生成模型入門

田中章詞　(RIKEN AIP/iTHEMS)

@計算物理 春の学校 2023

自己紹介

■ 所属

- 理研 AIP/iTHEMS
  └—— 上級研究員

■ 研究

- 機械学習 (理論/応用)
- 数理物理 (素粒子理論)

AkinoriTanaka-phys

学術変革領域研究（A） 2022〜2026年度

深層学習の数理と応用

PD募集中: JREC-in

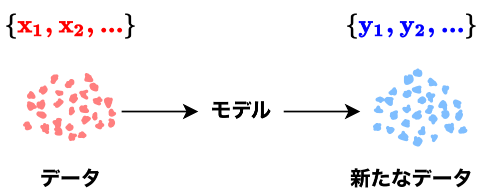
- メンバー
  - 唐木田 亮さん (産総研)
  - 瀧 雅人さん (立教大学)
- 目的
  - 深層学習の研究
    物理バックグラウンドの人も
    welcomeです！

1. 導入

1. 導入

■ 1-1. 生成モデルとは？

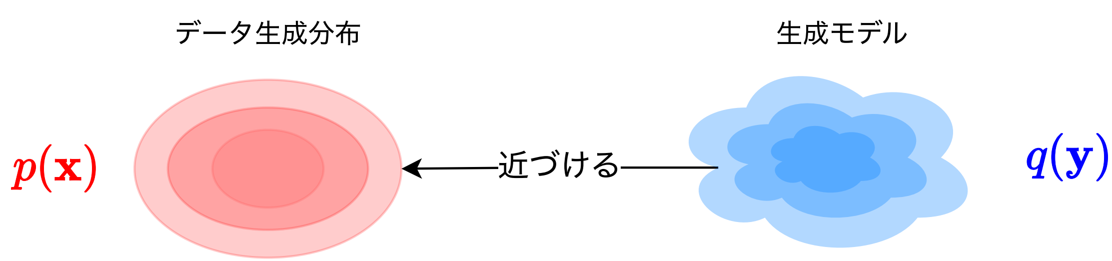
与えられたデータから、それに似たデータを作り出すモデル



■ 例： 文章生成

LARTIUS: O,'tis Marcius! Let's fetch him off, or make remain alike.	$\rightarrow$	AKINORI: Came, bloody decrees, By city, not her brows, he is bloody woe!

データの住むベクトル空間 X 上に確率分布があると考える：

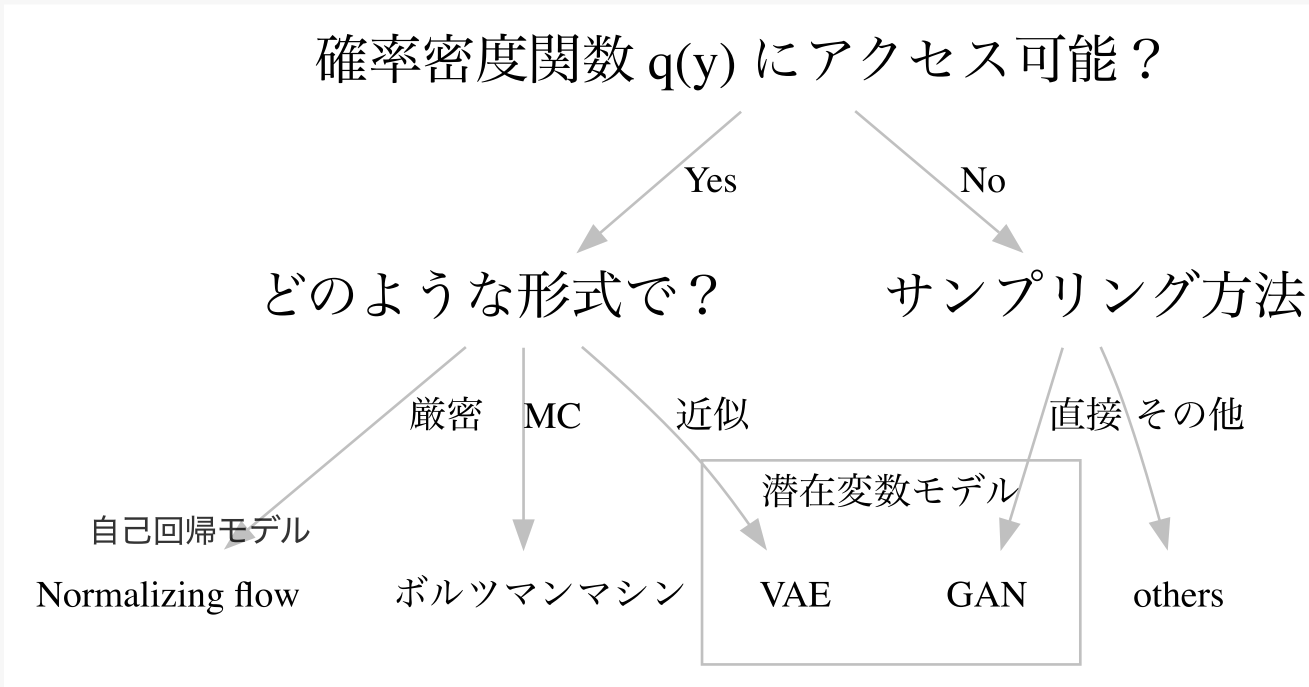


1. 導入

■ 1-2. 生成モデルの分類

（不完全な）分類図

https://arxiv.org/abs/1701.00160 より：



最近は 拡散モデル：確率密度にアクセス可能（数値積分が必要）な場合もある

1. 導入

■ 1-2. 生成モデルの分類

訓練方法による分類

KLダイバージェンス最小化によるもの

- ボルツマンマシン
- 自己回帰モデル（teacher-forcing）
- Normalizing flow

潜在変数モデル

- 変分自己符号化器（VAE）（理想的な場合はKL最小化）
- 敵対的生成ネットワーク(GAN)

拡散モデル

- スコアマッチング
- DDPM（理想的な場合はKL最小化?）
- SDE
- フローマッチング

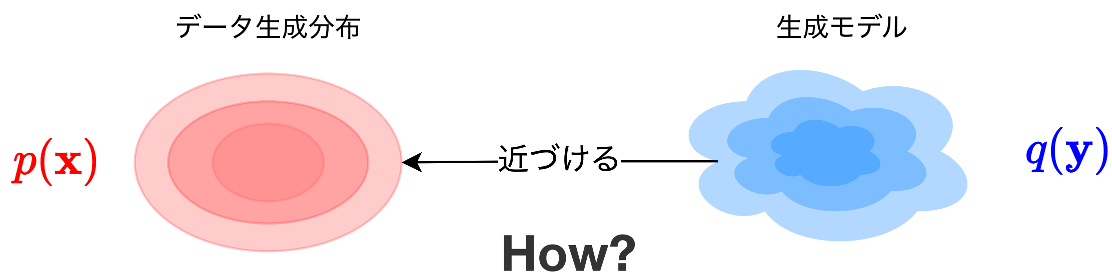
内容

2. KLダイバージェンス最小化によるもの

- ボルツマンマシン
- 自己回帰モデル
- Normalizing flow

2. KLダイバージェンス最小化によるもの

■ 2-1. 基本的な考え方

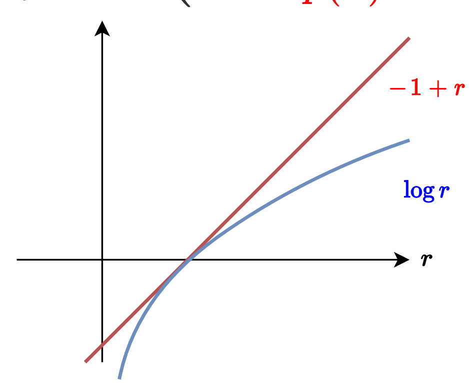


KLダイバージェンス最小化

$$\min_{q} \underbrace{D_{KL}(p\|q)}_{\left\langle \log \frac{p(\mathbf{x})}{q(\mathbf{x})} \right\rangle_p}$$

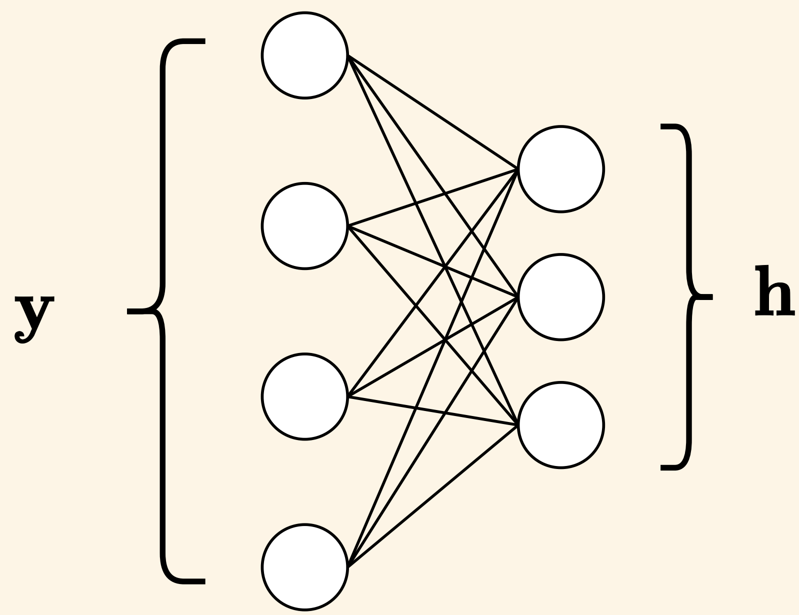
最小値（=0）でのみ $p = q$ となる：

$$D_{KL}(p\|q) = \int p(\mathbf{x}) \left(-\log \frac{q(\mathbf{x})}{p(\mathbf{x})} - 1 + \frac{q(\mathbf{x})}{p(\mathbf{x})} \right) d\mathbf{x} \geq 0$$



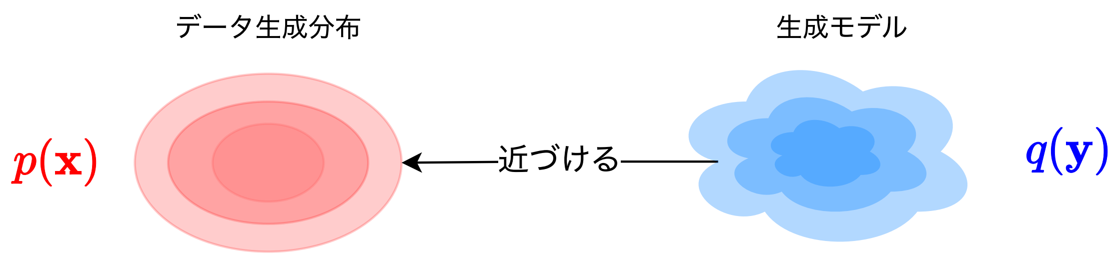
2. KLダイバージェンス最小化によるもの

■ 2-2. ボルツマンマシン



2. KLダイバージェンス最小化によるもの

■ 2-2. ボルツマンマシン



確率密度関数：　ボルツマン分布：$q_\theta(\mathbf{y}) = \dfrac{e^{-E_\theta(\mathbf{y})}}{Z(\theta)}$

やりたいこと

$$\min_\theta D_{KL}(p \| q_\theta)$$

難しいので 以下を繰り返す

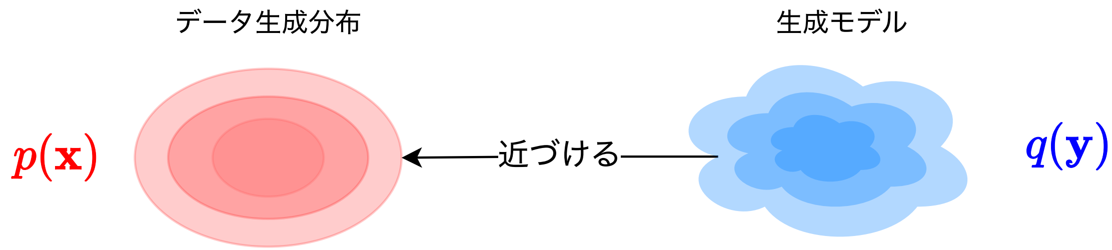
$$\theta_{t+1} = \theta_t - \epsilon_t \underline{\nabla_\theta D_{KL}(p \| q_\theta)}$$

ちょっと計算すると以下がわかります：

$$\underline{\nabla_\theta D_{KL}(p \| q_\theta)} = -\langle \nabla_\theta E_\theta(\mathbf{x}) \rangle_{p(\mathbf{x})} + \langle \nabla_\theta E_\theta(\mathbf{y}) \rangle_{q_\theta(\mathbf{y})}$$

2. KLダイバージェンス最小化によるもの

■ 2-2. ボルツマンマシン



確率密度関数： ボルツマン分布：$q_\theta(\mathbf{y}) = \dfrac{e^{-E_\theta(\mathbf{y})}}{Z(\theta)}$

理想的な学習ルール

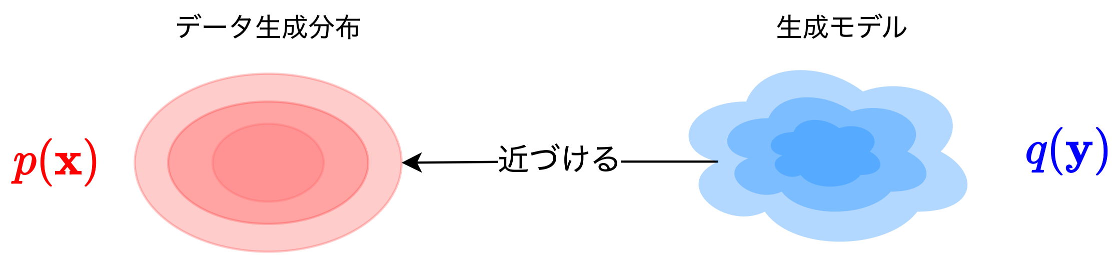
$$\theta_{t+1} = \theta_t + \epsilon_t \left(\langle \nabla_\theta E_\theta(\mathbf{x}) \rangle_{p(\mathbf{x})} - \langle \nabla_\theta E_\theta(\mathbf{y}) \rangle_{q_\theta(\mathbf{y})} \right)$$

実用上の問題：

- $\langle \ldots \rangle_{p(\mathbf{x})}$ は不可能 ➡ データ平均に置き換えて近似
- $\langle \ldots \rangle_{q_\theta(\mathbf{y})}$ は困難 ➡ モンテカルロ(MC)サンプルで置き換えて近似

2. KLダイバージェンス最小化によるもの

■ 2-2. ボルツマンマシン



確率密度関数：　ボルツマン分布：$q_\theta(\mathbf{y}) = \dfrac{e^{-E_\theta(\mathbf{y})}}{Z(\theta)}$

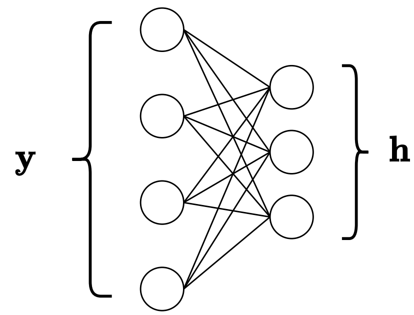
実行可能な学習ルール

$$\theta_{t+1} = \theta_t + \epsilon_t \left(\frac{1}{N_{\mathrm{data}}} \sum_{i=1}^{N_{\mathrm{data}}} \nabla_\theta E_\theta(\underbrace{\mathbf{x}_i}_{\text{データ点}}) - \frac{1}{N_{\mathrm{gen}}} \sum_{j=1}^{N_{\mathrm{gen}}} \nabla_\theta E_\theta(\underbrace{\mathbf{y}_j}_{MC\text{サンプル}}) \right)$$

残る問題：MC サンプル の効率

2. KLダイバージェンス最小化によるもの

■ 2-3. 制限ボルツマンマシン



「補助変数 $\mathbf{h}$」を設定することで 学習がうまくいくようにできる：

制限ボルツマンマシン

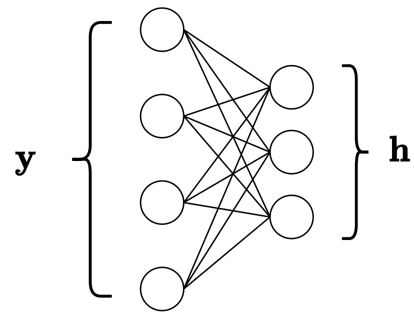
$$E_\theta(\mathbf{y}) = -\log \sum_{\mathbf{h}} e^{-E_\theta(\mathbf{y},\mathbf{h})}$$

エネルギー関数の微分値？

$$\nabla_\theta E_\theta(\mathbf{y}) = \frac{\sum_{\mathbf{h}} \nabla_\theta E_\theta(\mathbf{y},\mathbf{h}) e^{-E_\theta(\mathbf{y},\mathbf{h})}}{\sum_{\mathbf{h}'} e^{-E_\theta(\mathbf{y},\mathbf{h}')}} = \langle \nabla_\theta E_\theta(\mathbf{y},\mathbf{h}) \rangle_{\mathbf{h}\sim q_\theta(\mathbf{h}|\mathbf{y})}$$

2. KLダイバージェンス最小化によるもの

■ 2-3. 制限ボルツマンマシン



「補助変数 $\mathbf{h}$」を設定することで 学習がうまくいくようにできる：

理想的な学習ルール

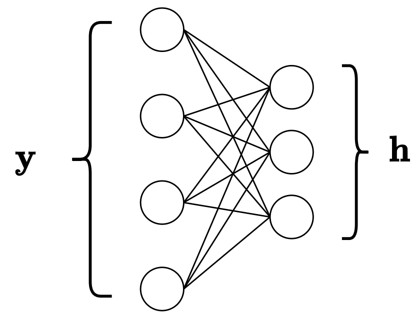
$$\theta_{t+1} = \theta_t + \epsilon_t \left(\langle \nabla_\theta E_\theta(\mathbf{x}, \mathbf{h}) \rangle_{p(\mathbf{x}) q_\theta(\mathbf{h}|\mathbf{x})} - \langle \nabla_\theta E_\theta(\mathbf{y}, \mathbf{h}) \rangle_{q_\theta(\mathbf{y}) q_\theta(\mathbf{h}|\mathbf{y})} \right)$$

