーダーで消失する条件から、系の運動を支配する Euler-Lagrange 方程式が導出される。

この古典的描像から量子論へと視点を移そう。量子力学の基本方程式は Schrödinger 方程式であり、

$$i\hbar \frac{\partial}{\partial t} \psi(t, \mathbf{x}) = \hat{H} \psi(t, \mathbf{x}) \quad (7)$$

と記述される。ここで \hat{H} は量子系の Hamiltonian であり、点粒子系では $\hat{H} = -\frac{\hbar^2}{2m} \nabla^2 + V(\mathbf{x})$ という形で与えられる。Schrödinger 方程式 (7) の一般解は、初期状態の波動関数 $\psi(0, \mathbf{x})$ を用いて

$$\psi(t, \mathbf{x}) = \int G(t, \mathbf{x} | 0, \mathbf{x}') \psi(0, \mathbf{x}') d\mathbf{x}'$$

と表現できる。ここに現れる関数 $G(t, \mathbf{x} | 0, \mathbf{x}')$ は伝播関数（あるいはグリーン関数）と呼ばれ、初期位置 \mathbf{x}' から時間 t 後に位置 \mathbf{x} へと遷移する量子力学的振幅を表す。この伝播関数は、Feynman の経路積分を用いて

$$G(t, \mathbf{x} | 0, \mathbf{x}') = \int_{(0, \mathbf{x}') \rightarrow (t, \mathbf{x})} [D\mathbf{x}_t] \exp\left(\frac{i}{\hbar} \mathcal{A}[\mathbf{x}_t]\right)$$

という形で表現される。この式の右辺が「Feynman の経路積分」と呼ばれる表式である。始点 $(0, \mathbf{x}')$ から終点 (t, \mathbf{x}) に至るすべての可能な経路に渡って、位相因子 $e^{\frac{i}{\hbar} \mathcal{A}}$ の重み付きで足し上げを行うことを意味する。

この定式化は量子力学の本質を明確に示している。古典力学では粒子は一意的な経路を辿るのに対し、量子力学では粒子はあらゆる可能な経路を「同時に辿る」。各経路には作用 \mathcal{A} に応じた複素位相が付与され、これらの経路の寄与が干渉し合うことで量子的振る舞いが生じるのである。

古典力学と量子力学の接続は、Planck 定数 $\hbar \rightarrow 0$ の極限過程を通じて理解できる。 \hbar が微小な場合、経路 \mathbf{x}_t が微小変化 $\delta \mathbf{x}_t$ を受ただけでも位相因子 $e^{\frac{i}{\hbar} \mathcal{A}}$ は大きく変動する。その結果、古典経路から逸脱した経路の寄与は互いに打ち消し合い、作用 \mathcal{A} が停留値

を取る経路（すなわち古典的軌道）のみが有意な寄与を残す。これにより、 $\hbar \rightarrow 0$ の極限では経路積分から自然に古典力学の最小作用原理が導出される。

古典論に対して量子効果を逐次的に取り入れるために、Planck 定数 \hbar に関するべき展開を考えることができる。この手法は物理学において Wentzel-Kramers-Brillouin (WKB) 近似として知られている。具体的には、粒子の経路を古典軌道 \mathbf{x}_t^{cl} とそこからの量子的揺らぎ $\delta \mathbf{x}_t$ に分解し、 $\mathbf{x}_t = \mathbf{x}_t^{\text{cl}} + \delta \mathbf{x}_t$ として表現し、作用汎関数 $\mathcal{A}[\{\mathbf{x}_t\}_{t \in [0, T]}]$ を古典軌道 \mathbf{x}_t^{cl} の周りでテイラー展開すると、

$$\begin{aligned} \mathcal{A}[\{\mathbf{x}_t\}_{t \in [0, T]}] &= \mathcal{A}[\{\mathbf{x}_t^{\text{cl}}\}_{t \in [0, T]}] \\ &+ \int \frac{\delta \mathcal{A}[\mathbf{x}_t]}{\delta \mathbf{x}_t} \Big|_{\mathbf{x}_t = \mathbf{x}_t^{\text{cl}}} \cdot \delta \mathbf{x}_t dt + \frac{1}{2} \iint \sum_{i,j} \frac{\delta^2 \mathcal{A}[\mathbf{x}_t]}{(\delta \mathbf{x}_t)_i (\delta \mathbf{x}_{t'})_j} (\delta \mathbf{x}_t)_i (\delta \mathbf{x}_{t'})_j dt dt' + \mathcal{O}((\delta \mathbf{x}_t)^3) \end{aligned}$$

