

拡散モデルやその周辺と非平衡熱力学との接点

伊藤創祐

1. 公開にあたって

この原稿は「数理科学」2024年10月号 No.736, サイエンス社, 2024."中の記事"伊藤創祐, 拡散モデルやその周辺と非平衡熱力学との接点"の原稿バージョンになります。発売から1ヶ月以上経過してからサイエンス社の公開の条件に則って記事を公開しています。公開にあたって引用情報などの一部修正を行っています。

2. 生成 AI の数理と非平衡熱力学の接点

近年, 生成 AI は社会的な注目を集めている。特に2022年ごろに発表された様々な画像生成 AI の出現は, 生成 AI が社会的に大きく注目を集めるきっかけとなり, AI における創造性の有無に関する議論が交わされるようになった。特に画像生成 AI の中でもテキストから画像を生成する text-to-image モデルの出現は, 絵を専門としない人々でも文章をいじるだけで巧みな画像を生成できることから, 創作や著作権やフェアユースとは何かを社会的に問いかける事態にまで発展しつつある。

本記事ではこのような社会的な生成 AI の善悪について議論するつもりはない。しかしながらこのような社会的な問題に発展しう理由として, 現状の生成 AI が生成する画像のクオリティが非常に高いからという点が挙げられると思う。特に Stability AI によって2022年に公開された Stable

Diffusion と呼ばれる画像生成 AI は, 革命的な画像生成 AI として社会的に受け止められたと思う。

この Stable Diffusion の核となる画像生成手法は, 現在**拡散モデル**¹⁾と呼ばれる手法の一種として知られている。この拡散モデルは拡散のダイナミクスによって画像構造などのデータ構造を学習し, さらにその学習したデータ構造に似た構造をダイナミクスを用いて再構成することによって達成される。この拡散モデルは2020年前後に画像生成タスクにおいて最良の性能を示した^{2, 3)}ことや, 画像データのみならず動画データや音楽データなどにも応用可能なことがわかってきたことから, 現在に至るまで注目されている手法である。

この拡散モデルは拡散の名の通り, 物理現象である拡散現象を用いている。さらにこの学習の手法は非平衡熱力学の理論から想起されている。例えば2015年の拡散モデルの原論文¹⁾のタイトルは"Deep unsupervised learning using nonequilibrium thermodynamics"であり, 極めて明確に非平衡熱力学すなわち"nonequilibrium thermodynamics"を使っていると書かれている。

しかしながら現在の拡散モデルの手法は改良を重ねた結果, 非平衡熱力学の理解が不要な形になっている。そのため現在拡散モデルと非平衡熱力学の関係に関する知見は, 研究者レベルでも浸透していないと感じている。そのため本記事では, 拡散モデルの発想の元となった非平衡熱力学の解説を行い, それがいかに拡散モデルの手法に繋がって

いったかを解説することで、現在の生成 AI や拡散モデルの知見を広げる助けになれば幸いである。

3. 拡散系のダイナミクスと非平衡熱力学

拡散現象のダイナミクスと非平衡熱力学の研究は、古くは 1905 年の Einstein によるブラウン運動の理論に遡る。この理論に基づいて 1908 年に Langevin はブラウン運動の拡散ダイナミクスを記述する **Langevin 方程式** を構築した。特に緩和時間が十分早く慣性項が無視できる極限 (over-damped 極限) で、Langevin 方程式は次のように記述できる。

$$\dot{\mathbf{x}}_t = \mu \mathbf{F}_t(\mathbf{x}_t) + \sqrt{2D_t} \boldsymbol{\xi}_t. \quad (1)$$

ここで、 $\mathbf{x}_t (\in \mathbb{R}^d)$ と $\dot{\mathbf{x}}_t (\in \mathbb{R}^d)$ は時刻 t における d -次元ユークリッド空間上のブラウン粒子の位置と速度、 μ は易動度、 $\mathbf{F}_t(\mathbf{x}_t)$ は時刻 t 、位置 $\mathbf{x}_t$ でのブラウン粒子に加わる外力、 $D_t (> 0)$ は時刻 t における拡散係数、 $\boldsymbol{\xi}_t = (\xi_{1;t}, \dots, \xi_{d;t})$ は白色ガウスノイズで表されるブラウン粒子に加わるランダム力の項であり、期待値 $\langle \dots \rangle$ を用いてその統計性は $\langle \xi_{i;t} \rangle = 0$ 、 $\langle \xi_{i;t} \xi_{j;t'} \rangle = \delta(t - t') \delta_{ij}$ で記述される。ただし $\delta(t - t')$ はデルタ関数、 δ_{ij} は Kronecker のデルタを表す。この Langevin 方程式は確率的に振る舞う量 $\boldsymbol{\xi}_t$ を含んだ微分方程式であるため**確率微分方程式**の一種とみなせる。Einstein のブラウン運動の業績として着目すべき点は、この Langevin 方程式における拡散係数 D_t を温度 T_t の関数として $D_t = \mu k_B T_t$ とした Einstein の関係式を導出したことにある。ただし k_B はボルツマン定数である。よって Einstein の業績は熱力学的な関数である温度と、確率的な拡散のダイナミクスを結びつけた初期の例とみなすことができるだろう。