実用上は：

- $\langle \ldots \rangle_{p(\mathbf{x}) q_\theta(\mathbf{h}|\mathbf{x})}$ ➡ データ $\mathbf{x} \to q_\theta(\mathbf{h}|\mathbf{x})$ で $\mathbf{h}$ をサンプル
- $\langle \ldots \rangle_{q_\theta(\mathbf{y}) q_\theta(\mathbf{h}|\mathbf{y})}$ ➡ ブロック化ギブスサンプリング

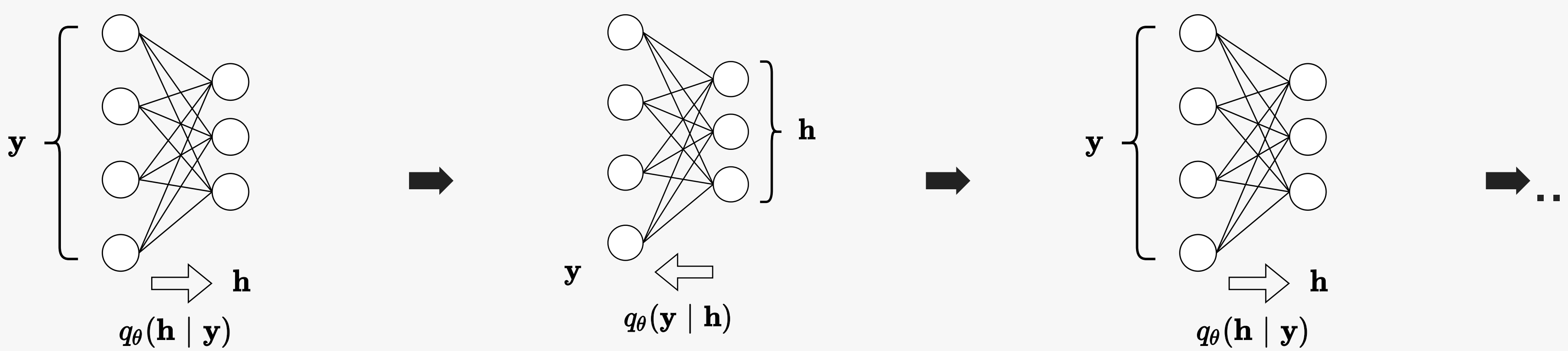
2. KLダイバージェンス最小化によるもの

■ 2-3. 制限ボルツマンマシン



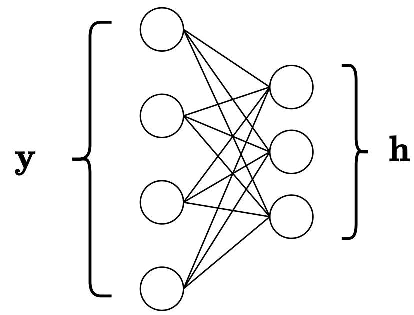
「補助変数 $\mathbf{h}$」を設定することで 学習がうまくいくようにできる：

ブロック化ギブスサンプリング



2. KLダイバージェンス最小化によるもの

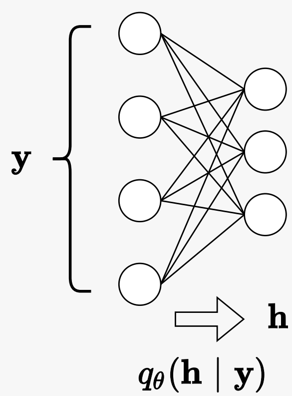
■ 2-3. 制限ボルツマンマシン



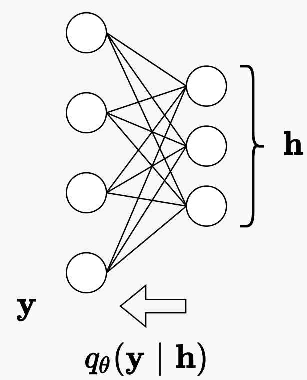
「補助変数 $\mathbf{h}$」を設定することで 学習がうまくいくようにできる：

コントラスティブ・ダイバージェンス法 (CD1)

$\mathbf{y} \sim p(\mathbf{x})$ ➡



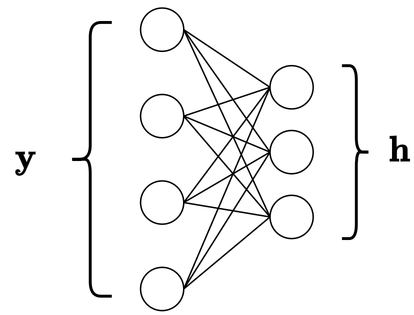
➡



最後の $\mathbf{y}, \mathbf{h}$ を
MCサンプルにする

2. KLダイバージェンス最小化によるもの

■ 2-3. 制限ボルツマンマシン



「補助変数 $\mathbf{h}$」を設定することで 学習がうまくいくようにできる：

実行可能な学習ルール

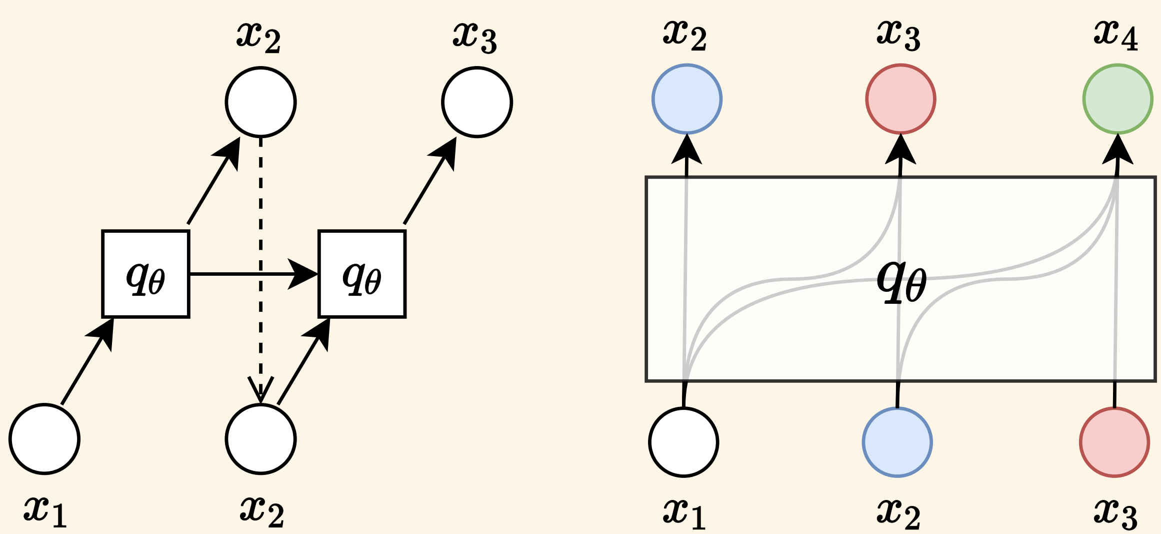
$$\theta_{t+1} = \theta_t + \epsilon_t \left(\frac{1}{N_{\text{data}}} \sum_{i=1}^{N_{\text{data}}} \nabla_\theta E_\theta(\underbrace{\mathbf{x}_i}_{\text{データ点}}, \underbrace{\mathbf{h}_i}_{\sim q_\theta(\mathbf{h}|\mathbf{x}_i)}) - \frac{1}{N_{\text{gen}}} \sum_{j=1}^{N_{\text{gen}}} \nabla_\theta E_\theta(\underbrace{\mathbf{y}_j, \mathbf{h}_j}_{MC\text{サンプル}}) \right)$$

MCサンプルは CD1 で十分とされていたが、近年見直し？

- Langevinサンプリング による訓練の報告：https://arxiv.org/abs/2210.10318

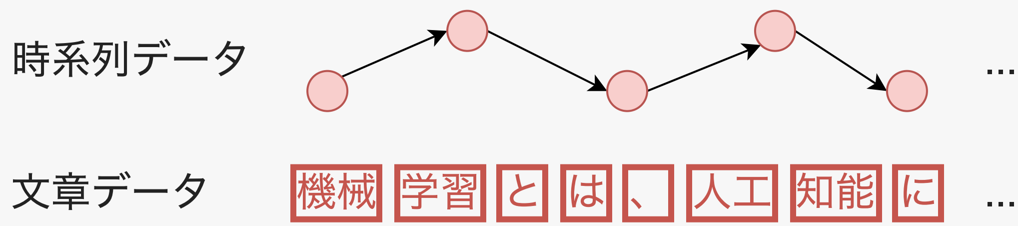
2. KLダイバージェンス最小化によるもの

■ 2-4. 自己回帰モデル



2. KLダイバージェンス最小化によるもの

■ 2-4. 自己回帰モデル



共に連続するベクトルからなるデータと考えられる：

$$\mathbf{x}^{t=1}, \mathbf{x}^{t=2}, \mathbf{x}^{t=3}, \mathbf{x}^{t=4}, \mathbf{x}^{t=5}, \ldots, \mathbf{x}^{t=T}$$

$$\Downarrow \text{略記}$$

$$\mathbf{x}^{1}, \mathbf{x}^{2}, \mathbf{x}^{3}, \mathbf{x}^{4}, \mathbf{x}^{5}, \ldots, \mathbf{x}^{T}$$

従って、データ生成分布も、モデルも

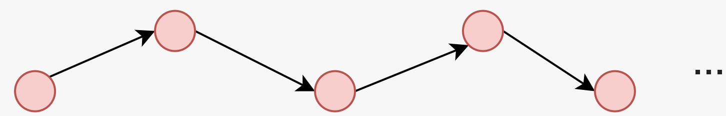
$$p\left(\mathbf{x}^{1}, \mathbf{x}^{2}, \mathbf{x}^{3}, \mathbf{x}^{4}, \mathbf{x}^{5}, \ldots, \mathbf{x}^{T}\right)$$
$$q_\theta\left(\mathbf{x}^{1}, \mathbf{x}^{2}, \mathbf{x}^{3}, \mathbf{x}^{4}, \mathbf{x}^{5}, \ldots, \mathbf{x}^{T}\right)$$

となる。

2. KLダイバージェンス最小化によるもの

■ 2-4. 自己回帰モデル

時系列データ

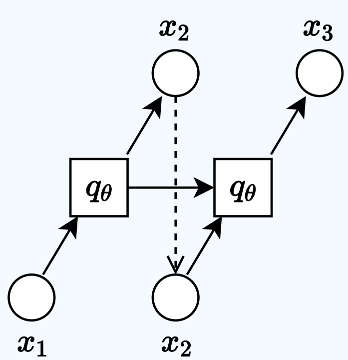


文章データ　機械 学習 と は 、 人工 知能 に …

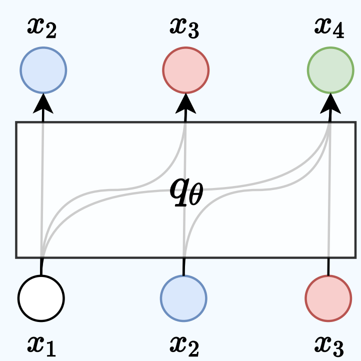
モデルには因果的な構造を課す場合が多い：

$$q_\theta\left(\mathbf{x}^1, \mathbf{x}^2, \mathbf{x}^3, \mathbf{x}^4, \mathbf{x}^5, \ldots, \mathbf{x}^T\right)$$
$$= q_\theta(\mathbf{x}^1) \cdot q_\theta(\mathbf{x}^2|\mathbf{x}^1) \cdot q_\theta(\mathbf{x}^3|\mathbf{x}^1, \mathbf{x}^2) \cdot q_\theta(\mathbf{x}^4|\mathbf{x}^1, \mathbf{x}^2, \mathbf{x}^3) \cdot q_\theta(\mathbf{x}^5|\mathbf{x}^1, \mathbf{x}^2, \mathbf{x}^3, \mathbf{x}^4) \cdots$$

リカレントニューラルネット

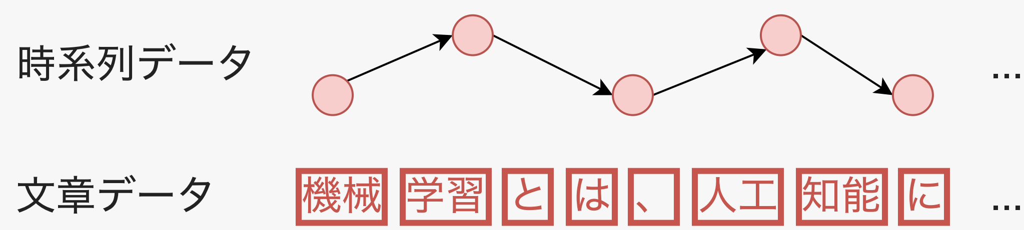


Transformer (位置エンコーディング)



2. KLダイバージェンス最小化によるもの

■ 2-4. 自己回帰モデル



モデルには因果的な構造を課す場合が多い：

$$
\begin{aligned}
&q_\theta\left(\mathbf{x}^1, \mathbf{x}^2, \mathbf{x}^3, \mathbf{x}^4, \mathbf{x}^5, \ldots, \mathbf{x}^T\right) \\
&= q_\theta(\mathbf{x}^1) \cdot q_\theta(\mathbf{x}^2|\mathbf{x}^1) \cdot q_\theta(\mathbf{x}^3|\mathbf{x}^1, \mathbf{x}^2) \cdot q_\theta(\mathbf{x}^4|\mathbf{x}^1, \mathbf{x}^2, \mathbf{x}^3) \cdot q_\theta(\mathbf{x}^5|\mathbf{x}^1, \mathbf{x}^2, \mathbf{x}^3, \mathbf{x}^4) \cdots
\end{aligned}
$$

やりたいこと

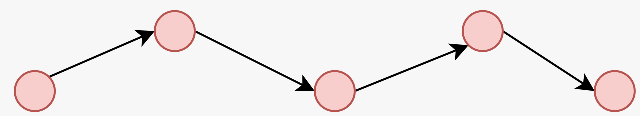
$$
\min_\theta D_{KL}(p \| q_\theta)
$$

因果構造があるので、少し分解できる

2. KLダイバージェンス最小化によるもの

■ 2-4. 自己回帰モデル

時系列データ



…

文章データ　機械 学習 と は 、 人工 知能 に …

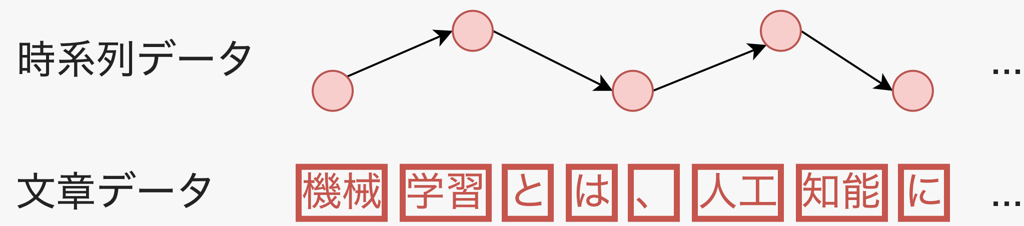
モデルには因果的な構造を課す場合が多い：

$$q_\theta\left(\mathbf{x}^1, \mathbf{x}^2, \mathbf{x}^3, \mathbf{x}^4, \mathbf{x}^5, \ldots, \mathbf{x}^T\right)$$
$$= q_\theta(\mathbf{x}^1) \cdot q_\theta(\mathbf{x}^2|\mathbf{x}^1) \cdot q_\theta(\mathbf{x}^3|\mathbf{x}^1, \mathbf{x}^2) \cdot q_\theta(\mathbf{x}^4|\mathbf{x}^1, \mathbf{x}^2, \mathbf{x}^3) \cdot q_\theta(\mathbf{x}^5|\mathbf{x}^1, \mathbf{x}^2, \mathbf{x}^3, \mathbf{x}^4) \cdots$$

$$\begin{aligned} D_{KL}(p\|q_\theta) + \underbrace{S(p)}_{\text{エントロピー}} &= -\left\langle \log q_\theta\left(\mathbf{x}^1, \mathbf{x}^2, \mathbf{x}^3, \ldots, \mathbf{x}^T\right)\right\rangle_p \\ &= -\left\langle \log q_\theta(\mathbf{x}^1) \cdot q_\theta(\mathbf{x}^2|\mathbf{x}^1) \cdot q_\theta(\mathbf{x}^3|\mathbf{x}^1, \mathbf{x}^2) \cdots \right\rangle_p \\ &= -\sum_{t=1}^{T} \left\langle \log q_\theta(\mathbf{x}^t|\mathbf{x}^1, \ldots, \mathbf{x}^{t-1})\right\rangle_p \end{aligned}$$

2. KLダイバージェンス最小化によるもの

■ 2-4. 自己回帰モデル



モデルには因果的な構造を課す場合が多い：

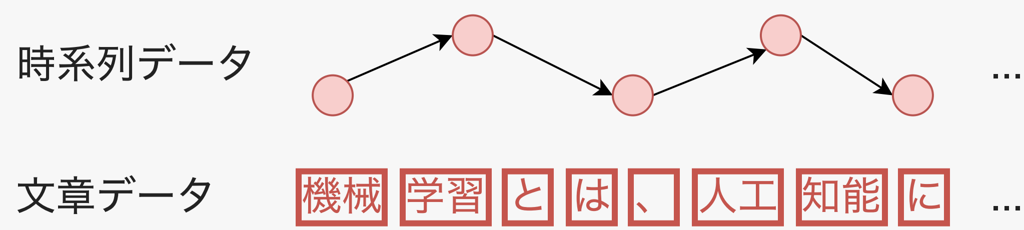
$$
\begin{aligned}
&q_\theta\left(\mathbf{x}^1, \mathbf{x}^2, \mathbf{x}^3, \mathbf{x}^4, \mathbf{x}^5, \dots, \mathbf{x}^T\right) \\
&= q_\theta(\mathbf{x}^1) \cdot q_\theta(\mathbf{x}^2|\mathbf{x}^1) \cdot q_\theta(\mathbf{x}^3|\mathbf{x}^1, \mathbf{x}^2) \cdot q_\theta(\mathbf{x}^4|\mathbf{x}^1, \mathbf{x}^2, \mathbf{x}^3) \cdot q_\theta(\mathbf{x}^5|\mathbf{x}^1, \mathbf{x}^2, \mathbf{x}^3, \mathbf{x}^4) \cdots
\end{aligned}
$$