という表式が得られる。ここで $(\delta \mathbf{x}_t)_i$ は $\delta \mathbf{x}_t$ の i 成分を表す。展開の第 2 項は \mathbf{x}_t^{cl} が古典運動方程式を満たすためゼロとなる。量子揺らぎ $\delta \mathbf{x}_t$ に対して 2 次までの項を考慮すると、 $\delta \mathbf{x}_t$ に関する経路積分はガウス型となり、伝搬関数は

$$G(t, \mathbf{x} | 0, \mathbf{x}') = e^{\frac{i}{\hbar} \mathcal{A}[\mathbf{x}_t^{\text{cl}}]} \det \left(\frac{\delta^2 \mathcal{A}}{(\delta \mathbf{x}_t)_i (\delta \mathbf{x}_{t'})_j} \Big|_{\mathbf{x}_t = \mathbf{x}_t^{\text{cl}}} \right)^{-1/2}$$

と半古典的に近似される。第 1 因子は古典作用に基づく急速振動項、第 2 因子はガウス積分から生じる揺らぎの寄与を表している。

経路積分形式は、点粒子系のみならず、量子電磁力学や量子色力学などのゲージ理論などを含む量子系の記述方法として、不可欠の理論的枠組みとなっている。また、時間変数の解析接続 $t \rightarrow -i\tau$ (ユークリッド化と呼ばれる操作) を適用すると、

$$\int [D\mathbf{x}_t] e^{\frac{i}{\hbar} \mathcal{A}} \dots$$

という振動積分形式は、

$$\int [D\mathbf{x}_t] e^{-\frac{1}{\hbar} \mathcal{A}} \dots$$

という減衰積分形式へと変換される。この変換により、量子力学の経路積分は統計力学における分配関数の表現と形式的に等価となる。すなわち、経路積分は量子力学と統計物理学を統一的に扱うための数学的橋渡しとしても機能するのである。

3.2 経路積分による拡散モデル

本節では、拡散モデルを経路積分の枠組みを用いて定式化する方法⁸⁾について述べる。拡散モデルの順過程は確率微分方程式 (Langevin 方程式) に従う確率過程であり、このような過程は経路積分形式による表現が可能である。実際、 \mathbf{x}_t が Langevin 方程式 $\dot{\mathbf{x}}_t = \mathbf{f}(\mathbf{x}_t, t) + \boldsymbol{\xi}_t$ に従うとすると、経路 $\{\mathbf{x}_t\}_{t \in [0, T]}$ に依存する任意の物理量 $\mathcal{O}(\{\mathbf{x}_t\})$ の期待値は、経路積分を用いて以下のように表現できる：

$$\mathbb{E}[\mathcal{O}(\{\mathbf{x}_t\})] = \int [D\mathbf{x}_t] [D\boldsymbol{\xi}_t] \mathcal{O}(\{\mathbf{x}_t\}) e^{-\int \frac{\|\boldsymbol{\xi}_t\|^2}{2g(t)^2} dt} p_0(\mathbf{x}_0) \prod_t \delta(\mathbf{x}_t - \mathbf{x}_t^{\text{sol}})$$