拡散現象において、粒子のランダムな挙動を記述せずに粒子の分布の時間発展を考えることもある。 $P_t(\mathbf{x})$ は時刻 t での位置 $\mathbf{x} (\in \mathbb{R}^d)$ の確率分布とすると、Langevin 方程式に対応する $P_t(\mathbf{x})$ の時間発展は次の **Fokker-Planck 方程式** で記述される。

$$\partial_t P_t(\mathbf{x}) = -\nabla \cdot (\boldsymbol{\nu}_t(\mathbf{x}) P_t(\mathbf{x})), \quad (2)$$

$$\boldsymbol{\nu}_t(\mathbf{x}) = \mu \mathbf{F}_t(\mathbf{x}) - D_t \nabla \ln P_t(\mathbf{x}), \quad (3)$$

ここで式 (2) は連続の式になっており、 $\boldsymbol{\nu}_t(\mathbf{x})$ は速度場を表す量である。また式 (3) の速度場における $\mu \mathbf{F}_t(\mathbf{x})$ の項は外力によるドリフトの寄与を意味し、また $D_t \nabla \ln P_t(\mathbf{x})$ の項は拡散による寄与を表している。

この拡散現象は非平衡現象であり、拡散現象を記述する Langevin 方程式や Fokker-Planck 方程式における非平衡熱力学を議論する研究が近年**ゆらぎの熱力学**⁴⁾という名の下で行われている。このゆらぎの熱力学において考察の中心となる量が、熱力学的な散逸の指標である**エントロピー生成率** σ_t という量である。いまボルツマン定数を $k_B = 1$ と規格化した場合、エントロピー生成率は

$$\sigma_t = \frac{1}{D_t} \int d\mathbf{x} \|\boldsymbol{\nu}_t(\mathbf{x})\|^2 P_t(\mathbf{x}), \quad (4)$$

で与えられる。この量は Einstein の関係式 $D_t = \mu T_t$ と式 (3) の表現を用いることで、

$$\begin{aligned} \sigma_t = & \frac{1}{T_t} \int d\mathbf{x} \mathbf{F}_t(\mathbf{x}) \cdot \boldsymbol{\nu}_t(\mathbf{x}) P_t(\mathbf{x}) \\ & - \int d\mathbf{x} \nabla \ln P_t(\mathbf{x}) \cdot \boldsymbol{\nu}_t(\mathbf{x}) P_t(\mathbf{x}), \end{aligned} \quad (5)$$

と表される。ここで外力 $\mathbf{F}_t(\mathbf{x})$ によって単位時間あたりに散逸する熱を $-\dot{Q}(t) = \int d\mathbf{x} \mathbf{F}_t(\mathbf{x}) \cdot \boldsymbol{\nu}_t(\mathbf{x}) P_t(\mathbf{x})$ と定義し、また第二項 $-\int d\mathbf{x} \nabla \ln P_t(\mathbf{x}) \cdot \boldsymbol{\nu}_t(\mathbf{x}) P_t(\mathbf{x})$ を部分積分と連続の式 (2)、確率の規格化 ($\int d\mathbf{x} P_t(\mathbf{x}) = 1$) を用いて、微分エントロピーの時間変化 $\dot{S}_t = -\partial_t \int d\mathbf{x} P_t(\mathbf{x}) \ln P_t(\mathbf{x}) = -\int d\mathbf{x} (\partial_t P_t(\mathbf{x})) \ln P_t(\mathbf{x})$ とみなすと、エントロピー生成率は

$$\sigma_t = -\frac{\dot{Q}_t}{T_t} + \dot{S}_t, \quad (6)$$

と記述できる。よってこのエントロピー生成率の定義から自明に得られる非負性 $\sigma_t \geq 0$ は、 $\dot{S}_t \geq \dot{Q}_t/T_t$ と書き直せる。この非負性は**熱力学第二法則**の一種である (単位時間あたりの) クラウジウスの不等式と同じ形をしており、よってこの非負性 $\sigma_t \geq 0$ は拡散系における熱力学第二法則とみなせる。またエントロピー生成率を有限時間