勾配更新

$$
\theta_{t+1} = \theta_t - \epsilon_t \nabla_\theta \left(- \sum_{t=1}^{T} \langle \log q_\theta(\mathbf{x}^t|\mathbf{x}^1, \dots, \mathbf{x}^{t-1}) \rangle_p \right)
$$

2. KLダイバージェンス最小化によるもの

■ 2-4. 自己回帰モデル



モデルには因果的な構造を課す場合が多い：

$$
\begin{aligned}
&q_\theta\left(\mathbf{x}^1, \mathbf{x}^2, \mathbf{x}^3, \mathbf{x}^4, \mathbf{x}^5, \ldots, \mathbf{x}^T\right) \\
&= q_\theta(\mathbf{x}^1) \cdot q_\theta(\mathbf{x}^2|\mathbf{x}^1) \cdot q_\theta(\mathbf{x}^3|\mathbf{x}^1, \mathbf{x}^2) \cdot q_\theta(\mathbf{x}^4|\mathbf{x}^1, \mathbf{x}^2, \mathbf{x}^3) \cdot q_\theta(\mathbf{x}^5|\mathbf{x}^1, \mathbf{x}^2, \mathbf{x}^3, \mathbf{x}^4) \cdots
\end{aligned}
$$

実行可能な学習ルール

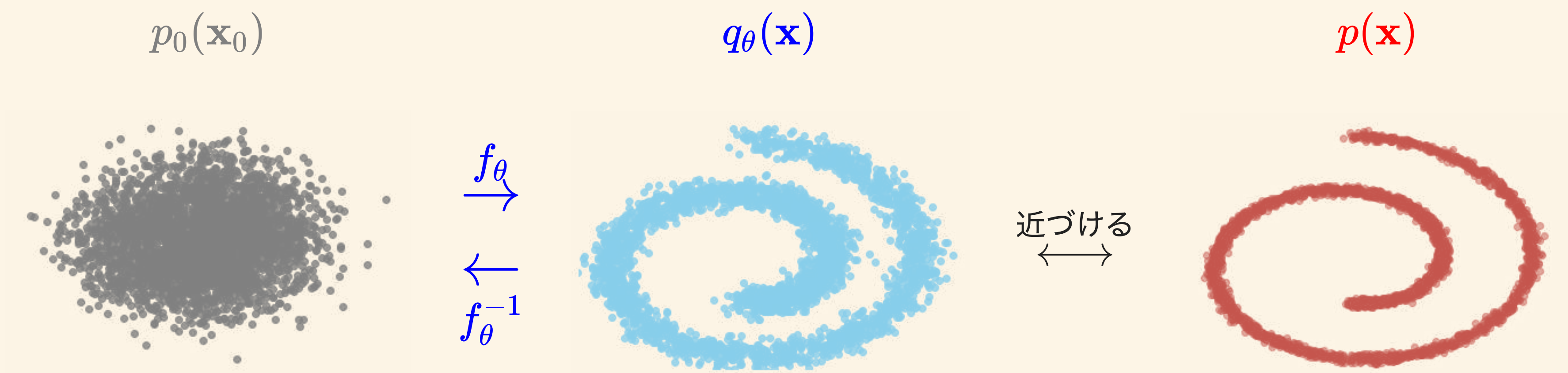
$$
\theta_{t+1} = \theta_t - \epsilon_t \nabla_\theta \frac{1}{N_{\mathrm{data}}} \sum_{i=1}^{N_{\mathrm{data}}} \left(- \sum_{t=1}^{T} \log q_\theta(\mathbf{x}_i^t | \mathbf{x}_i^1, \ldots, \mathbf{x}_i^{t-1}) \right)
$$

入力を自身の過去の生成ベクトルではなく、データ（教師）由来のものに取るため、しばしば "teacher-forcing" と呼ばれる。

2. KLダイバージェンス最小化によるもの

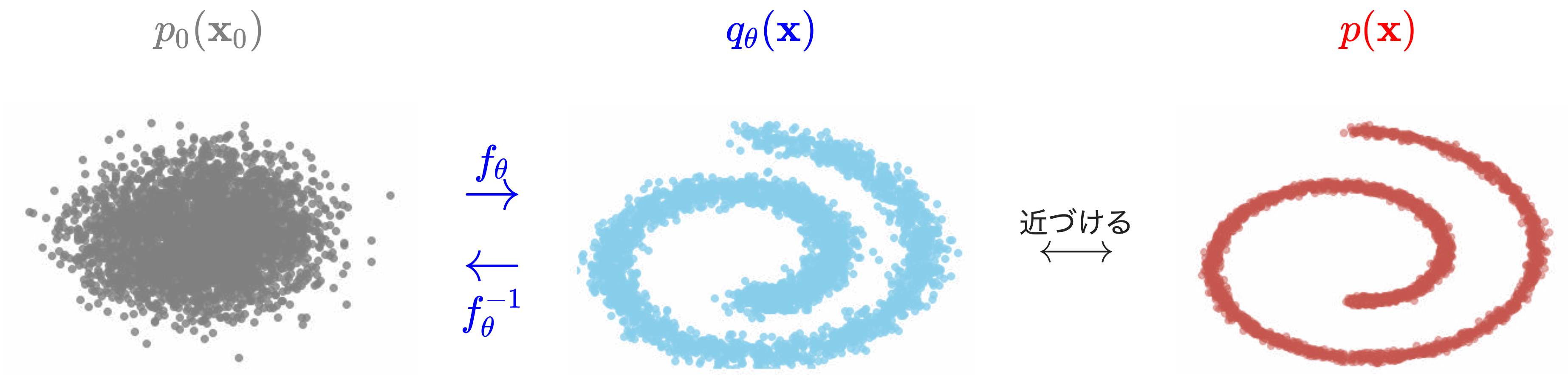
■ 2-5. Normalizing flow

レビュー論文：https://arxiv.org/abs/1912.02762



2. KLダイバージェンス最小化によるもの

■ 2-5. Normalizing flow



もし、可逆な関数 f_θ があれば、

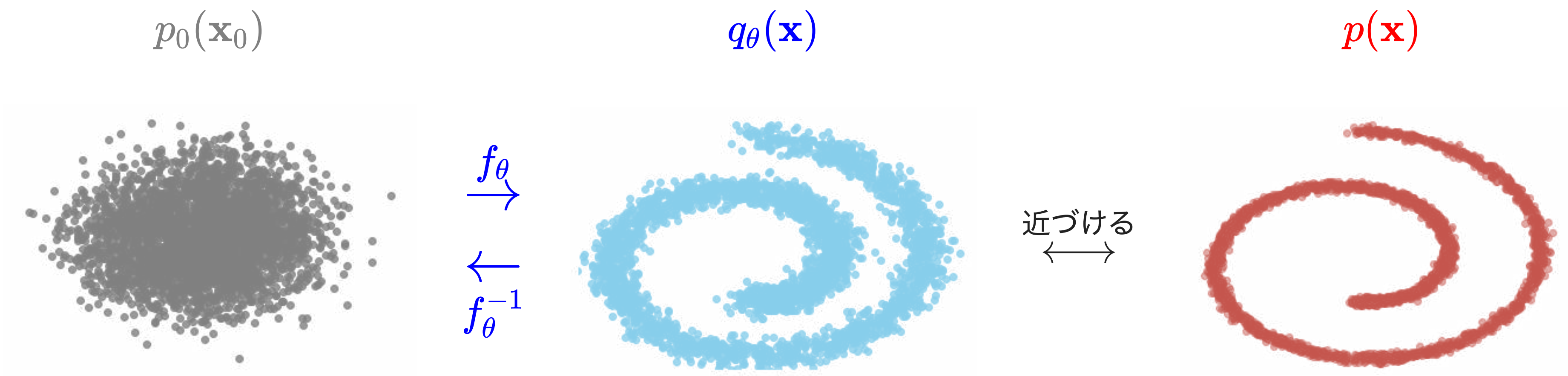
$$q_\theta(\mathbf{x}) := \int \delta(\mathbf{x} - f_\theta(\mathbf{x}_0)) p_0(\mathbf{x}_0) d\mathbf{x}_0$$

$$= \underbrace{\left| \det \nabla f_\theta^{-1}(\mathbf{x}) \right|}_{\text{Jacobian}} \cdot p_0(f_\theta^{-1}(\mathbf{x}))$$

Jacobian が計算できれば、密度関数を計算できる！

2. KLダイバージェンス最小化によるもの

■ 2-5. Normalizing flow



訓練の指針(a)

$$\min_\theta D_{KL}(p \| q_\theta)$$

or

訓練の指針(b)

$$\min_\theta D_{KL}(q_\theta \| p)$$

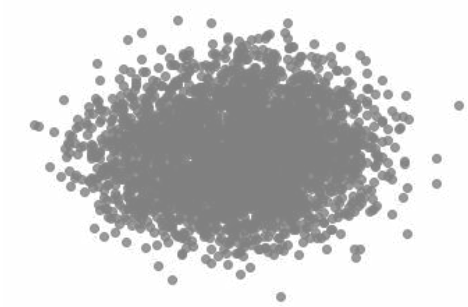
※ $D_{KL}(p^{(1)} \| p^{(2)}) + S(p^{(1)}) \approx -\frac{1}{N}\sum_{i=1}^{N} \log p^{(2)}(\mathbf{x}_i^{(1)})$ では、以下が必要

- $p^{(1)}$ からのサンプリング $\mathbf{x}_i^{(1)}$ が容易であること
- $p^{(2)}$ の密度関数が計算できること

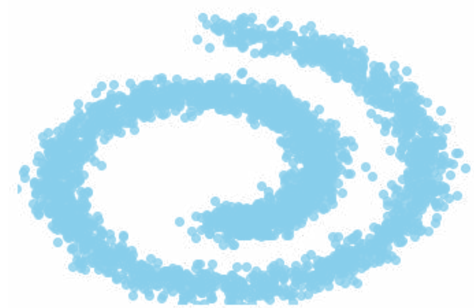
2. KLダイバージェンス最小化によるもの

■ 2-5. Normalizing flow

$p_0(\mathbf{x}_0)$ $\quad$ $q_\theta(\mathbf{x})$ $\quad$ $p(\mathbf{x})$



$\xrightarrow{f_\theta}$ $\quad$ $\xleftarrow[f_\theta^{-1}]{}$



近づける $\longleftrightarrow$

訓練の指針(a)

$$\min_{\theta} D_{KL}(p \| q_\theta)$$

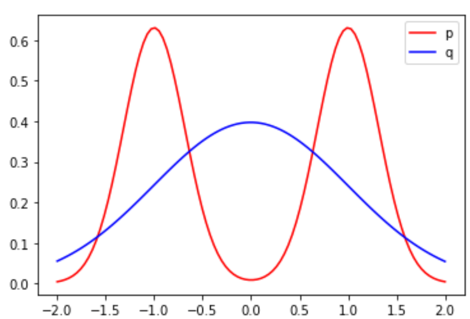
or

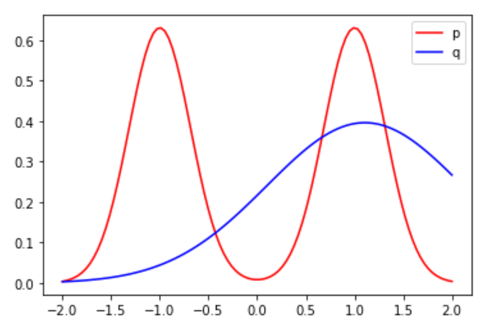
訓練の指針(b)

$$\min_{\theta} D_{KL}(q_\theta \| p)$$

q_θ : Gaussian

p : 2つピーク





※ Normalizing flowではなく

単なる勾配による訓練

3. 潜在変数モデル

- 変分自己符号化器（VAE）
- 敵対的生成ネットワーク(GAN)

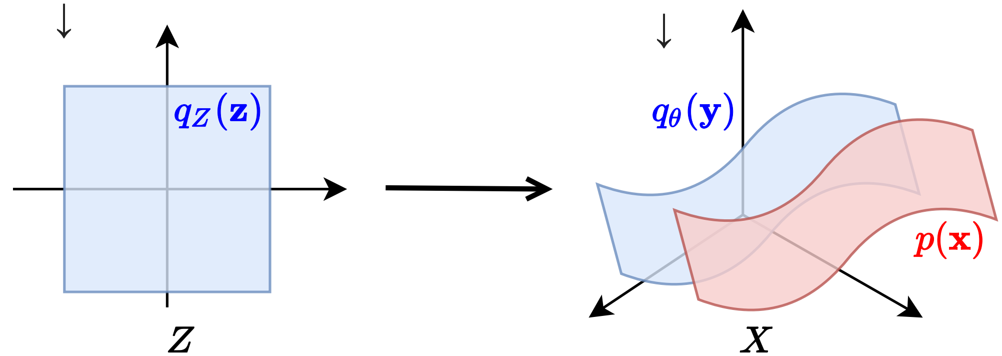
3. 潜在変数モデル

■ 3-1. 基本的な考え方

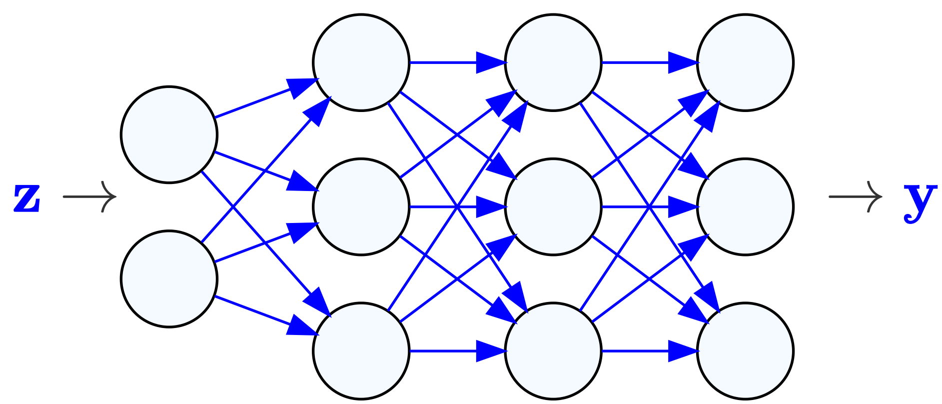
その上に確率密度関数を設定 ↓

それを X に「押し出す」 ↓

ターゲット空間 X より
低次元の空間 Z を設定
（**潜在空間**という）



→ には色々なものを使って良いが、最近は(深層)ニューラルネットを使うのが多い

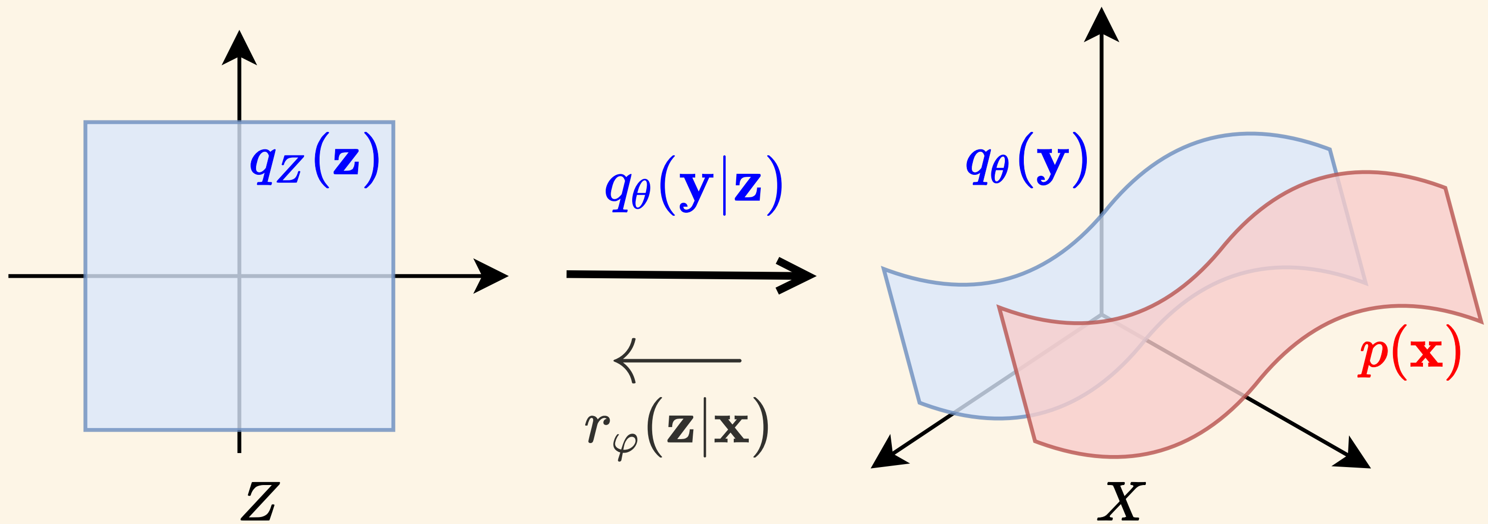


各層での重み行列＋バイアスベクトル ＝ θ

3. 潜在変数モデル

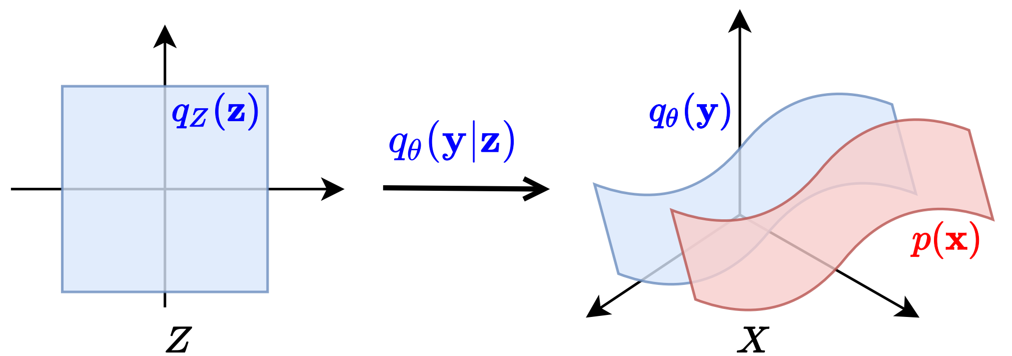
■ 3-2. 変分自己符号化器 (VAE)