ここで $p_0(\mathbf{x}_0)$ は初期状態の確率分布を表し、 $\delta(\cdot)$ は Dirac のデルタ関数であり、 $\mathbf{x}_t^{\text{sol}}$ は与えられたノイズ実現 $\{\boldsymbol{\xi}_t\}_{t \in [0, T]}$ に対する「運動方程式」の解である。指数関数内の項 $e^{-\int \frac{\|\boldsymbol{\xi}_t\|^2}{2g(t)^2} dt}$ はノイズ $\boldsymbol{\xi}_t$ のガウス分布重みを表し、デルタ関数 $\delta(\mathbf{x}_t - \mathbf{x}_t^{\text{sol}})$ は経路 \mathbf{x}_t が Langevin 方程式 $\dot{\mathbf{x}}_t = \mathbf{f}(\mathbf{x}_t, t) + \boldsymbol{\xi}_t$ の解であるという拘束条件を課している。

この経路積分表示をより扱いやすい形式に変換するために、デルタ関数の変数変換公式を適用する。

$$\delta(\mathbf{x}_t - \mathbf{x}_t^{\text{sol}}) = \left| \det \frac{\delta \text{EOM}_t}{\delta \mathbf{x}_{t'}} \right| \delta(\text{EOM}_t)$$

ここで $\text{EOM}_t := \dot{\mathbf{x}}_t - \mathbf{f}(\mathbf{x}_t, t) - \boldsymbol{\xi}_t$ と定義した。この式の Jacobian の寄与を以下のように表記する。

$$\prod_t \left| \det \frac{\delta \text{EOM}_t}{\delta \mathbf{x}_{t'}} \right| = e^{-J}$$

次に、ノイズ変数 $\boldsymbol{\xi}_t$ に関する積分を実行する。デルタ関数により $\boldsymbol{\xi}_t = -\mathbf{f}(\mathbf{x}_t, t) + \dot{\mathbf{x}}_t$ と表せるため、期待値 $\mathbb{E}[\mathcal{O}(\{\mathbf{x}_t\})]$ はさらに以下のように簡略化される。

$$\mathbb{E}[\mathcal{O}(\{\mathbf{x}_t\})] = \int [D\mathbf{x}_t] \mathcal{O}(\{\mathbf{x}_t\}) e^{-\mathcal{A}} p_0(\mathbf{x}_0)$$

ここで作用 \mathcal{A} は $\mathcal{A} = \int_0^T L(\dot{\mathbf{x}}_t, \mathbf{x}_t) dt + J$ とし、Lagrangian $L(\dot{\mathbf{x}}_t, \mathbf{x}_t)$ は以下で与えられる¹⁴⁾：

$$L(\dot{\mathbf{x}}_t, \mathbf{x}_t) := \frac{1}{2g(t)^2} \|\dot{\mathbf{x}}_t - \mathbf{f}(\mathbf{x}_t, t)\|^2$$

この作用の第 1 項を展開した際に現れる

$$- \int \frac{1}{g(t)^2} \mathbf{f}(\mathbf{x}_t, t) \cdot d\mathbf{x}_t$$

という項は、実は離散化スキームに依存することに注意が必要である。 $\mathbf{f}_t = \mathbf{f}(\mathbf{x}_t, t)$ と略記すると、伊藤

型およびストラトノビッチ型の積はそれぞれ以下のよう
に定義される。

$$\begin{aligned} \mathbf{f}(\mathbf{x}_t, t) \cdot d\mathbf{x}_t|_{\text{Itô}} &:= \mathbf{f}_t \cdot (\mathbf{x}_{t+dt} - \mathbf{x}_t), \\ \mathbf{f}(\mathbf{x}_t, t) \cdot d\mathbf{x}_t|_{\text{Stratonovich}} &:= \frac{\mathbf{f}_t + \mathbf{f}_{t+dt}}{2} \cdot (\mathbf{x}_{t+dt} - \mathbf{x}_t) \end{aligned}$$

この違いは、 $d\mathbf{x}_t$ に乗じる \mathbf{f}_t の評価時点にある。伊藤型では時刻 t での値を用い、ストラトノビッチ型では時刻 t と $t+dt$ における値の平均を採用する。Jacobian に由来する項 J も以下のように離散化スキームに依存する：

$$J = \begin{cases} 0 & (\text{Itô}) \\ \int_0^T \frac{1}{2} \nabla \cdot \mathbf{f}(\mathbf{x}_t, t) dt & (\text{Stratonovich}) \end{cases}$$