区間 $[0, \tau]$ で時間積分した量 $\Sigma_{[0, \tau]} = \int_0^\tau dt \sigma_t$ を **エントロピー生成**といい、その非負性 $\Sigma_{[0, \tau]} \geq 0$ も熱力学第二法則に相当する。

4. ゆらぎの定理

Fokker–Planck 方程式の速度場から導入されたエントロピー生成率 σ_t もしくはエントロピー生成 $\Sigma_{[0, \tau]}$ は、Langevin 方程式の確率的な挙動から議論することもできる。特に確率的な挙動からエントロピー生成率を定式化する方法は**ゆらぎの定理**⁵⁾と呼ばれる定理の形で議論されている。このゆらぎの定理は、詳細ゆらぎの定理、積分ゆらぎの定理及び Jarzynski 等式⁶⁾などさまざまな表現が知られている。このゆらぎの定理による表現は、拡散モデルの出発点ともなるような表現であり、実際に拡散モデルの原論文¹⁾では同様の計算とともに Jarzynski 等式の論文⁶⁾が引用されている。今回は式 (1) で記述される場合におけるゆらぎの定理の主張について紹介したい。

ゆらぎの定理とは、Langevin 方程式などで記述される確率的な状態の変化について、ある時間順方向にとった経路の達成確率と時間逆方向にとった経路の達成確率の違いをエントロピー生成と結びつけるものである。まずは簡単のため、経路ではなく瞬間的な1回の遷移に対して議論しよう。今 Langevin 方程式における時刻 t での状態が微小時間 $dt (> 0)$ たった後に $\mathbf{x}_t$ から $\mathbf{x}_{t+dt}$ に変化する状況を考える。この状態変化が起きる遷移確率は ξ_t のガウス性から

$$T_t(\mathbf{x}_{t+dt}|\mathbf{x}_t) = \frac{e^{-\frac{\|\mathbf{x}_{t+dt} - \mathbf{x}_t - \mu \mathbf{F}_t(\mathbf{x}_t)dt\|^2}{4D_t dt}}}{(4\pi D_t dt)^{d/2}}, \quad (7)$$

与えられる。よって分布 $P_t(\mathbf{x}_t)$ から出発して時刻 t で時刻遷移 ($\mathbf{x}_t \rightarrow \mathbf{x}_{t+dt}$) が起こるという同時確率分布は

$$\mathbb{P}_F(\mathbf{x}_t, \mathbf{x}_{t+dt}) = T_t(\mathbf{x}_{t+dt}|\mathbf{x}_t)P_t(\mathbf{x}_t), \quad (8)$$

与えられる。また分布 $P_{t+dt}(\mathbf{x}_{t+dt})$ から出発して時刻 t での遷移確率 T_t を用いて $\mathbf{x}_{t+dt}$ から $\mathbf{x}_t$

に状態が変化する状況を考えよう。この状況での同時確率分布は

$$\mathbb{P}_B(\mathbf{x}_t, \mathbf{x}_{t+dt}) = T_t(\mathbf{x}_t|\mathbf{x}_{t+dt})P_{t+dt}(\mathbf{x}_{t+dt}), \quad (9)$$

与えられる。

この二つの同時確率分布 $\mathbb{P}_F, \mathbb{P}_B$ の違いから、エントロピー生成を考えることができる。この分布の違いは、分布間の違いを定量化する指標である

Kullback-Leibler ダイバージェンス

$$\begin{aligned} D_{\text{KL}}(\mathbb{P}_F||\mathbb{P}_B) \\ = \int d\mathbf{x}_t \int d\mathbf{x}_{t+dt} \mathbb{P}_F(\mathbf{x}_t, \mathbf{x}_{t+dt}) \ln \frac{\mathbb{P}_F(\mathbf{x}_t, \mathbf{x}_{t+dt})}{\mathbb{P}_B(\mathbf{x}_t, \mathbf{x}_{t+dt})}, \end{aligned} \quad (10)$$

を用いて記述され、エントロピー生成との間に以下の式が成り立つ。

$$\Sigma_{[t, t+dt]} = D_{\text{KL}}(\mathbb{P}_F||\mathbb{P}_B) + O((dt)^2). \quad (11)$$

ただし、 $O((dt)^2)$ はランダウの O 記号であり、以降はこの $O((dt)^2)$ の項は十分に微小な量として無視して議論を行うこととする。さらに確率的なエントロピー生成を

$$\hat{\Sigma}_{[t, t+dt]}(\mathbf{x}_t, \mathbf{x}_{t+dt}) = \ln \frac{\mathbb{P}_F(\mathbf{x}_t, \mathbf{x}_{t+dt})}{\mathbb{P}_B(\mathbf{x}_t, \mathbf{x}_{t+dt})} \quad (12)$$