レビュー論文：https://arxiv.org/abs/1906.02691



3. 潜在変数モデル

■ 3-2. 変分自己符号化器 (VAE)



X 上の 密度：$q_\theta(\mathbf{y}) = \int_Z q_\theta(\mathbf{y}|\mathbf{z}) q_Z(\mathbf{z}) d\mathbf{z}$ ➡ $\min_\theta D_{KL}(p \| q_\theta)$ は難

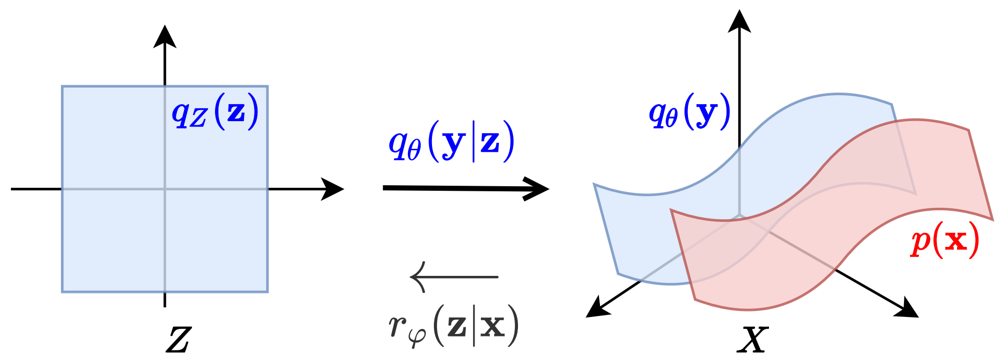
変分によるバウンド (→ 証明)

任意の逆向きプロセス $r(\mathbf{z}|\mathbf{x})$ について：

$$D_{KL}(p \| q_\theta) + \underbrace{S(p)}_{\text{エントロピー}} \leq -\left\langle \log \frac{q_\theta(\mathbf{x}|\mathbf{z}) q_Z(\mathbf{z})}{r(\mathbf{z}|\mathbf{x})} \right\rangle_{p(\mathbf{x}) r(\mathbf{z}|\mathbf{x})}$$

3. 潜在変数モデル

■ 3-2. 変分自己符号化器 (VAE)



X 上の 密度：$q_\theta(\mathbf{y}) = \int_Z q_\theta(\mathbf{y}|\mathbf{z}) q_Z(\mathbf{z}) d\mathbf{z}$ ➡ $\min_\theta D_{KL}(p \| q_\theta)$ は難

VAEの目的

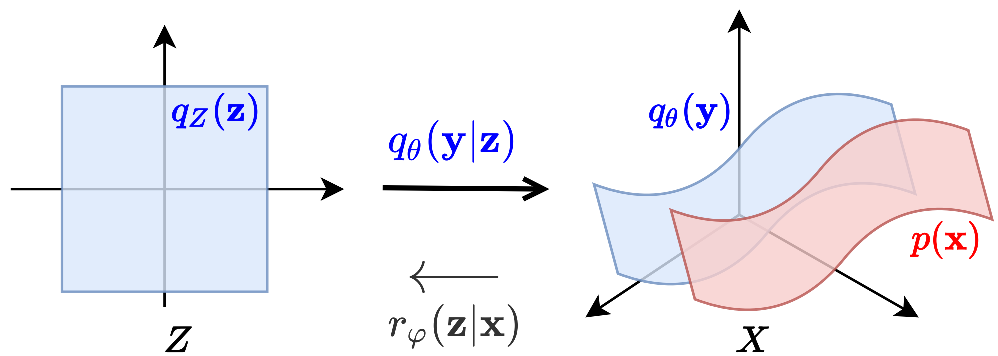
逆向きのモデル $r_\varphi(\mathbf{z}|\mathbf{x})$ を導入して：

$$\min_{\theta,\varphi} \left(- \left\langle \log \frac{q_\theta(\mathbf{x}|\mathbf{z}) q_Z(\mathbf{z})}{r_\varphi(\mathbf{z}|\mathbf{x})} \right\rangle_{p(\mathbf{x}) r_\varphi(\mathbf{z}|\mathbf{x})} \right)$$

※ 実際にはやはり 勾配更新 を行う（次ページ）

3. 潜在変数モデル

■ 3-2. 変分自己符号化器 (VAE)



VAEの勾配更新

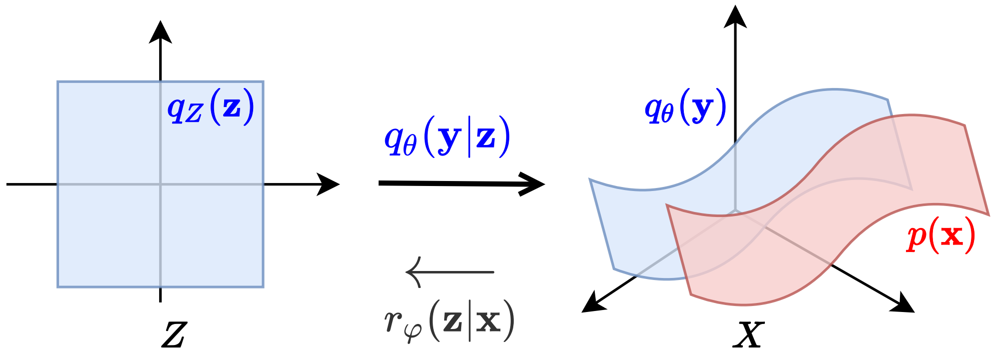
$$\theta_{t+1} = \theta_t + \epsilon_t \nabla_\theta \left\langle \log \frac{q_\theta(\mathbf{x}|\mathbf{z}) q_Z(\mathbf{z})}{r_\varphi(\mathbf{z}|\mathbf{x})} \right\rangle_{p(\mathbf{x}) r_\varphi(\mathbf{z}|\mathbf{x})}$$

$$\varphi_{t+1} = \varphi_t + \epsilon_t \nabla_\varphi \left\langle \log \frac{q_\theta(\mathbf{x}|\mathbf{z}) q_Z(\mathbf{z})}{r_\varphi(\mathbf{z}|\mathbf{x})} \right\rangle_{p(\mathbf{x}) r_\varphi(\mathbf{z}|\mathbf{x})}$$

※ 実際には 期待値は サンプリング で近似する

3. 潜在変数モデル

■ 3-2. 変分自己符号化器 (VAE)



注意

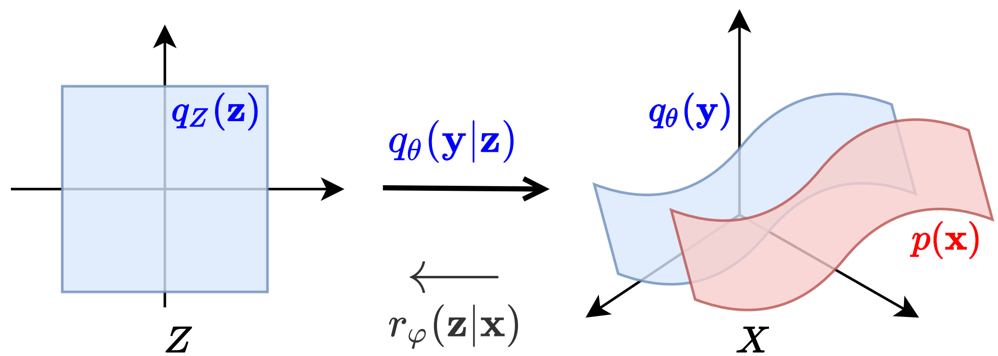
φ の微分について注意

- 👍 $\nabla_\theta \langle \dots \rangle_{p(\mathbf{x})r_\varphi(\mathbf{z}|\mathbf{x})} = \langle \nabla_\theta \dots \rangle_{p(\mathbf{x})r_\varphi(\mathbf{z}|\mathbf{x})} \overset{\text{大数の法則}}{\longleftarrow} \frac{1}{N_{\text{samples}}} \sum_{i=1}^{N_{\text{samples}}} \nabla_\theta \dots$
- 👎 $\nabla_\varphi \langle \dots \rangle_{p(\mathbf{x})r_\varphi(\mathbf{z}|\mathbf{x})} \neq \langle \nabla_\varphi \dots \rangle_{p(\mathbf{x})r_\varphi(\mathbf{z}|\mathbf{x})}$

うまく $\langle \dots \rangle_{p(\mathbf{x})r_\varphi(\mathbf{z}|\mathbf{x})} = \langle f(\dots, \varepsilon, \varphi) \rangle_{p(\mathbf{x})s(\varepsilon)}$ となるような モデルを使う。
これを reparametrization trick という。

3. 潜在変数モデル

■ 3-2. 変分自己符号化器 (VAE)



幾つかのコメント：

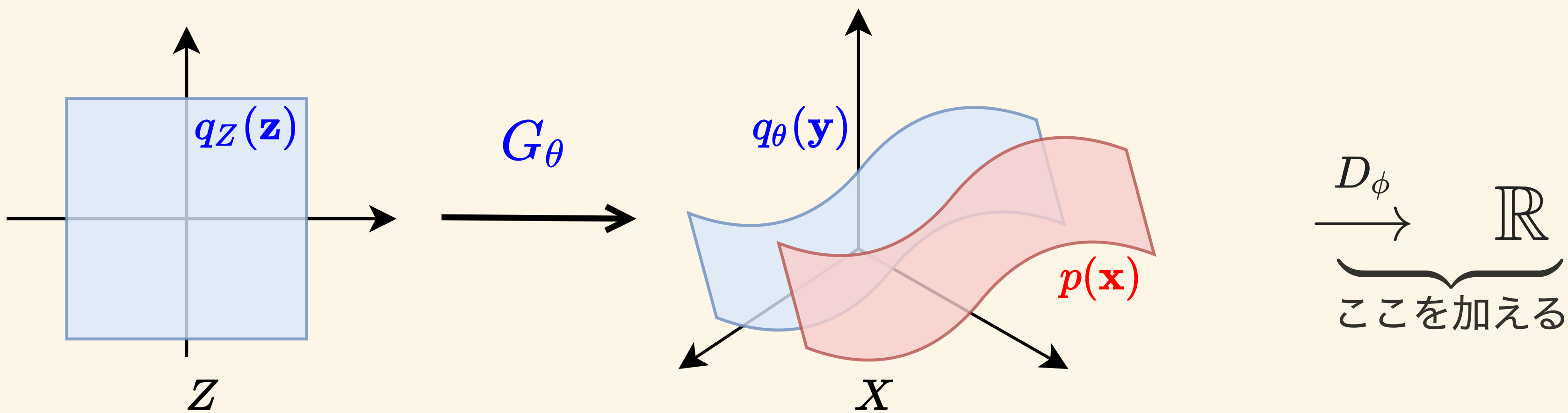
- 理想的なケースでは $q_{\theta^*}(\mathbf{z}|\mathbf{x}) = r_{\varphi^*}(\mathbf{z}|\mathbf{x})$ となる。（証明の最後の部分で＝）
- その場合、$-\left\langle \log \frac{q_{\theta^*}(\mathbf{x}|\mathbf{z})q_Z(\mathbf{z})}{r_{\varphi^*}(\mathbf{z}|\mathbf{x})} \right\rangle_{r_{\varphi^*}(\mathbf{z}|\mathbf{x})} = -\log q_{\theta^*}(\mathbf{x})$

このため「近似的に密度関数 $q(\mathbf{x})$ を計算できるモデル」と言える

3. 潜在変数モデル

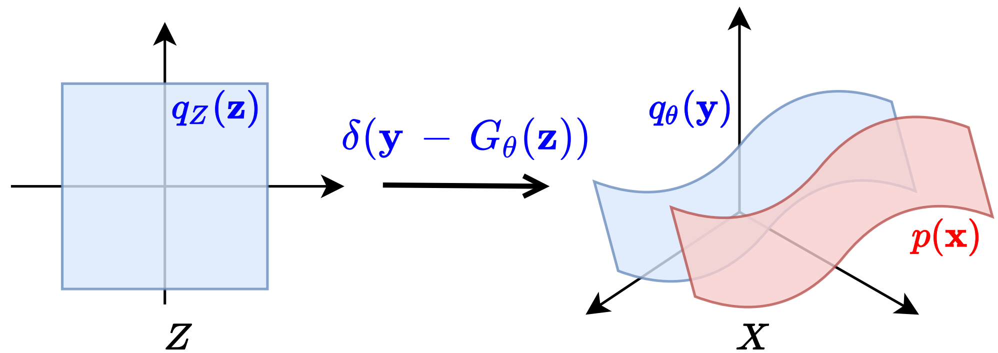
■ 3-3. 敵対的生成ネットワーク (GAN)

（たぶん良い）レビュー論文：https://arxiv.org/abs/2001.06937

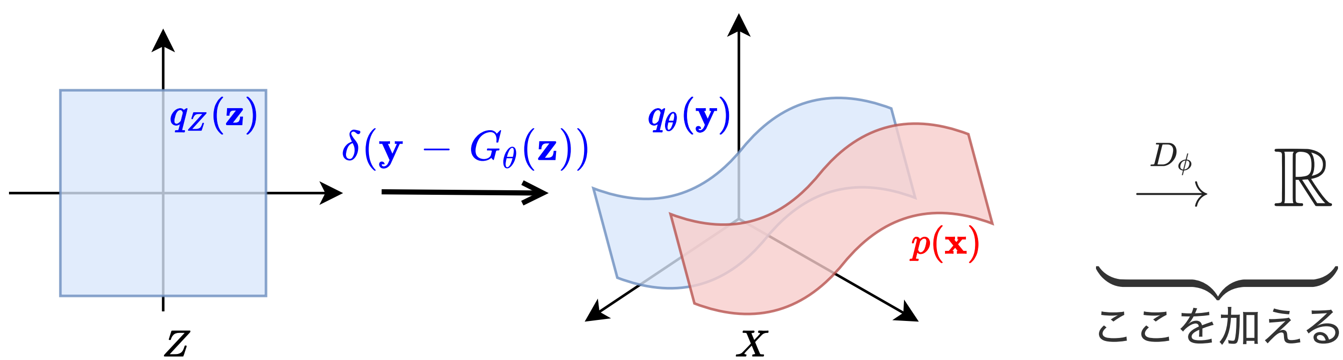


3. 潜在変数モデル

■ 3-3. 敵対的生成ネットワーク (GAN)

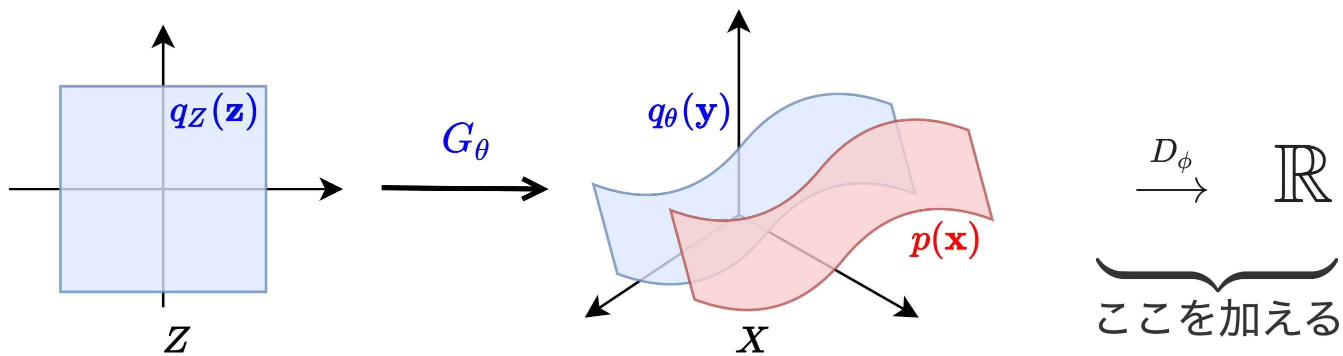


X 上の 密度：$q_\theta(\mathbf{y}) = \int_Z \delta(\mathbf{y} - G_\theta(\mathbf{z}))q_Z(\mathbf{z})d\mathbf{z}$ ➡ $\min_\theta D_{KL}(p \| q_\theta)$ は難



3. 潜在変数モデル

■ 3-3. 敵対的生成ネットワーク (GAN)



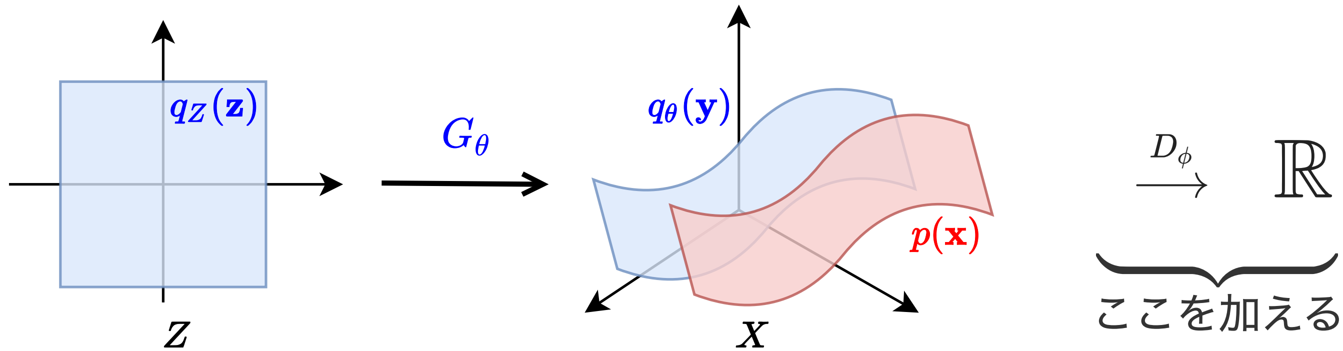
Jensen-Shannon (JS) ダイバージェンス の 汎関数表示 (→ 証明)