作用 $\mathcal{A} = \int_0^T L(\dot{\mathbf{x}}_t, \mathbf{x}_t) dt + J$ 個々の項は離散化スキームに依存するが、作用全体はスキーム不変であることに留意されたい。

3.3 逆過程の導出

経路積分表示は拡散モデルの多様な側面を分析する強力なツールとなる。その一例として、経路積分を用いた逆過程の導出を示そう。 $\mathbb{E}[\mathcal{O}(\{\mathbf{x}_t\})]$ の経路積分表示を以下のように変形する。

$$\begin{aligned} & \int [D\mathbf{x}_t] \mathcal{O}(\{\mathbf{x}_t\}) e^{-\mathcal{A}} p_0(\mathbf{x}_0) \\ &= \int [D\mathbf{x}_t] \mathcal{O}(\{\mathbf{x}_t\}) e^{-\mathcal{A} + \log p_0(\mathbf{x}_0) - \log p_T(\mathbf{x}_T)} p_T(\mathbf{x}_T) \\ &=: \int [D\mathbf{x}_t] \mathcal{O}(\{\mathbf{x}_t\}) e^{-\tilde{\mathcal{A}}} p_T(\mathbf{x}_T) \end{aligned}$$

ここで新たな作用 $\tilde{\mathcal{A}} := \mathcal{A} + \log p_0(\mathbf{x}_0) - \log p_T(\mathbf{x}_T)$ を導入した。さらに

$$\log p_T(\mathbf{x}_T) - \log p_0(\mathbf{x}_0) = \int_0^T \frac{d}{dt} p_t(\mathbf{x}_t) dt$$

と表現し、確率分布の時間微分を Fokker-Planck 方程式と伊藤の公式を用いて変形すると、作用 $\tilde{\mathcal{A}}$ は以下の形式に書き換えられる：

$$\tilde{\mathcal{A}} = \int_0^T \left[\frac{1}{2g(t)^2} \|\dot{\mathbf{x}}_t - \tilde{\mathbf{f}}(\mathbf{x}_t, t)\|^2 - \nabla \cdot \tilde{\mathbf{f}}(\mathbf{x}_t, t) \right] dt$$

ここで新たなドリフト項

$$\tilde{\mathbf{f}}(\mathbf{x}_t, t) := \mathbf{f}(\mathbf{x}_t, t) - g(t)^2 \nabla \log p_t(\mathbf{x}_t)$$

を定義した。 $\tilde{\mathcal{A}}$ の第 2 項を第 1 項に組み込んで Jacobian 項が出ない形で作用を書くと、 $\tilde{\mathbf{f}}(\mathbf{x}_t, t) \cdot d\mathbf{x}_t$ の部分が

$$\tilde{\mathbf{f}}(\mathbf{x}_t, t) \cdot d\mathbf{x}_t|_{\text{Reverse Itô}} := \tilde{\mathbf{f}}_{t+dt} \cdot (\mathbf{x}_{t+dt} - \mathbf{x}_t)$$

という形の積となることがわかる。この表現は時間反転した際に伊藤型になることから「リバース伊藤型」と呼んでいる。このような形となることは、作用 $\tilde{\mathcal{A}}$ が逆過程を記述するものとして自然であることを示唆している。

3.4 確率フロー ODE と古典極限の関係性

拡散モデルによる生成プロセスにおいては、ノイズレベルは調節可能であり、ノイズを含む確率的生成とノイズを含まない確率フロー ODE(5) による生成をパラメータ \hbar により連続的に繋げることを 2.3 節で見た。ここでは経路積分の枠組みでこの現象を考察してみよう。ノイズレベル \hbar におけるモデル生成過程での $\mathcal{O}(\{\mathbf{x}_t\})$ の期待値は以下のように表現できる：

$$\mathbb{E}[\mathcal{O}(\{\mathbf{x}_t\})] = \int [D\mathbf{x}_t] \mathcal{O}(\{\mathbf{x}_t\}) e^{-\tilde{\mathcal{A}}_\theta^\hbar / \hbar} q_T(\mathbf{x}_T)$$