と定義しよう。確率 $\mathbb{P}_F$ による期待値を $\langle \cdots \rangle_{\mathbb{P}_F} = \int d\mathbf{x}_t \int d\mathbf{x}_{t+dt} \mathbb{P}_F(\mathbf{x}_t, \mathbf{x}_{t+dt}) \cdots$ と定義すると、式 (11) は $\langle \hat{\Sigma}_{[t, t+dt]} \rangle_{\mathbb{P}_F} = \Sigma_{[t, t+dt]}$ を意味するため、この確率的なエントロピー生成の期待値はエントロピー生成となる。

ゆらぎの定理は確率的なエントロピー生成の定義の式 (12) を変形することで得られる。まず次の形の式

$$\frac{\mathbb{P}_F(\mathbf{x}_t, \mathbf{x}_{t+dt})}{\mathbb{P}_B(\mathbf{x}_t, \mathbf{x}_{t+dt})} = e^{\hat{\Sigma}_{[t, t+dt]}(\mathbf{x}_t, \mathbf{x}_{t+dt})} \quad (13)$$

を**詳細ゆらぎの定理**という。これは $\mathbb{P}_F, \mathbb{P}_B$ の比が確率的なエントロピー生成と結びつくという結果である。また式 (12) を変形することで得られる $\mathbb{P}_F(\mathbf{x}_t, \mathbf{x}_{t+dt})e^{-\hat{\Sigma}_{[t, t+dt]}(\mathbf{x}_t, \mathbf{x}_{t+dt})} = \mathbb{P}_B(\mathbf{x}_t, \mathbf{x}_{t+dt})$ という式に対して、両辺を積分す

ることで

$$\langle e^{-\hat{\Sigma}_{[t,t+dt]}} \rangle_{\mathbb{P}_F} = 1 \quad (14)$$

が, $\mathbb{P}_B$ の規格化条件から得られる. この結果を**積分揺らぎの定理**という. Jarzynski 等式はこの積分揺らぎの定理の特殊な場合に相当する.

この議論は有限時間区間 $[0, \tau]$ に対しても拡張できる. ただし拡張には $\tau = Ndt$ としたときの経路 $\Gamma = (\mathbf{x}_0, \mathbf{x}_{dt}, \mathbf{x}_{2dt}, \dots, \mathbf{x}_{Ndt})$ を考え, この経路に対する $P_0(\mathbf{x}_0)$ から出発したときの確率を

$$\mathcal{P}_F(\Gamma) = \prod_{n=1}^N T_{(n-1)dt}(\mathbf{x}_{ndt} | \mathbf{x}_{(n-1)dt}) P_0(\mathbf{x}_0), \quad (15)$$

とする. また $P_\tau(\mathbf{x}_\tau)$ から出発して時間逆方向の経路が達成される確率を

$$\mathcal{P}_B(\Gamma) = \prod_{n=1}^N T_{(n-1)dt}(\mathbf{x}_{(n-1)dt} | \mathbf{x}_{ndt}) P_\tau(\mathbf{x}_\tau), \quad (16)$$

とする. このとき, 同様に確率的なエントロピー生成を $\hat{\Sigma}_{[0,\tau]}(\Gamma) = \ln[\mathcal{P}_F(\Gamma)/\mathcal{P}_B(\Gamma)]$ と定義すると, $\hat{\Sigma}_{[0,\tau]}(\Gamma) = \sum_{n=1}^N \hat{\Sigma}_{[(n-1)dt, ndt]}(\mathbf{x}_{(n-1)dt}, \mathbf{x}_{ndt})$ という式が成立する. よって, $\langle \dots \rangle_{\mathcal{P}_F} = \int d\Gamma \mathcal{P}_F(\Gamma) \dots$ という期待値に対して同様に $\langle \hat{\Sigma}_{[0,\tau]} \rangle_{\mathcal{P}_F} = \Sigma_{[0,\tau]}$ かつ, 積分揺らぎの定理 $\langle e^{-\hat{\Sigma}_{[0,\tau]}} \rangle_{\mathcal{P}_F} = 1$ が成立する. Kullback-Leibler ダイバージェンスを用いると, $\langle \hat{\Sigma}_{[0,\tau]} \rangle_{\mathcal{P}_F} = \Sigma_{[0,\tau]}$ は $\Sigma_{[0,\tau]} = D_{\text{KL}}(\mathcal{P}_F || \mathcal{P}_B)$ とも記述可能である. すなわちエントロピー生成率は時間順方向の経路と時間逆方向の経路がどれくらい確率的に異なるかを定量化するものになっており, 拡散のダイナミクスがどの程度不可逆に起こるかの指標とも言える (図1). このようなゆらぎの定理における, 時間順方向の経路と時間逆方向の経路を用いるという非平衡熱力学の考え方に恐らくインスパイアされたのが, 拡散モデルと呼ばれる生成モデルの手法である.