$\sigma(x) = \frac{1}{1+e^{-x}}$ として

$$\begin{aligned} D_{JS}(p, q_\theta) &:= D_{KL}\left(p \middle\| \frac{p+q_\theta}{2}\right) + D_{KL}\left(q \middle\| \frac{p+q_\theta}{2}\right) \\ &= \max_D \Big(\big\langle \log \sigma\big(D(\mathbf{x})\big) \big\rangle_{p(\mathbf{x})} + \big\langle \log \big\{1 - \sigma\big(D(\mathbf{y})\big)\big\} \big\rangle_{q_\theta(\mathbf{y})} \Big) + 2 \log 2 \end{aligned}$$

3. 潜在変数モデル

■ 3-3. 敵対的生成ネットワーク (GAN)



GANの目的

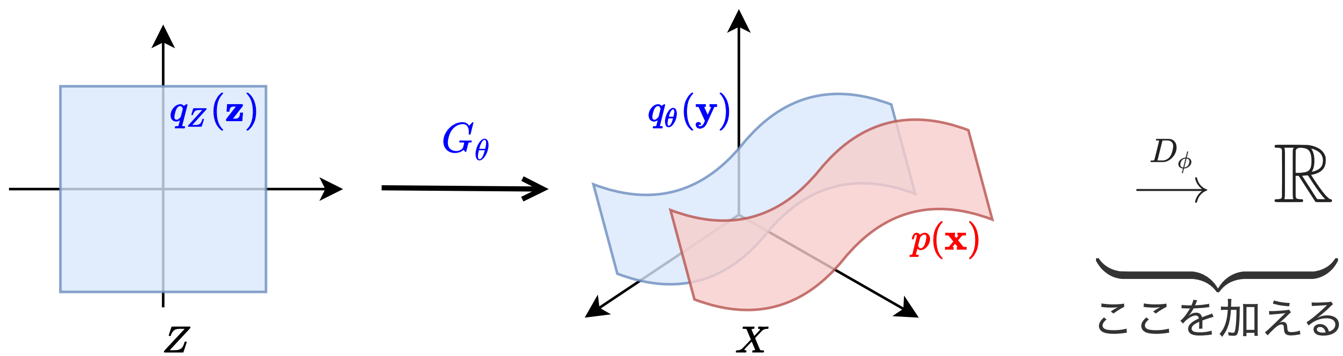
$\sigma(x) = \frac{1}{1+e^{-x}}$ として

$$\min_{\theta} \max_{\phi} \left(\left\langle \log \sigma\left(D_\phi(\mathbf{x})\right) \right\rangle_{p(\mathbf{x})} + \left\langle \log \left\{ 1 - \sigma\left(D_\phi(\mathbf{y})\right) \right\} \right\rangle_{q_\theta(\mathbf{y})} \right)$$

※ 実際にはやはり 勾配更新 を行う（次ページ）

3. 潜在変数モデル

■ 3-3. 敵対的生成ネットワーク (GAN)



GANの勾配更新

$$\theta_{t+1} = \theta_t + \epsilon_t \nabla_\theta \left(\left\langle \log \sigma\left(D_\phi(\mathbf{x})\right) \right\rangle_{p(\mathbf{x})} + \left\langle \log \left\{ 1 - \sigma\left(D_\phi(\mathbf{y})\right) \right\} \right\rangle_{q_\theta(\mathbf{y})} \right)$$

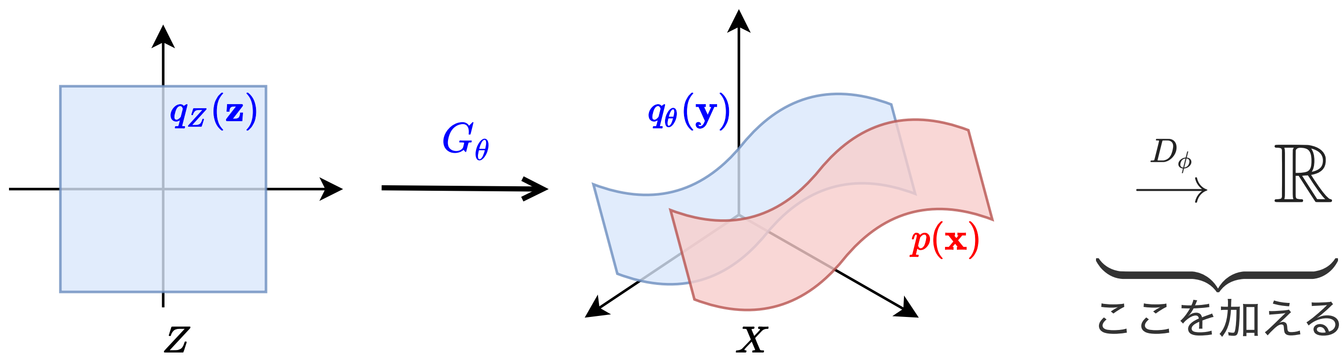
$$\phi_{t+1} = \phi_t + \eta_t \nabla_\phi \left(\left\langle \log \sigma\left(D_\phi(\mathbf{x})\right) \right\rangle_{p(\mathbf{x})} + \left\langle \log \left\{ 1 - \sigma\left(D_\phi(\mathbf{y})\right) \right\} \right\rangle_{q_\theta(\mathbf{y})} \right)$$

※ 実際には 期待値は サンプリング で近似する

※ $\min_\theta \max_\phi$ 問題がこれで解けるためには $\left(\dots \right)$ が 鞍点型 である必要がある

3. 潜在変数モデル

■ 3-3. 敵対的生成ネットワーク (GAN)



幾つかのコメント

- 理想的なケースでは $e^{D^*(\mathbf{x})} = \frac{p(\mathbf{x})}{q_\theta(\mathbf{x})}$ となる (証明の中盤)
- つまり $q_\theta(\mathbf{x})$ そのものにはアクセスできないが、ターゲットとの密度比の近似はできると考えられる
- Metropolis-Hastings テストなどを近似するなど：https://arxiv.org/abs/1811.11357

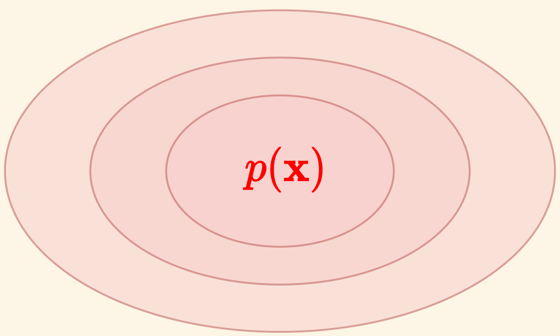
4. 拡散モデル

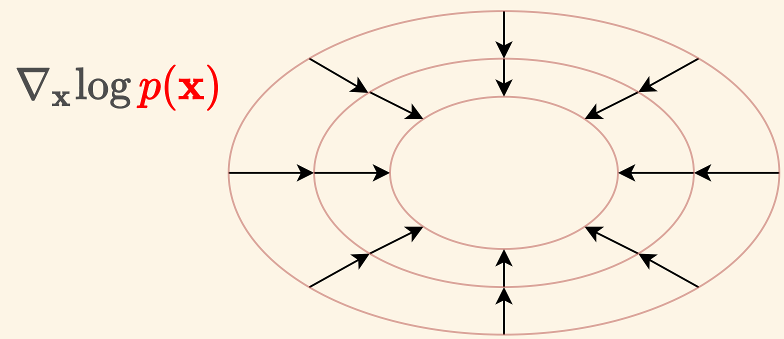
- スコアマッチング，デノイジング スコアマッチング
- DDPM
- SDE
- フローマッチング

4. 拡散モデル

■ 4-1. スコアマッチング

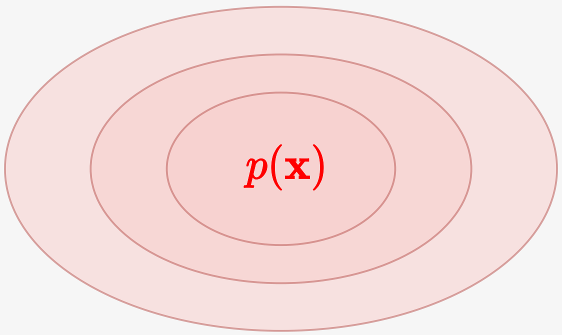
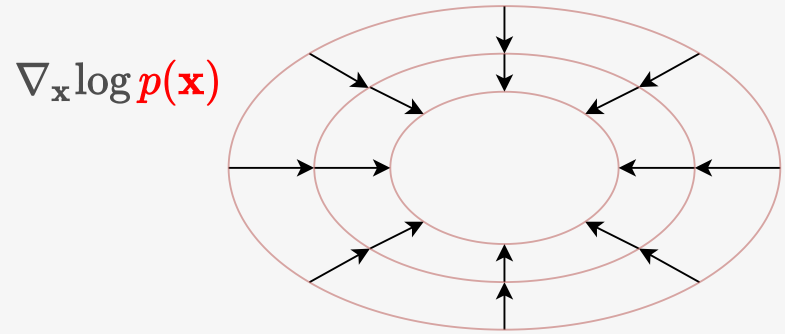
原論文： https://jmlr.org/papers/v6/hyvarinen05a.html





4. 拡散モデル

■ 4-1. スコアマッチング

密度関数	スコア※
	

スコアがわかれば、適当な $\mathbf{x}_0$ から Langevin MC

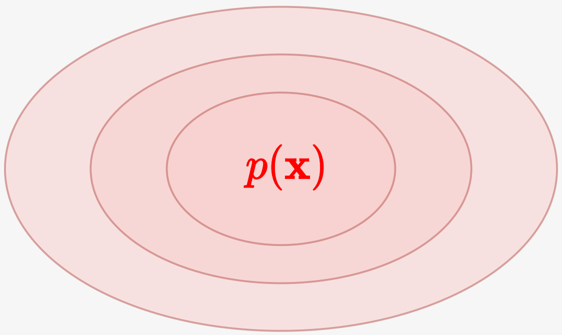
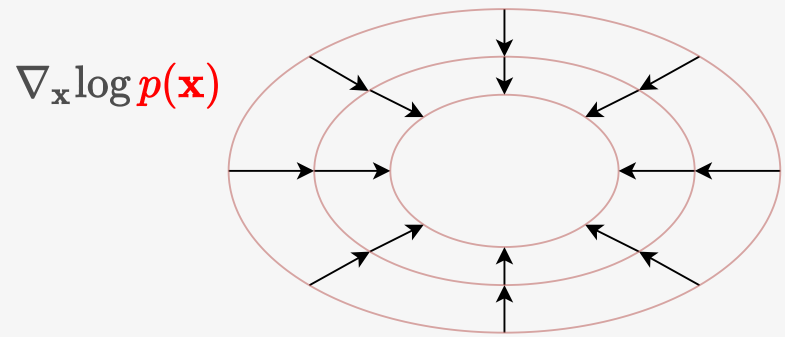
$$\mathbf{x}_t = \mathbf{x}_{t-1} + \epsilon \underbrace{\nabla_{\mathbf{x}} \log p(\mathbf{x}_{t-1})}_{s_\theta(\mathbf{x}_{t-1})\text{でモデル化}} + \sqrt{2\epsilon}\mathbf{z}_t, \quad \mathbf{z}_t \sim \mathcal{N}(0, I)$$

$\Rightarrow$（十分小さな ϵ で） $\mathbf{x}_\infty \sim p(\mathbf{x})$ (→証明)

※ 本来スコアは パラメトリックな確率密度関数 $p_\theta(\mathbf{x})$ の パラメータ微分値 $\nabla_\theta \log p_\theta(\mathbf{x})$ のことを指す？

4. 拡散モデル

■ 4-1. スコアマッチング

密度関数	スコア※
	

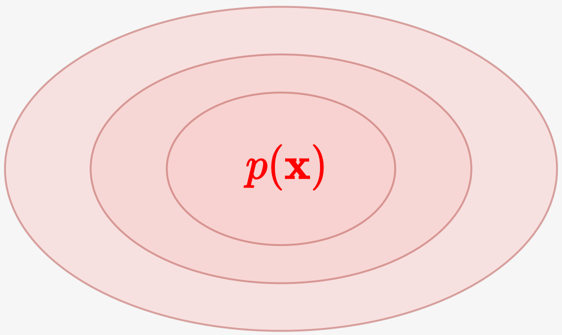
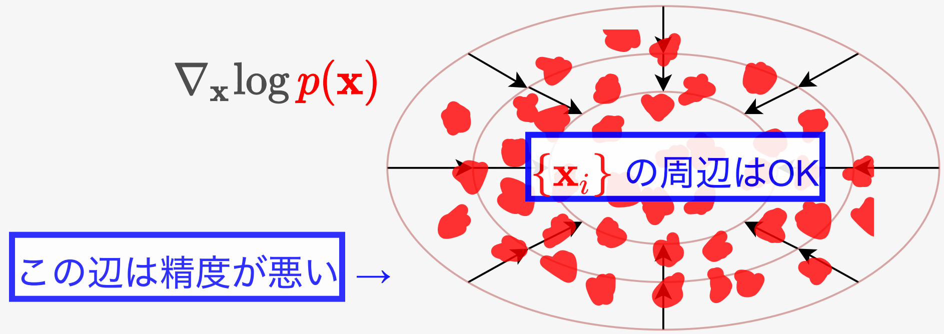
スコアマッチングの目的 (→証明)

$$\min_{\theta} \underbrace{\left\langle \left(s_\theta(\mathbf{x}) - \nabla_{\mathbf{x}} \log p(\mathbf{x}) \right)^2 \right\rangle_{p(\mathbf{x})}}_{\left\langle s_\theta(\mathbf{x})^2 + 2\nabla_{\mathbf{x}} \cdot s_\theta(\mathbf{x}) \right\rangle_{p(\mathbf{x})} + \text{const}}$$

$\underbrace{\cdots}$ の変形のおかげで、期待値をサンプル平均に置き換えられる

4. 拡散モデル

■ 4-1. スコアマッチング

密度関数	スコア※
	

実際にやること

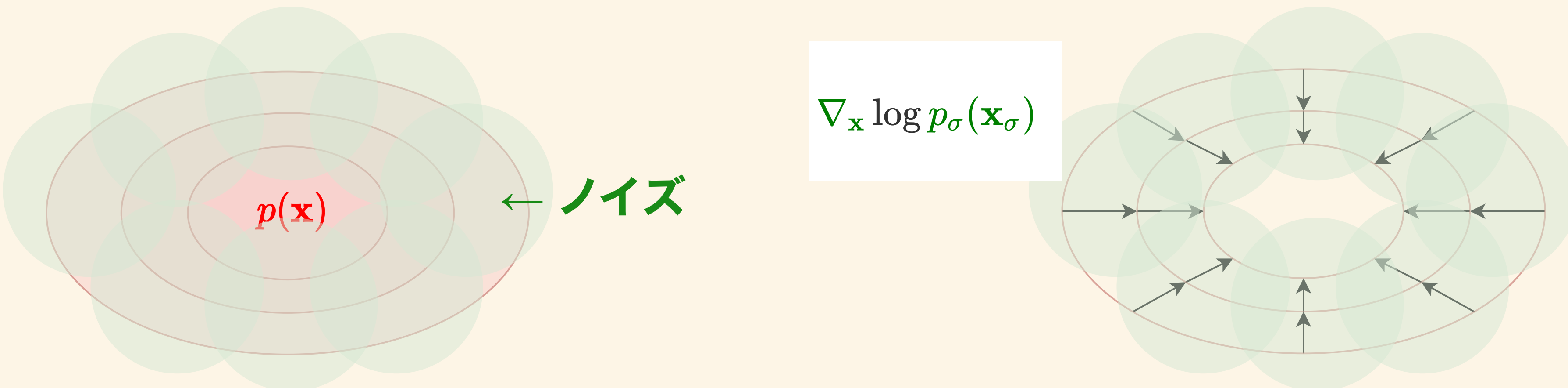
$$\min_{\theta} \frac{1}{N_{\mathrm{data}}} \sum_{i=1}^{N_{\mathrm{data}}} \left(s_{\theta}(\mathbf{x}_i)^2 + 2\nabla_{\mathbf{x}} \cdot s_{\theta}(\mathbf{x}_i) \right)$$

ただし、データサンプルがない点での精度の保証ができない（上図を参照）

➡ Langevin サンプリングの初期値によっては問題

4. 拡散モデル

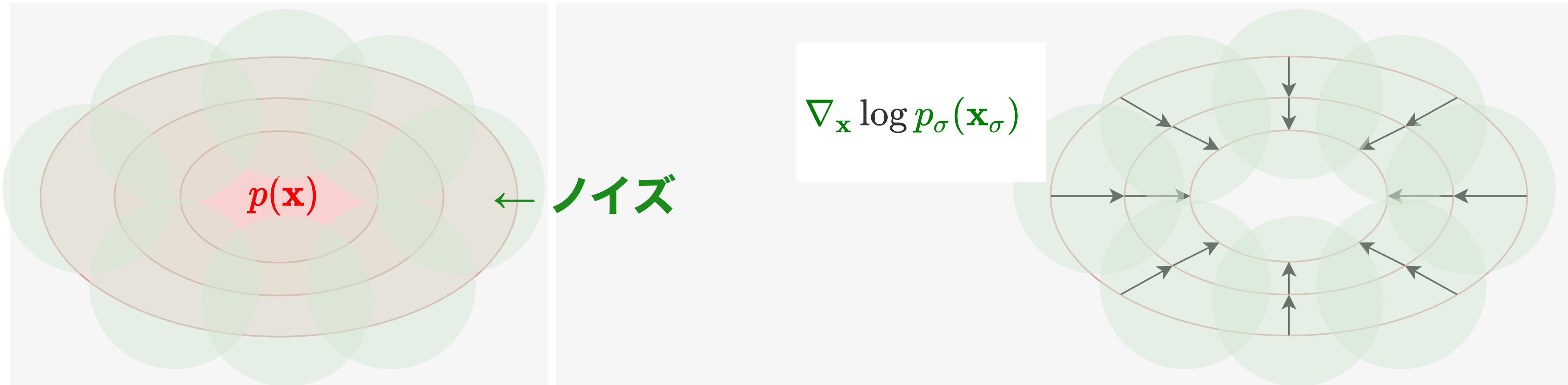
■ 4-2. デノイジング スコアマッチング



4. 拡散モデル

■ 4-2. デノイジング スコアマッチング

密度関数	スコア※



ターゲットに **ノイズ** を振る：

$$p_\sigma(\mathbf{x}_\sigma) := \int \underbrace{p_\sigma(\mathbf{x}_\sigma|\mathbf{x})}_{\text{ノイズ}} p(\mathbf{x}) d\mathbf{x}$$