ここでパラメータ \hbar に依存するモデルの逆過程の作用を

$$\tilde{\mathcal{A}}_\theta^\hbar := \int_0^T \tilde{L}_\theta^\hbar(\dot{\mathbf{x}}_t, \mathbf{x}_t) dt$$

と定義し、 $\tilde{L}_\theta^\hbar(\dot{\mathbf{x}}_t, \mathbf{x}_t)$ は次のように与えられる：

$$\tilde{L}_\theta^\hbar(\dot{\mathbf{x}}_t, \mathbf{x}_t) := \frac{\left\| \dot{\mathbf{x}}_t - \mathbf{f}(\mathbf{x}_t, t) + \frac{1+\hbar}{2} g(t)^2 \mathbf{s}_\theta(\mathbf{x}_t, t) \right\|^2}{2g(t)^2}$$

量子力学における経路積分が次の形式で表されることを思い出そう：

$$\int [D\mathbf{x}_t] \mathcal{O}(\{\mathbf{x}_t\}) e^{i\mathcal{A}/\hbar} \dots$$

これと拡散モデルの経路積分表示を比較すると、構造的な類似性が明らかになる。つまり、拡散モデルにおいて生成過程のノイズレベルを特徴づけるパラメータ \hbar は、量子系の経路積分における Planck 定数 \hbar に対応する役割を担っている：

$$e^{i\mathcal{A}/\hbar} \leftrightarrow e^{-\tilde{\mathcal{A}}_\theta^\hbar / \hbar}$$

このような対応関係に基づくと、量子力学の「古典極限」の拡散モデルにおける対応物を考えることができる。 $\hbar \rightarrow 0$ の極限で、経路の実現確率は以下のように書くことができる：

$$\begin{aligned} P(\{\mathbf{x}_t\}_{t \in [0, T]}) &= [D\mathbf{x}_t] e^{-\tilde{\mathcal{A}}_\theta^\hbar / \hbar} q_T(\mathbf{x}_T) \\ &\rightarrow [D\mathbf{x}_t] \prod_t \delta(\mathbf{x}_t - \mathbf{x}_t^{\text{PF}}) q_T(\mathbf{x}_T) \end{aligned}$$

ここで \mathbf{x}_t^{PF} は確率フロー ODE $\dot{\mathbf{x}}_t = \tilde{\mathbf{f}}^{\text{PF}}(\mathbf{x}_t, t)$ の解である。したがって、 $\hbar \rightarrow 0$ の極限では、確率フロー ODE によって決定される決定論的な経路が選択される。この状況は、量子力学の経路積分表現において $\hbar \rightarrow 0$ とすると古典論における運動方程式が導出されることと類似している。

$\hbar \rightarrow 0$ の極限に加え、 \hbar による級数展開も行うことができる。すなわち、WKB 近似の類似概念を適用できる。これにより、 $\hbar \neq 0$ の場合でも、確率フロー ODE を用いた場合と同様に、モデルの対数尤度をノイズレベル \hbar のべき展開の形で評価することが可能となる⁸⁾。

4. 終わりに

本稿では、機械学習における生成モデルの新たな潮流である拡散モデルと物理学の概念との接点について論じてきた。拡散モデルの中核となる確率微分方程式は Langevin 方程式として物理学で長らく用いられてきたものであり、さらに量子力学の定式化に用いられる経路積分の方法によって拡散モデルを統一的に記述できることを示した。

特に注目すべき点は、拡散モデルの生成過程における確率的・決定論的サンプリング間を補間するパラメータ \hbar が、量子力学のプランク定数 \hbar に類似した役割を果たしていることである。 $\hbar \rightarrow 0$ の極限では確率フロー ODE による決定論的な経路が選択され、これは量子力学における古典極限と構造的に同一である。この対応関係に基づき、WKB 近似の類似概念を拡散モデルに適用することで、異なるサンプリング手法の性能差を理論的に解析することが可能となる。