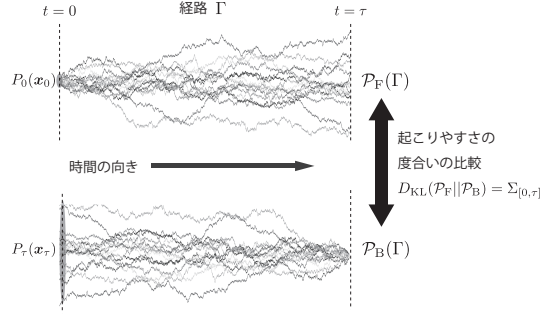


図1 ゆらぎの定理に基づくエントロピー生成の概念図. 拡散のダイナミクスの経路が時間順方向と時間逆方向でどれだけ確率的に起きやすい/起きにくいかの違いを定量化している.

5. 拡散モデルの考え方

ここで一旦非平衡熱力学の話から離れて, 生成AIの手法である拡散モデルの考え方を説明しよう. 拡散モデルとは端的に言えば, 入力データや分布を拡散してデータ構造を破壊する過程を行いながら, その破壊されていく様子を記録することで時間を巻き戻す方法を構築し, 別の構造のないデータに対してこの巻き戻しによって同様のデータ構造を新たに生成する手法である.

より具体的には次のようなものである. ある初期状態の分布 $P_0(\mathbf{x})$ を入力するデータの分布とし, これに対して Fokker-Planck 方程式や Langevin 方程式による $[0, \tau]$ の時間区間の拡散ダイナミクスで十分拡散させる過程を考える. 例えば $\mathbf{F}_t(\mathbf{x}) = \mathbf{0}$ のような外力が加わらない拡散ダイナミクスにおいては, 終状態の分布 $P_\tau(\mathbf{x})$ は十分 τ が大きければ一様分布に近くなるため, これはデータ構造が十分拡散によって破壊された分布とみなせるだろう. そして拡散のダイナミクスの様子を逐一覚えておけば, 原理的にはデータ構造が破壊された分布 $P_\tau(\mathbf{x})$ から出発して初期状態の分布 $P_0(\mathbf{x})$ をダイナミクスによって再現できるような時間逆方向のダイナミクスを構築することができる. 特に真の終状態の分布 $P_\tau(\mathbf{x})$ とは異なるデータ構造のない分布 $P'_\tau(\mathbf{x}) (\neq P_\tau(\mathbf{x}))$ から出発し, さらに不完全な時

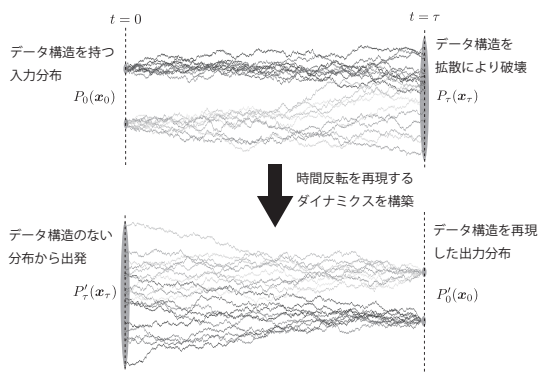


図2 拡散モデルの概念図. 時間反転した拡散ダイナミクスを再現するような時間逆向きのダイナミクスを構築して、データ構造を再現した出力を生成する。

時間逆向きのダイナミクスを用いる場合には、当初入力した分布 $P_0(\mathbf{x})$ とは別の分布 $P'_0(\mathbf{x}) (\neq P_0(\mathbf{x}))$ が不完全な時間逆向きのダイナミクスの結果得られるだろう。ここで得られた別の分布 $P'_0(\mathbf{x})$ を入力データ $P_0(\mathbf{x})$ の模倣として生成された出力の分布とみなすのが、生成AIの手法としての拡散モデルの考え方である(図2)。

この拡散モデルの考え方は、ゆらぎの定理の主張に出てくる逆向きの経路の確率の影響を受けたものであると恐らく考えられる。拡散モデルにおける原論文¹⁾では、ある分布(例えば定常分布) $P'_\tau(\mathbf{x})$ から出発して、時間逆向きのダイナミクスにおける遷移確率 $T'_t(\mathbf{x}_t|\mathbf{x}_{t+dt})$ を用いることで、

$$P_R(\Gamma) = \prod_{n=1}^N T'_{(n-1)dt}(\mathbf{x}_{(n-1)dt}|\mathbf{x}_{ndt})P'_\tau(\mathbf{x}_\tau), \quad (17)$$