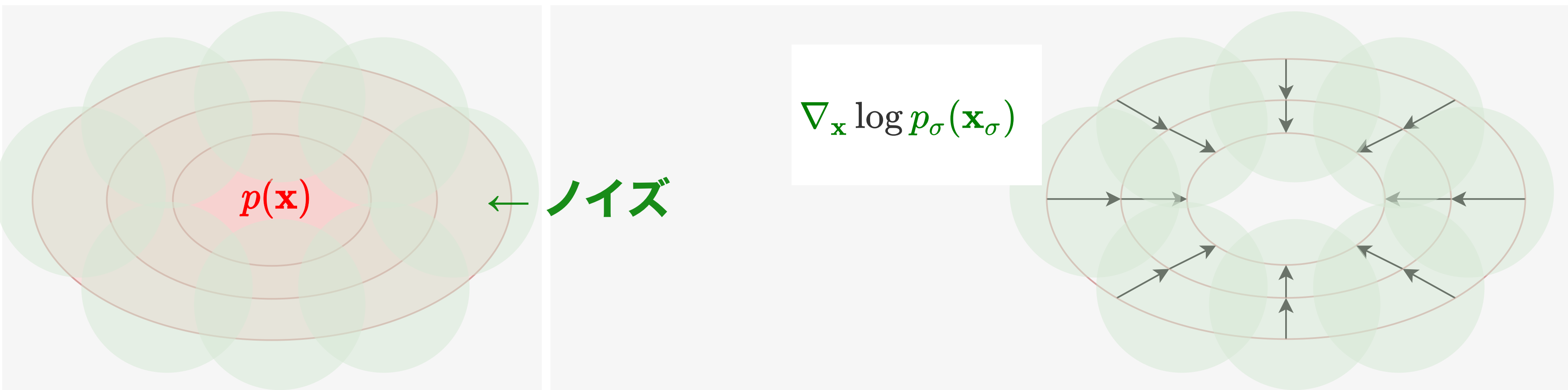
分布が広がるので、より広い領域のベクトル場を推定できる

➡ $p_\sigma(\mathbf{x}_\sigma)$ のスコア推定 をする

4. 拡散モデル

■ 4-2. デノイジング スコアマッチング

密度関数	スコア※



デノイジング スコアマッチングの目的 (→証明) （原論文：https://doi.org/10.1162/NECO_a_00142 ）

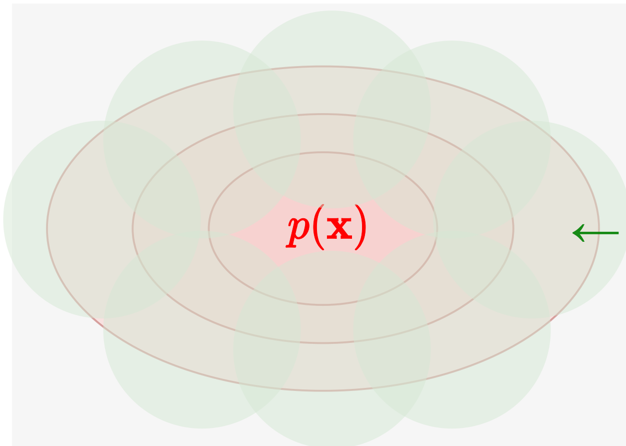
$p_\sigma(\mathbf{x}_\sigma) := \int p_\sigma(\mathbf{x}_\sigma|\mathbf{x})p(\mathbf{x})d\mathbf{x}$ として、

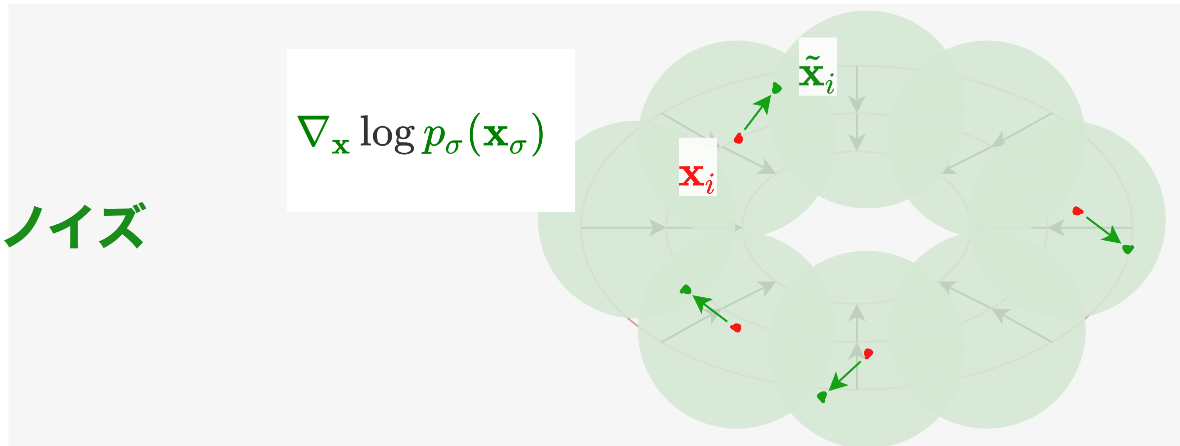
$$\min_{\theta} \underbrace{\left\langle \left(s_\theta(\mathbf{x}_\sigma) - \nabla_{\mathbf{x}} \log p_\sigma(\mathbf{x}_\sigma) \right)^2 \right\rangle_{p_\sigma(\mathbf{x}_\sigma)}}_{\left\langle \left(s_\theta(\tilde{\mathbf{x}}) - \nabla_{\mathbf{x}} \log p_\sigma(\mathbf{x}_\sigma|\mathbf{x}) \right)^2 \right\rangle_{p_\sigma(\mathbf{x}_\sigma|\mathbf{x})p(\mathbf{x})} + \text{const}}$$

4. 拡散モデル

■ 4-2. デノイジング スコアマッチング

密度関数	スコア※





実際にやること

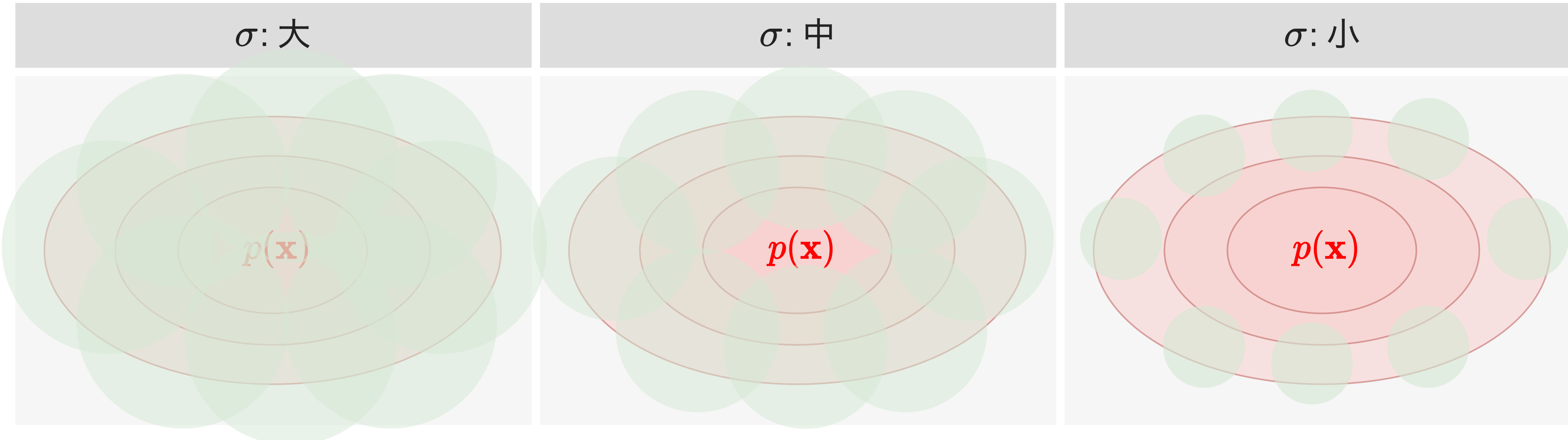
$$\min_{\theta} \frac{1}{N_{\mathrm{data}}} \sum_{i=1}^{N_{\mathrm{data}}} \left(s_\theta(\tilde{\mathbf{x}}_i) - \nabla_{\tilde{\mathbf{x}}} \log p_\sigma(\tilde{\mathbf{x}}_i | \mathbf{x}_i) \right)^2$$

- ノイズの大きさ σ 大 $\Leftrightarrow$ 広い範囲で s_θ 訓練可能だが、p から遠ざかる
- ノイズの大きさ σ 小 $\Leftrightarrow$ p に近づくが、狭い範囲でのみ s_θ 訓練可能

4. 拡散モデル

■ 4-2. デノイジング スコアマッチング

ノイズをアニーリング＋モデルに σ 依存性をつける



改良版（原論文： https://arxiv.org/abs/1907.05600 **）**

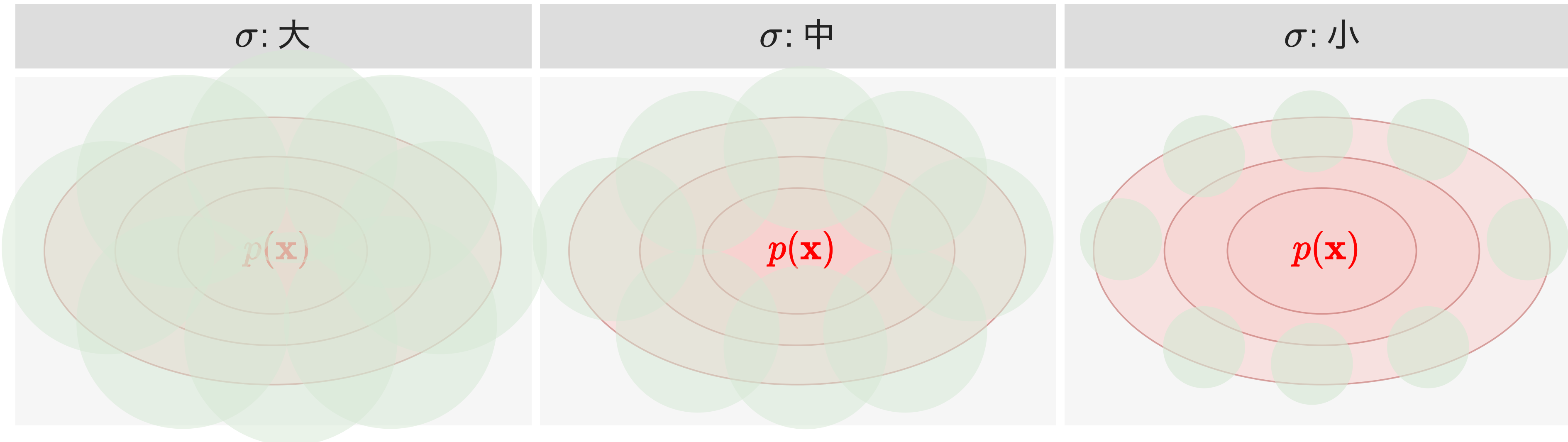
$p_\sigma(\mathbf{x}_\sigma) := \int p_\sigma(\mathbf{x}_\sigma|\mathbf{x})p(\mathbf{x})d\mathbf{x}$ として、

$$\min_\theta \int d\sigma \underbrace{\lambda(\sigma)}_{\text{重み}} \left\langle \left(s_\theta(\mathbf{x}_\sigma, \underbrace{\sigma}_{\sigma\text{ 依存性を持たせる}}) - \nabla_\mathbf{x} \log p_\sigma(\mathbf{x}_\sigma|\mathbf{x}) \right)^2 \right\rangle_{p_\sigma(\mathbf{x}_\sigma|\mathbf{x})p(\mathbf{x})}$$

4. 拡散モデル

■ 4-2. デノイジング スコアマッチング

ノイズをアニーリング＋モデルに σ 依存性をつける



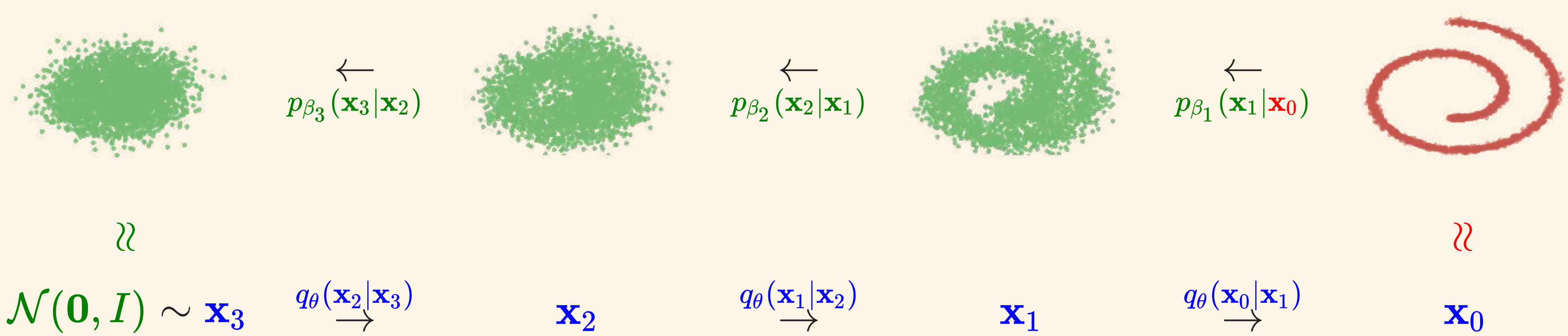
訓練後のデータサンプリング $\xrightarrow{s_\theta(\sigma)}$ は Langevin MC を表す。

$$
\begin{array}{ccccccccccccccccc}
\underset{\text{固定分布}}{\pi(\tilde{\mathbf{x}})} & \sim & \tilde{\mathbf{x}}_0 & \xrightarrow{s_\theta(\sigma_L)} & \tilde{\mathbf{x}}_{T_L} & \xrightarrow{s_\theta(\sigma_{L-1})} & \tilde{\mathbf{x}}_{T_{L-1}} & \rightarrow & \cdots & \rightarrow & \tilde{\mathbf{x}}_{T_2} & \xrightarrow{s_\theta(\sigma_1)} & \tilde{\mathbf{x}}_{T_1} & & & & \\
 & & & & \wr & & \wr & & \wr & & \wr & & \wr & & & & \\
 & & & & \mathbf{x}_{\sigma_L} & \xleftarrow{\text{ノイズ}} & \mathbf{x}_{\sigma_{L-1}} & \leftarrow & \cdots & \leftarrow & \mathbf{x}_{\sigma_2} & \xleftarrow{\text{ノイズ}} & \mathbf{x}_{\sigma_1} & \xleftarrow{\text{ノイズ}} & \mathbf{x}_0 & \sim & p(\mathbf{x}) \\
 & & & \text{分散} & \sigma_L & & \sigma_{L-1} & & & & \sigma_2 & & \sigma_1 & & & &
\end{array}
$$

4. 拡散モデル

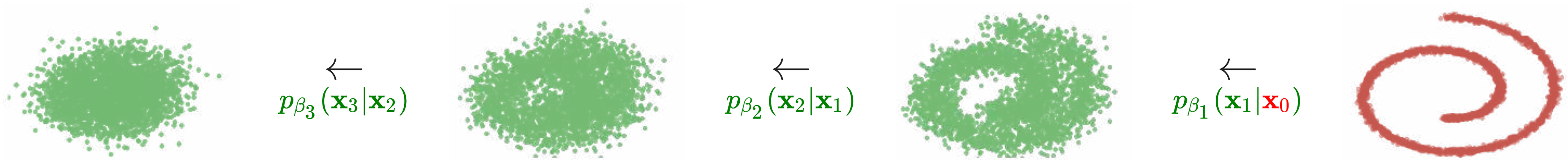
■ 4-3. デノイジング拡散確率モデル (DDPM)

原論文：https://arxiv.org/abs/2006.11239



4. 拡散モデル

■ 4-3. デノイジング拡散確率モデル (DDPM)



別のノイズのかけ方：

$$p_{\beta_t}(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{\alpha_t}\mathbf{x}_{t-1}, \beta_t I), \quad \alpha_t = 1 - \beta_t, \quad \beta_t \in (0, 1)$$

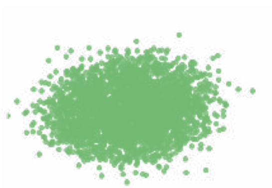
時刻 t での分布 ([→証明](#))

$$p(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}\left(\mathbf{x}_t; \sqrt{\prod_{\tau=1}^{t} \alpha_\tau}\,\mathbf{x}_0, \left(1 - \prod_{\tau=1}^{t} \alpha_\tau\right) I\right)$$

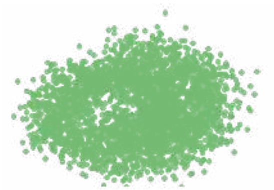
特に $\mathbf{x}_\infty \sim \mathcal{N}(\mathbf{0}, I)$ となる。

4. 拡散モデル

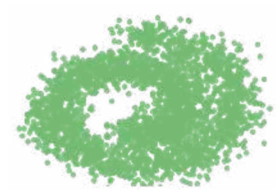
■ 4-3. デノイジング拡散確率モデル (DDPM)