また、拡散モデルの発展に物理学からのインスピレーションがさらに持ち込まれている。慣性項を導入したアンダーダンブ系の Langevin 方程式を採用したモデル¹²⁾や、生成サンプル間に相互作用を導入することで多様性を向上させる手法¹⁵⁾など、物理学の知見を活かした改良が報告されている。さらに、拡散モデルの生成過程をある種の相転移として理解しようとする試み¹⁶⁾もある。

拡散モデルは単なる実用的な生成技術に留まらず、物理学や統計力学といった広範な理論領域と深く結びついている。この結びつきを探究することで、生成モデルの本質的理解が一層深まり、より洗練されたモデルの開発につながることが期待される。同時に、拡散モデルの研究を通じて得られた知見が物理学にフィードバックされ、両分野の相互発展が促進されるであ

ろう。

物理学と機械学習の境界領域は依然として多くの未探索課題を内包しており、魅力的な研究分野となっている。この分野の発展が、生成 AI の次なる革新をもたらすとともに、物理学の新たな適用領域を開拓していくことを期待したい。

参考文献

- 1) Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., Ganguli, S. (2015): Deep unsupervised learning using nonequilibrium thermodynamics, In: International Conference on Machine Learning.
- 2) Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y. (2020): Generative adversarial networks, Commun. ACM, Vol.63(11), 139.
- 3) Kingma, D. P., Welling, M. (2014): Auto-encoding variational bayes, Proc. 2nd Int. Conf. Learning Representations (ICLR) 2014.
- 4) Rezende, D. J., Mohamed, S., Wierstra, D. (2014): Stochastic backpropagation and approximate inference in deep generative models, Proc. 31st Int. Conf. Machine Learning, 2014, Vol.32, p.1278.
- 5) Jacob, A., Johnson, D. D., Ho, J., Tarlow, D., van den Berg, R. (2021): Structured denoising diffusion models in discrete state-spaces. In: Advances in Neural Information Processing Systems (NeurIPS), 2021.
- 6) Li, X. L., Mohit, I., Yejin, W., Jing, Y. (2022): Diffusion-LM improves controllable text generation, arXiv preprint arXiv:2203.09753.
- 7) Lou, A., Meng, C., Ermon, S. (2023): Discrete diffusion modeling by estimating the ratios of the data distribution, arXiv preprint arXiv:2310.16834.
- 8) Hirono, Y., Tanaka A., Fukushima, K. (2024): Understanding diffusion models by feynman's path integral, in Proceedings of the 41st International Conference on Machine Learning, Proc. Mach. Learn. Res., Vol.235 (PMLR, 2024), pp.18324–18351.
- 9) Oksendal, B. (2013): Stochastic differential equations: An introduction with applications, Springer Science & Business Media.
- 10) Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., Poole, B. (2021): Score-based generative modeling through stochastic differential equations, 9th International Conference on Learning Representations, 2021.
- 11) Ho, J., Jain, A., Abbeel, P. (2020): Denoising

- diffusion probabilistic models, NeurIPS, 2020.
- 12) Dockhorn, T., Vahdat, A., Kreis, K. (2022): Score-based generative modeling with critically-damped langevin diffusion, ICLR, 2022.
- 13) Karras, T., Aittala, M., Aila, T., Laine, S. (2022): Elucidating the design space of diffusion-based generative models, NeurIPS, 2022.
- 14) Onsager, L., Machlup, S. (1953): Fluctuations and irreversible processes, Phys. Rev., Vol.91, 1505.
- 15) Corso, G., Xu, Y., De Bortoli, V., Barzilay, R., Jaakkola, T. (2023): Particle guidance: non-iid diverse sampling with diffusion models. In: International Conference on Learning Representations.
- 16) Biroli, G., Bonnaire, T., de Bortoli, V., Mézard, M. (2024): Dynamical regimes of diffusion models, Nature Commun., Vol.15, Article number: 9957.