という時間逆向きのダイナミクスの経路の確率を考えている。ただしこの遷移確率 T'_t は時間逆向きのダイナミクスについて定義されるため、時間順方向の遷移確率である T_t とは異なるものである。よって P_R は P_B とは本質的に異なるものであることに注意したい。以上より、拡散モデルの手法はゆらぎの定理の影響を受けているものの本質的には異なる操作を行っており、ゆらぎの定理と拡散モデルの手法との関係はアナロジ的な関係

に過ぎないというのが本当のところである。

しかしながら、数理の点においてはそのアナロジーから得るものがあるであろう。実際に拡散モデルの原論文¹⁾においてはゆらぎの定理と同様の計算を行いながら、時間逆向きのダイナミクスにおける遷移確率 T'_t の学習を行うことを考えている。この遷移確率 T'_t の学習は、生成される出力の分布 $P'_0(\mathbf{x})$ と入力の分布 $P_0(\mathbf{x})$ をできるだけ近づけるように次のように行われる。まず生成される出力の分布 $P'_0(\mathbf{x})$ は時刻 $t=0$ を除いた経路 $\Gamma_{>0} = (\mathbf{x}_{dt}, \mathbf{x}_{2dt}, \dots, \mathbf{x}_{Ndt})$ に対する周辺化で $P'_0(\mathbf{x}) = \int d\Gamma_{>0} P_R(\Gamma)$ のように得られる。次に入力するデータの分布 $P_0(\mathbf{x})$ と生成された出力の分布 $P'_0(\mathbf{x})$ の間の負のクロスエントロピー

$$-H(P_0\|P'_0) = \int d\mathbf{x}_0 P_0(\mathbf{x}_0) \ln P'_0(\mathbf{x}_0), \quad (18)$$

を考える。この負のクロスエントロピーの上限は $P'_0 = P_0$ の時達成される $-H(P_0\|P_0)$ であるため、この $-H(P_0\|P'_0)$ がどれだけこの上限に近いかが、出力された分布 $P'_0(\mathbf{x})$ が入力された $P_0(\mathbf{x})$ にどれだけ近づけているかを表す。ただし P'_0 は $P'_0(\mathbf{x}) = \int d\Gamma_{>0} P_R(\Gamma)$ で与えられる T'_t の複雑な関数なため、そのまま $-H(P_0\|P'_0)$ を計算したものを T'_t に対して最大化することで上限に近づけるのは得策ではない。ここでゆらぎの定理で用いられているような計算により、 $-H(P_0\|P'_0)$ に対する計算しやすい下限 $K (\leq -H(P_0\|P'_0))$ を考える。そして $-H(P_0\|P'_0)$ の値の範囲を下から抑えるこの下限 K を最大化するような遷移確率 T'_t を得ることで、 $P'_0 = P_0$ に近いような時間逆向きのダイナミクスを構築している。

このように拡散モデルの手法は元々非平衡熱力学、特にゆらぎの定理に想起されるような手法によって考えられてきたものであるが、現在では非平衡熱力学の手法、特にゆらぎの定理の手法に立脚することは主流ではなくなっている。現在では拡散モデルの原論文¹⁾の手法は大きく改良され、昨今の主流の方法では K のような量の最大化によって時間方向逆向きのダイナミクスを構築するわけ

ではない。例えば外力のない状況 ($\mathbf{F}_t(\mathbf{x}) = \mathbf{0}$) で
の速度場は $\boldsymbol{\nu}_t(\mathbf{x}) = -D_t \nabla \ln P_t(\mathbf{x})$ で与えられ
る。よってこの速度場 $\boldsymbol{\nu}_t(\mathbf{x})$ もしくはスコア関数
と呼ばれる $\nabla \ln P_t(\mathbf{x})$ という量を知っていれば、
時間逆方向のダイナミクスも連続の式 (2) を時間
反転をするだけで原理的に構成可能なはずである。
よって、この速度場 $\boldsymbol{\nu}_t(\mathbf{x})$ そのものの量の推定、も
しくはスコア関数 $\nabla \ln P_t(\mathbf{x})$ の推定によって、時
間逆向きのダイナミクスを再現するということが
実際に行われている。このうち速度場 $\boldsymbol{\nu}_t(\mathbf{x})$ に立
脚するものをフローベース生成モデル⁷⁾、スコア関
数 $\nabla \ln P_t(\mathbf{x})$ に立脚するものをスコアベース生成
モデル⁸⁾と呼び、これらの手法が現在は主流となっ
ている。