$\overleftarrow{\;p_{\beta_3}(\mathbf{x}_3|\mathbf{x}_2)\;}$



$\overleftarrow{\;p_{\beta_2}(\mathbf{x}_2|\mathbf{x}_1)\;}$



$\overleftarrow{\;p_{\beta_1}(\mathbf{x}_1|\mathbf{x}_0)\;}$

$\wr$ 　　逆拡散モデル　　逆拡散モデル　　逆拡散モデル　　$\wr$

$$\mathcal{N}(\mathbf{0}, I) \sim \mathbf{x}_3 \xrightarrow{q_\theta(\mathbf{x}_2|\mathbf{x}_3)} \mathbf{x}_2 \xrightarrow{q_\theta(\mathbf{x}_1|\mathbf{x}_2)} \mathbf{x}_1 \xrightarrow{q_\theta(\mathbf{x}_0|\mathbf{x}_1)} \mathbf{x}_0$$

$$q_\theta(\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_L) := \mathcal{N}(\mathbf{x}_L; \mathbf{0}, I) q_\theta(\mathbf{x}_{L-1}|\mathbf{x}_L) \cdots q_\theta(\mathbf{x}_0|\mathbf{x}_1) \xrightarrow{\int d\mathbf{x}_1 \ldots d\mathbf{x}_L} q_\theta(\mathbf{x}_0)$$

➡ $\min_\theta D_{KL}(p \| q_\theta)$ は難

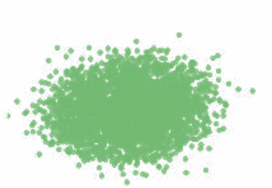
バウンド (→証明)

$p(\mathbf{x}_1, \ldots, \mathbf{x}_L|\mathbf{x}_0) := p_{\beta_L}(\mathbf{x}_L|\mathbf{x}_{L-1}) \cdots p_{\beta_1}(\mathbf{x}_1|\mathbf{x}_0)$ について：

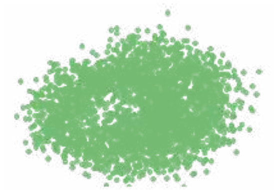
$$D_{KL}(p \| q_\theta) + \underbrace{S(p)}_{\text{エントロピー}} \leq -\left\langle \log \frac{q_\theta(\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_L)}{p(\mathbf{x}_1, \ldots, \mathbf{x}_L|\mathbf{x}_0)} \right\rangle_{p(\mathbf{x}_1, \ldots, \mathbf{x}_L|\mathbf{x}_0) p(\mathbf{x}_0)}$$

4. 拡散モデル

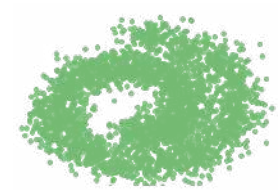
■ 4-3. デノイジング拡散確率モデル (DDPM)



$\overleftarrow{\quad}$ $p_{\beta_3}(\mathbf{x}_3|\mathbf{x}_2)$



$\overleftarrow{\quad}$ $p_{\beta_2}(\mathbf{x}_2|\mathbf{x}_1)$



$\overleftarrow{\quad}$ $p_{\beta_1}(\mathbf{x}_1|\mathbf{x}_0)$

≀ 逆拡散モデル 逆拡散モデル 逆拡散モデル ≀

$$\mathcal{N}(\mathbf{0}, I) \sim \mathbf{x}_3 \xrightarrow{q_\theta(\mathbf{x}_2|\mathbf{x}_3)} \mathbf{x}_2 \xrightarrow{q_\theta(\mathbf{x}_1|\mathbf{x}_2)} \mathbf{x}_1 \xrightarrow{q_\theta(\mathbf{x}_0|\mathbf{x}_1)} \mathbf{x}_0$$

バウンドをさらに分解できる: （この辺り煩雑なので、論文を見るか、解説記事をみてください... 🙇）

$$-\left\langle \log \frac{q_\theta(\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_L)}{p(\mathbf{x}_1, \ldots, \mathbf{x}_L|\mathbf{x}_0)} \right\rangle_{p(\mathbf{x}_1, \ldots, \mathbf{x}_L|\mathbf{x}_0)p(\mathbf{x}_0)}$$

$$= \text{定数} + \sum_{t=2}^{L} \left\langle D_{KL}(p(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) \| q_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)) \right\rangle_{p(\mathbf{x}_t|\mathbf{x}_0)p(\mathbf{x}_0)} - \left\langle \log q_\theta(\mathbf{x}_0|\mathbf{x}_1) \right\rangle_{p(\mathbf{x}_1|\mathbf{x}_0)p(\mathbf{x}_0)}$$

$$\downarrow\ q_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}\left(\mathbf{x}_{t-1}; \frac{1}{\sqrt{\alpha_t}}(\mathbf{x}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(\mathbf{x}_t, t), \sigma_t^2 I\right) \text{ にとる}$$

$$= \text{定数} + \sum_{t=2}^{L} \frac{\beta_t^2}{2\sigma_t^2(1 - \prod_{\tau=1}^{t} \alpha_\tau)} \left\langle \left(\epsilon - \epsilon_\theta(\mathbf{x}_t(\mathbf{x}_0, \epsilon, t)\right)^2 \right\rangle_{p(\mathbf{x}_0)\mathcal{N}(\epsilon;0,I)} - \left\langle \log q_\theta(\mathbf{x}_0|\mathbf{x}_1) \right\rangle_{p(\mathbf{x}_1|\mathbf{x}_0)p(\mathbf{x}_0)}$$

4. 拡散モデル

■ 4-4. 確率微分方程式による解釈

原論文： https://arxiv.org/abs/2011.13456

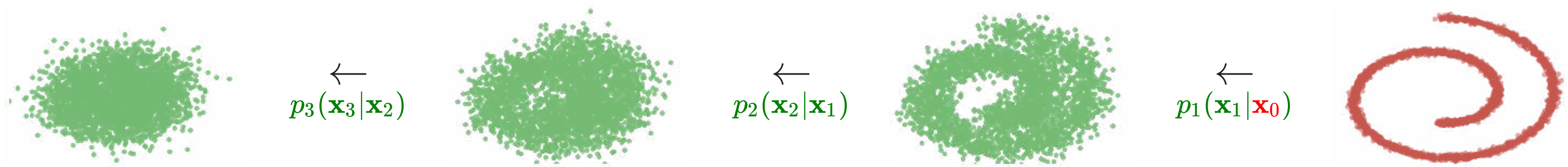
$$d\mathbf{x}(t) = \mathbf{f}\big(\mathbf{x}(t), t\big)dt + g(t)d\mathbf{w}(t)$$

$$\Downarrow$$

$$\frac{\partial p_t(\mathbf{x})}{\partial t} = -\nabla_{\mathbf{x}} \cdot (\mathbf{f}(\mathbf{x}, t)p_t(\mathbf{x})) + \frac{g(t)^2}{2}\nabla_{\mathbf{x}}^2 p_t(\mathbf{x})$$

4. 拡散モデル

■ 4-4. 確率微分方程式による解釈



ノイズをより一般化

$$p_t(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t|\mathbf{x}_{t-1} + \mathbf{f}(\mathbf{x}_{t-1}, t)\Delta t, g(t)^2\Delta t I)$$

$$\Updownarrow$$

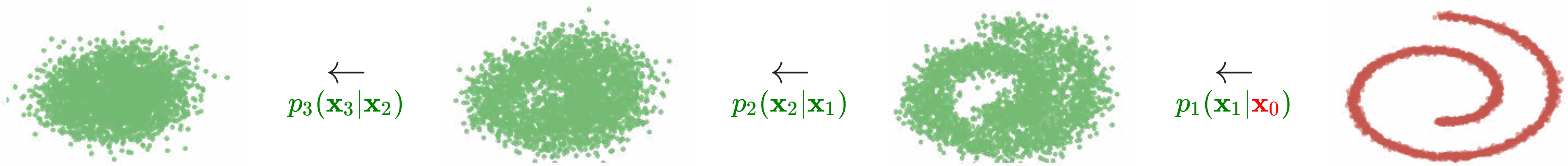
$$\mathbf{x}_t = \mathbf{x}_{t-1} + \mathbf{f}(\mathbf{x}_{t-1}, t)\Delta t + g(t)\sqrt{\Delta t}\mathbf{w}_t, \quad \mathbf{w}_t \sim \mathcal{N}(\mathbf{0}, I)$$

$\Delta t \to +0$ を取ると 確率微分方程式になる。

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

4. 拡散モデル

■ 4-4. 確率微分方程式による解釈



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

これまでの例：

- デノイジング スコアマッチング (Variance Exploding, VE)

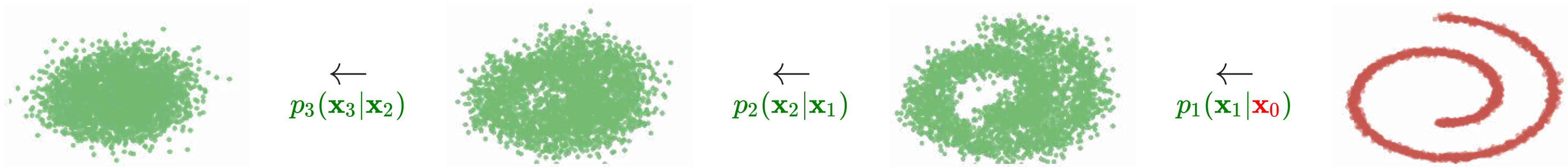
$$\mathbf{f}(\mathbf{x}, t) = \mathbf{0}, \quad g(t) = \sqrt{\frac{\sigma_t^2 - \sigma_{t-1}^2}{\Delta t}} = \sqrt{\frac{d\sigma^2(t)}{dt}}$$

- DDPM (Variance Preserving, VP)

$$\mathbf{f}(\mathbf{x}, t) = (\sqrt{1-\beta_t}\mathbf{x} - \mathbf{x})/\Delta t \approx -\frac{1}{2}\frac{\beta_t}{\Delta t}\mathbf{x}, \quad g(t) = \sqrt{\frac{\beta_t}{\Delta t}}$$

4. 拡散モデル

■ 4-4. 確率微分方程式による解釈



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

Kolmogorov forward/backward 方程式

- forward: 時刻 t の拡散された粒子の分布 $p_t(\mathbf{x})$ は以下を満たす (**Fokker-Planck 方程式**)(→証明)

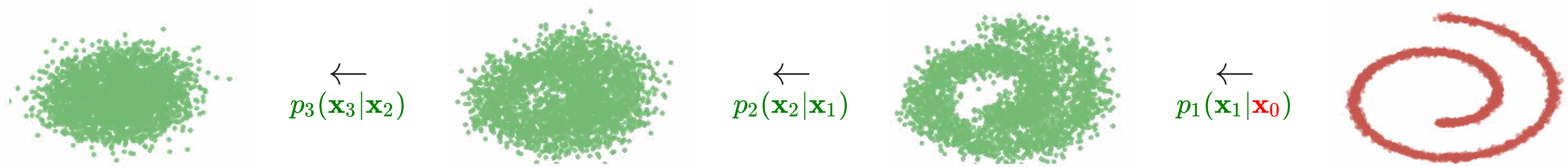
$$\frac{\partial p_t(\mathbf{x})}{\partial t} = -\nabla_{\mathbf{x}} \cdot (\mathbf{f}(\mathbf{x}, t)p_t(\mathbf{x})) + \frac{g(t)^2}{2}\nabla^2_{\mathbf{x}}p_t(\mathbf{x})$$

- backward: 時刻 t で位置 $\mathbf{x}$ にいる条件で、時刻 $s > t$ で $\mathbf{x}_s$ に見出す確率 $p_t(\mathbf{x}_s|\mathbf{x})$ は以下を満たす (→証明)

$$-\frac{\partial p_t(\mathbf{x}_s|\mathbf{x})}{\partial t} = \mathbf{f}(\mathbf{x}, t) \cdot \nabla_{\mathbf{x}}p_t(\mathbf{x}_s|\mathbf{x}) + \frac{g(t)^2}{2}\nabla^2_{\mathbf{x}}p_t(\mathbf{x}_s|\mathbf{x})$$

4. 拡散モデル

■ 4-4. 確率微分方程式による解釈



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

逆拡散の確率微分方程式 (→証明)

$\tau = T - t$ として 以下を考えると逆拡散になる。

$$d\mathbf{x}(\tau) = -[\mathbf{f}(\mathbf{x}(\tau), t) - g(t)^2 \nabla_{\mathbf{x}} \log p_t(\mathbf{x}(\tau))]d\tau + g(t)d\bar{\mathbf{w}}(\tau)$$

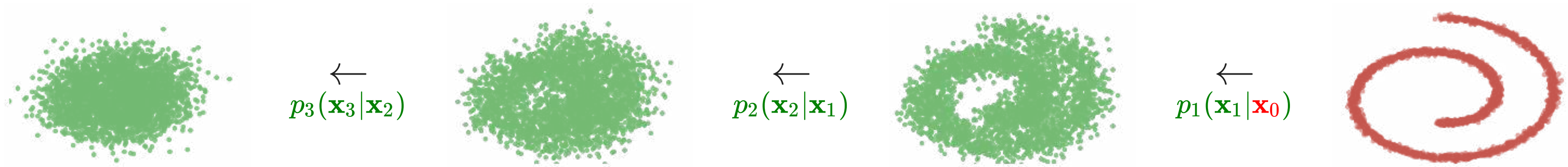
なので、スコアマッチングで $\nabla_{\mathbf{x}} \log p_t(\mathbf{x}) \approx s_\theta(\mathbf{x}, t)$ として

$$d\mathbf{x}(\tau) = -[\mathbf{f}(\mathbf{x}(\tau), t) - g(t)^2 s_\theta(\mathbf{x}(\tau), t)]d\tau + g(t)d\bar{\mathbf{w}}(\tau)$$

を解けば良い。

4. 拡散モデル

■ 4-4. 確率微分方程式による解釈



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

ところで Fokker-Planck方程式は以下のように書き直せる：

$$\frac{\partial p_t(\mathbf{x})}{\partial t} = -\nabla_{\mathbf{x}} \cdot (\mathbf{f}(\mathbf{x}, t)p_t(\mathbf{x})) + \frac{g(t)^2}{2} \underbrace{\nabla_{\mathbf{x}}^2 p_t(\mathbf{x})}_{\nabla_{\mathbf{x}} \cdot (p_t(\mathbf{x}) \nabla_{\mathbf{x}} \log p_t(\mathbf{x}))}$$

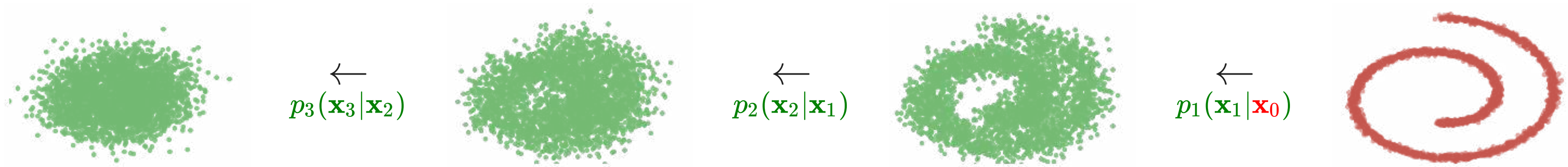
$$= -\nabla_{\mathbf{x}} \cdot \left(\left(\mathbf{f}(\mathbf{x}, t) - \frac{g(t)^2}{2} \nabla_{\mathbf{x}} \log p_t(\mathbf{x}) \right) p_t(\mathbf{x}) \right)$$

同じFP方程式を導く $\Uparrow$

$$d\mathbf{x}(t) = \left(\mathbf{f}(\mathbf{x}, t) - \frac{g(t)^2}{2} \nabla_{\mathbf{x}} \log p_t(\mathbf{x}) \right) dt$$ 常微分方程式!

4. 拡散モデル

■ 4-4. 確率微分方程式による解釈



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

逆拡散の確率フロー (→証明)

$$\mathbf{x}(T) \sim p_T \Rightarrow d\mathbf{x}(t) = \left(\mathbf{f}(\mathbf{x}, t) - \frac{g(t)^2}{2} s(\mathbf{x}, t)\right) dt$$

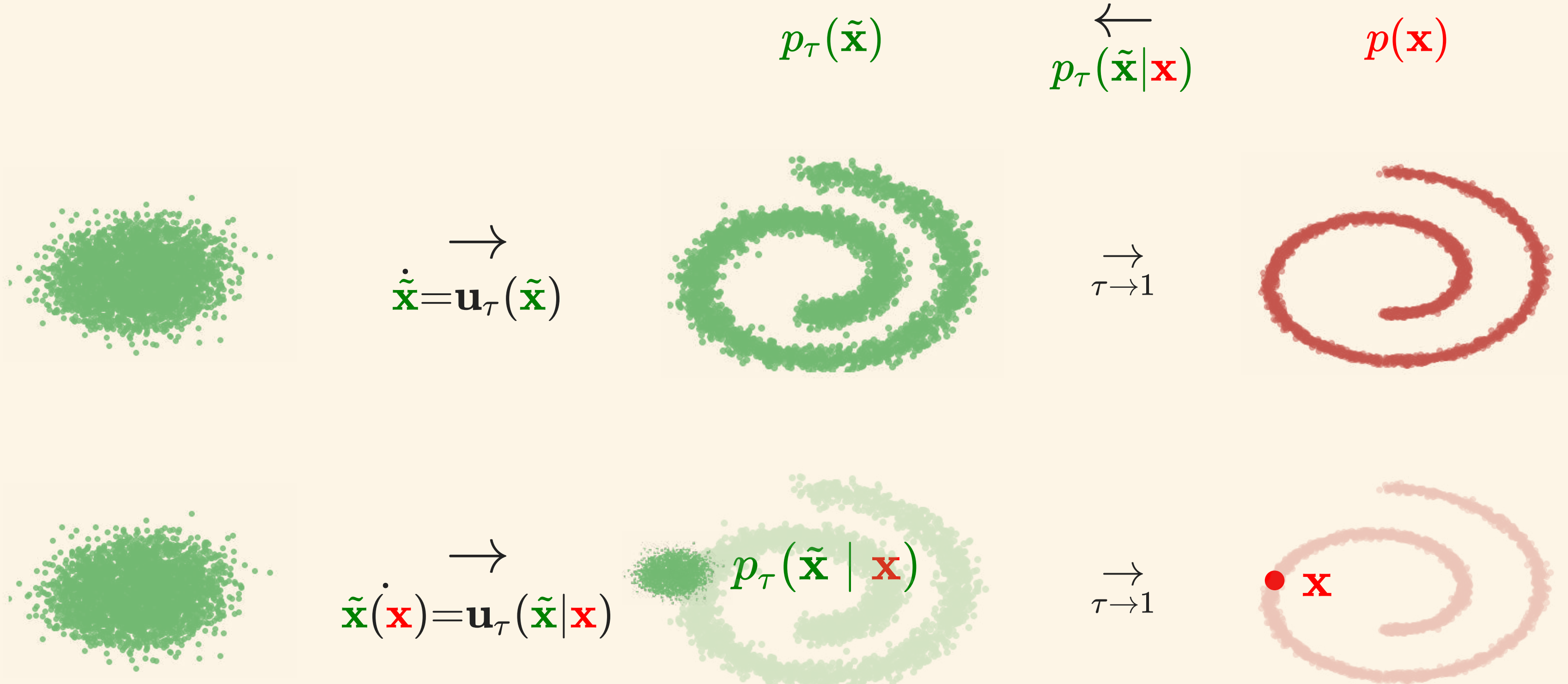
（ここでは時間座標が t なのに注意）この時モデルの密度関数が書き下せる：

$$\log p_0(\mathbf{x}) = \log p_T(\mathbf{x}(T)) + \int_0^T \nabla_{\mathbf{x}} \cdot \left(\mathbf{f}(\mathbf{x}(t), t) - \frac{g(t)^2}{2} s(\mathbf{x}(t), t)\right) dt$$