6. 拡散モデルと最適輸送

生成 AI の手法としての拡散モデルで重要とな
るのは、生成されたデータ構造が元のデータ構造
をきちんと再現することである。拡散モデルでは
拡散ダイナミクスに選択の余地があるため、この
データ構造の再現においてどのような拡散ダイナ
ミクスを選ぶとパフォーマンスが上がるか、とい
うのは拡散モデルにおける一つの重要なトピック
になる。拡散ダイナミクスは拡散係数 D_t で決ま
るため、 D_t としてどのような時間依存性を導入す
べきか、ということは拡散モデルの手法におけ
る一つの問題になっている。この D_t の時間依存
性はノイズスケジュールと呼ばれており、これま
で様々なノイズスケジュールが提案されてきた。

このノイズスケジュールとしてフローベースモ
デルにおいては、条件付き最適輸送 (cond-OT)⁷⁾
と呼ばれるノイズスケジュールが提案されている。
このノイズスケジュールを用いた際には、データ
数 n がデータの次元 d に比べて $n/\sqrt{d} \rightarrow 0$ となる
ような十分小さい時には、分布の時間変化が近似的
に 2-Wasserstein 距離と呼ばれる量に対応する
最適輸送¹⁰⁾になることが示されている⁹⁾。この
最適輸送の手法が拡散モデルの生成 AI としての
最適性を示しているかについてはわからないもの

の、最適輸送の理論は拡散のダイナミクスと極めて
相性の良いことが知られているため、非常に有望
な手法と言えるだろう。その相性の良さの例とし
ては、例えば Fokker-Planck 方程式における確
率分布の時間発展が、最適輸送理論において考え
られている 2-Wasserstein 距離と、エントロピー
生成に極めて関係が深い自由エネルギーという量
を用いた最適化問題の解として決定されることが
知られていることが挙げられる¹¹⁾。

最適輸送のダイナミクスはさらにエントロピー
生成と呼ばれる量に対するある種の最適化に相当
する、初期状態と終状態を固定した下での最小エン
トロピー生成を達成する条件になっていることが
知られている¹²⁾。本稿でこの最適輸送理論の数理
がいかに非平衡熱力学と関係づくかについて、我々
の研究を含む結果を紹介しながら締めたいと思う。

7. 最適輸送と非平衡熱力学

最適輸送理論¹⁰⁾と非平衡熱力学の関係を議論す
る上で、まずは 2-Wasserstein 距離と呼ばれる分布
間の距離の指標を導入しよう。Benamou-Brenier
公式¹³⁾と呼ばれるものに基づく、分布 P, Q 間の
2-Wasserstein 距離 $\mathcal{W}_2(P, Q)$ は連続の式を用い
て次のように定義される。

$$\mathcal{W}_2(P, Q) = \sqrt{\inf_{\tau} \int_0^{\tau} dt \int d\mathbf{x} \|\mathbf{u}_t(\mathbf{x})\|^2 P_t(\mathbf{x})}, \quad (19)$$

ただし下限 $\inf$ は次の連続の式

$$\begin{aligned} \partial_t P_t(\mathbf{x}) &= -\nabla \cdot (\mathbf{u}_t(\mathbf{x}) P_t(\mathbf{x})), \\ P_0(\mathbf{x}) &= P(\mathbf{x}), \quad P_{\tau}(\mathbf{x}) = Q(\mathbf{x}), \end{aligned} \quad (20)$$

を満たす任意のパス $(\mathbf{u}_t(\mathbf{x}), P_t(\mathbf{x}))$ の範囲でとる。
この 2-Wasserstein 距離が発散しないような分布
 P, Q に対して定義されることに注意したい。

もし拡散係数 D_t が時間に依存しない場合 ($D_t = \mu T$) は 2-Wasserstein 距離を用いて、エントロピー
生成に対する次のような下限を得ることができ
る¹²⁾。

$$\Sigma_{[0,\tau]} \geq \frac{[\mathcal{W}_2(P_0, P_\tau)]^2}{\mu T \tau} \quad (21)$$

この結果は、エントロピー生成率の定義式 (4) で出てくる $\nu_t(\mathbf{x})$ が Fokker-Planck 方程式を連続の式とみなしたときの速度場であることから得られる。さらに拡散係数 D_t が時間に依存する場合においてもエントロピー生成率に対して、同様の手法を用いることで次のような下限を導くことができる¹⁴⁾。

$$\sigma_t \geq \frac{v_t^2}{\mu T_t}, \quad (22)$$

$$v_t = \lim_{\Delta t \rightarrow +0} \frac{\mathcal{W}_2(P_t, P_{t+\Delta t})}{\Delta t}.$$

この不等式 (22) は外力 $\mathbf{F}_t(\mathbf{x})$ が何らかのポテンシャル関数 $U_t(\mathbf{x})$ の勾配 $-\nabla U_t(\mathbf{x})$ で記述できるポテンシャル力の場合に等号が成立する。さらに不等式 (21) が等号達成、すなわち最小エントロピー生成を達成する条件は、拡散ダイナミクスが 2-Wasserstein 距離の世界での測地線に沿っているとき、すなわち

$$v_t = \frac{\mathcal{W}_2(P_0, P_\tau)}{\tau} \quad (23)$$

が任意の $t \in [0, \tau]$ で成立することである。このように最適輸送のダイナミクスは非平衡熱力学において、エントロピー生成で表される熱力学的な散逸を最小化するという重要な位置付けになっている。また我々はこの最適輸送理論の幾何構造に基づいて様々な熱力学的な関係式を議論したり、Kullback-Leibler ダイバージェンスと相性が高い幾何学である情報幾何と最適輸送理論の幾何の関係についてエントロピー生成という量に基づいた議論を行ってきた¹⁵⁾。

8. 展望: 非平衡熱力学の知見を生かして

本稿で説明したように、非平衡熱力学と拡散モデルの手法の間には様々なアナロジーが成り立つことが見て取れるだろう。これは元々、拡散モデルの原論文が非平衡熱力学の手法であるゆらぎの定理の数理に着想を得たからという説明もできる

だろう。さらにそれだけでなく拡散のダイナミクスにおいて、非平衡熱力学の考え方はそもそも本質的であることも理由になると考えている。特に最適輸送などの数理においても Fokker-Planck 方程式で記述される拡散ダイナミクスを考えた際には、自由エネルギーという熱力学的な量が密接に関わることが示されている¹¹⁾。またその最適輸送の数理は実際に非平衡熱力学における最小エントロピー生成の問題と深く関わっている^{12, 14, 15)}。

よって自然な疑問として次のようなものが生じるだろう。非平衡熱力学の知見は拡散モデルという生成 AI の手法に対してどの程度有用になりうるだろうか、という疑問である。私自身は現在、東京大学の学部授業の一環として行なったゼミの学生たちとこの疑問に取り組んでいる。幸いなことに、その疑問の一部を解決できると考えるに至り、"拡散モデルのダイナミクスとして最適輸送を近似するダイナミクスを選ぶ必要があるのか"という疑問に対する、非平衡熱力学の手法を用いた回答を我々は提案しようとしている¹⁶⁾。このような非平衡熱力学の手法と拡散モデルとの手法の接点、及び拡散モデルに限らない生成 AI の手法との接点の研究はまだ始まったばかりであり、非平衡熱力学に立脚した生成 AI の研究が今後ますます発展していくことを期待したいと思う。

参考文献

- 1) J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, & S. Ganguli, *In International conference on machine learning*, 2256-2265. PMLR (2015).
- 2) Y. Song, & S. Ermon, *Advances in neural information processing systems*, **32** (2019).
- 3) J. Ho, A. Jain & P. Abbeel, *Advances in neural information processing systems*, **33**, 6840-6851 (2020).
- 4) U. Seifert, *Reports on progress in physics*, **75**, 126001 (2012).
- 5) D. J. Evans, E. G. D. Cohen, & G. P. Morriss, *Physical review letters*, **71**, 2401 (1993).
- 6) C. Jarzynski, *Physical Review Letters*, **78**, 2690 (1997).
- 7) Y. Lipman, R. T. Chen, H. Ben-Hamu, M. Nickel & M. Le, *In The Eleventh International Conference on Learning Representations* (2022).
- 8) Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Ku-

- mar, S. Ermon & B. Poole, *In International Conference on Learning Representations* (2020).
- 9) N. Shaul, R.T. Chen, M. Nickel, M. Le & Y. Lipman, *In International Conference on Machine Learning*, 30883-30907. PMLR (2023).
 - 10) C. Villani, *Optimal transport: old and new*, Vol. **338**, Springer (2009).
 - 11) R. Jordan, D. Kinderlehrer & F. Otto, *SIAM journal on mathematical analysis*, **29**, 1-17 (1998).
 - 12) E. Aurell, K. Gawedzki, C. Mejia-Monasterio, R. Mohayaei & P. Muratore-Ginanneschi, *Journal of statistical physics*, **147**, 487-505 (2012).
 - 13) J. D. Benamou & Y. Brenier, *Numerische Mathematik*, **84**, 375-393 (2000).
 - 14) M. Nakazato & S. Ito, *Physical Review Research*, **3**, 043093 (2021).
 - 15) S. Ito, *Information Geometry*, **7(Suppl 1)**, 441-483 (2024).
 - 16) K. Ikeda, T. Uda, D. Okanohara and S. Ito, *Physical Review X*, **15**, 031031 (2025).

(いとう そうすけ, 東京大学大学院理学系研究科)