4. 拡散モデル

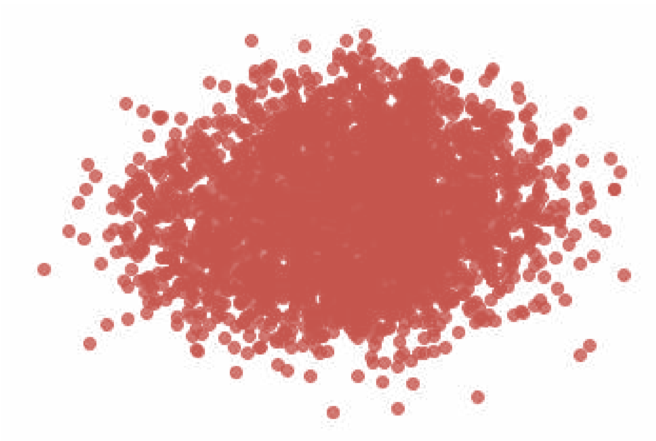
■ 4-5. フローマッチング

原論文： https://arxiv.org/abs/2210.02747



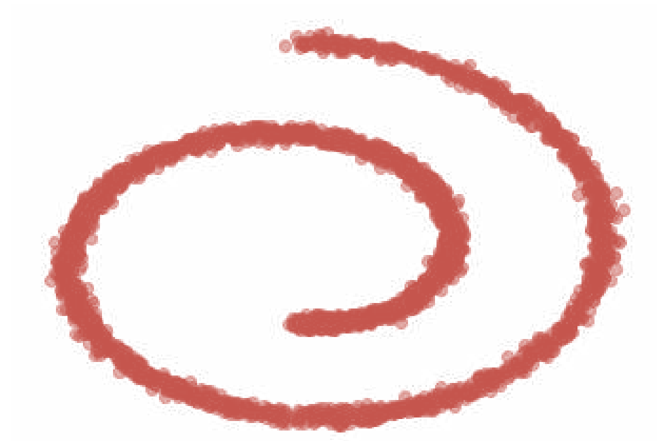
4. 拡散モデル

■ 4-5. フローマッチング



$$\xrightarrow{}$$

$$\dot{\mathbf{x}}(\tau)=\mathbf{u}_\tau(\mathbf{x}(\tau))$$



ODE からのアプローチ

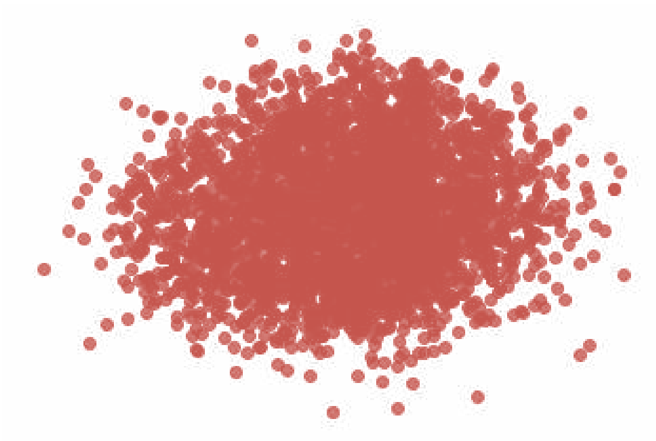
連続の方程式 (→証明)

$\mathbf{x}(0) \sim p_0(\mathbf{x}), \quad \dot{\mathbf{x}}(\tau) = \mathbf{u}_\tau(\mathbf{x}(\tau))$ で出来る分布 $p_\tau(\mathbf{x})$ は以下を満たす：

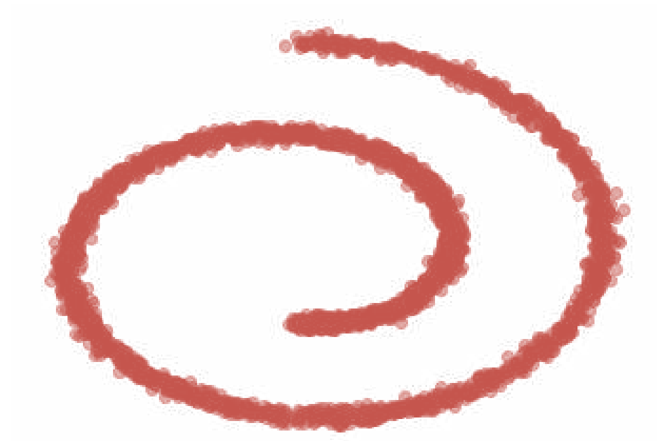
$$\frac{\partial p_\tau(\mathbf{x})}{\partial \tau} + \nabla \cdot (p_\tau(\mathbf{x})\mathbf{u}_\tau(\mathbf{x})) = 0$$

4. 拡散モデル

■ 4-5. フローマッチング



$$\overrightarrow{\dot{\mathbf{x}}(\tau)=\mathbf{u}_\tau(\mathbf{x}(\tau))}$$



なので $\mathbf{u}_\tau$ が学習できれば良い：

$$\left\langle \left(\color{blue}{\mathbf{v}_{\tau,\theta}}\color{blue}{(}\color{red}{\mathbf{x}}\color{blue}{)} - \mathbf{u}_\tau(\color{red}{\mathbf{x}})\right)^2 \right\rangle_{\tau\sim U[0,1],\color{red}{\mathbf{x}\sim p_\tau}}$$

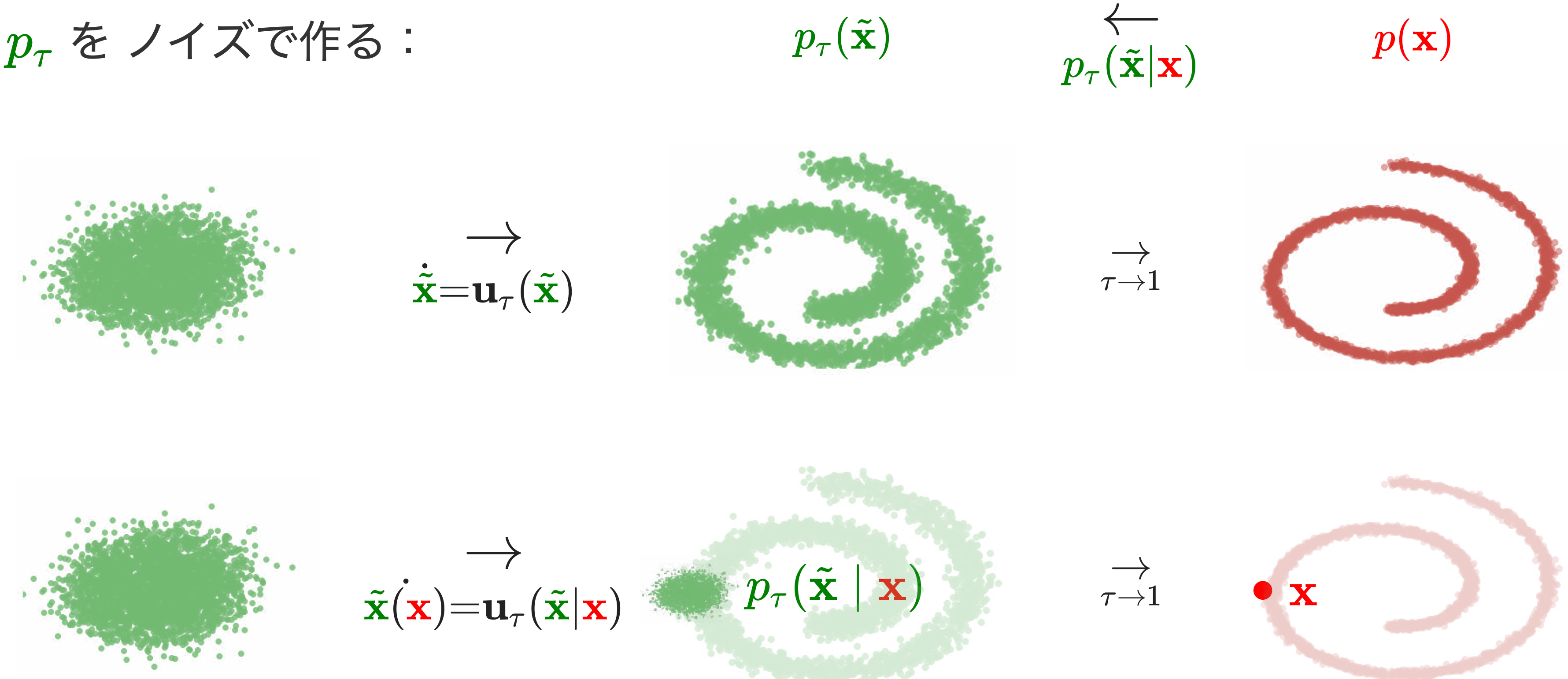
ただし、 $\color{red}{p_1 = p}$ (ターゲット分布)となるような $\mathbf{u}_\tau$ ...

絵に描いた餅：そのような $\mathbf{u}_\tau$ がわかれば 学習の必要などない。

4. 拡散モデル

■ 4-5. フローマッチング

p_τ を ノイズで作る：



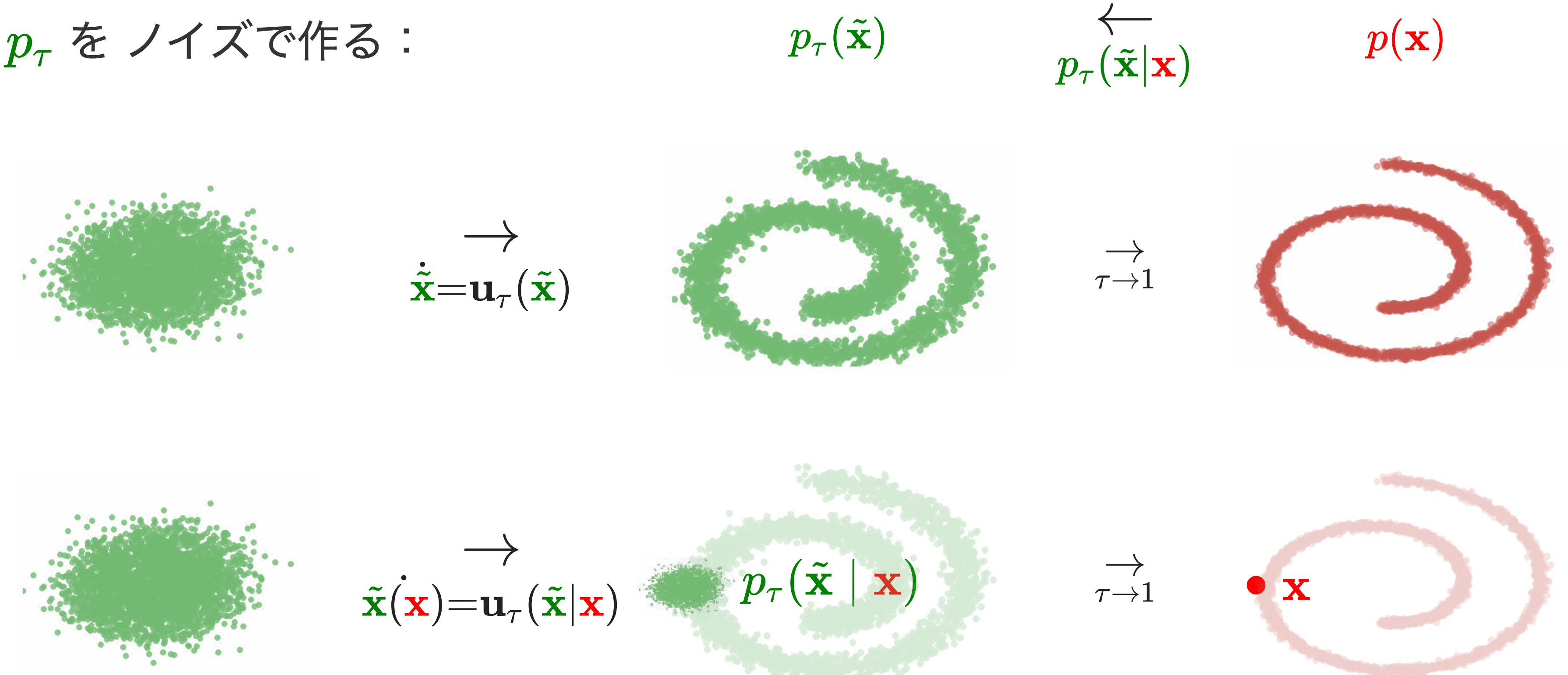
条件付きフロー (→証明)

$$p_\tau(\tilde{\mathbf{x}}|\mathbf{x}) = \mathcal{N}(\tilde{\mathbf{x}}|\mu_\tau(\mathbf{x}), \sigma_\tau(\mathbf{x})^2 I) \quad \text{ならば} \quad \mathbf{u}_\tau(\tilde{\mathbf{x}}|\mathbf{x}) = \frac{\dot{\sigma}_\tau(\mathbf{x})}{\sigma_\tau(\mathbf{x})}(\tilde{\mathbf{x}} - \mu_\tau(\mathbf{x})) + \dot{\mu}_\tau(\mathbf{x})$$

4. 拡散モデル

■ 4-5. フローマッチング

p_τ を ノイズで作る：



ノイズ入りフローマッチング (→証明)

$$\left\langle \left(\mathbf{v}_{\tau,\theta}(\tilde{\mathbf{x}}) - \mathbf{u}_\tau(\tilde{\mathbf{x}})\right)^2 \right\rangle_{\tau, p_\tau(\tilde{\mathbf{x}})} = \left\langle \left(\mathbf{v}_{\tau,\theta}(\tilde{\mathbf{x}}) - \mathbf{u}_\tau(\tilde{\mathbf{x}}|\mathbf{x})\right)^2 \right\rangle_{\tau, p_\tau(\tilde{\mathbf{x}}|\mathbf{x})p(\mathbf{x})} + \text{const.}$$

終わり

Appendix

■ 変分バウンドの証明

$$
\begin{aligned}
D_{KL}(p\|q_\theta) + S(p) &= -\langle \log q_\theta(\mathbf{x}) \rangle_{p(\mathbf{x})} \\
&= -\langle \log \underbrace{q_\theta(\mathbf{x})}_{\frac{q_\theta(\mathbf{x}|\mathbf{z})q_Z(\mathbf{z})}{q_\theta(\mathbf{z}|\mathbf{x})}} \rangle_{p(\mathbf{x})r(\mathbf{z}|\mathbf{x})} \\
&= -\left\langle \log \frac{q_\theta(\mathbf{x}|\mathbf{z})q_Z(\mathbf{z})}{q_\theta(\mathbf{z}|\mathbf{x})} \frac{r(\mathbf{z}|\mathbf{x})}{r(\mathbf{z}|\mathbf{x})} \right\rangle_{p(\mathbf{x})r(\mathbf{z}|\mathbf{x})} \\
&= -\left\langle \log \frac{q_\theta(\mathbf{x}|\mathbf{z})q_Z(\mathbf{z})}{r(\mathbf{z}|\mathbf{x})} \frac{r(\mathbf{z}|\mathbf{x})}{q_\theta(\mathbf{z}|\mathbf{x})} \right\rangle_{p(\mathbf{x})r(\mathbf{z}|\mathbf{x})} \\
&= -\left\langle \log \frac{q_\theta(\mathbf{x}|\mathbf{z})q_Z(\mathbf{z})}{r(\mathbf{z}|\mathbf{x})} \right\rangle_{p(\mathbf{x})r(\mathbf{z}|\mathbf{x})} - \langle D_{KL}(r(\mathbf{z}|\mathbf{x})\|q_\theta(\mathbf{z}|\mathbf{x})) \rangle_{p(\mathbf{x})} \\
&\leq -\left\langle \log \frac{q_\theta(\mathbf{x}|\mathbf{z})q_Z(\mathbf{z})}{r(\mathbf{z}|\mathbf{x})} \right\rangle_{p(\mathbf{x})r(\mathbf{z}|\mathbf{x})}
\end{aligned}
$$

↩ バウンドの式に戻る

↩ コメントに戻る

■ JSダイバージェンス汎関数表示の証明

1. $\log \sigma(x)$ や $\log\{1-\sigma(x)\}$ が x について上に凸
2. なので、D について変分をとって0とおけば $\max_D$ での値が求まる：

$$
\begin{aligned}
V(D) &:= \Big\langle \log \sigma\Big(D(\mathbf{x})\Big)\Big\rangle_{p(\mathbf{x})} + \Big\langle \log \Big\{1-\sigma\Big(D(\mathbf{y})\Big)\Big\}\Big\rangle_{q_\theta(\mathbf{y})} \\
&= \int_X \Big\{ p(\mathbf{x}) \log \frac{1}{1+e^{-D(\mathbf{x})}} + q_\theta(\mathbf{x}) \log \frac{1}{1+e^{+D(\mathbf{x})}} \Big\} d\mathbf{x} \\
&\Downarrow \\
\delta V(D) &= \int_X \frac{\delta D(\mathbf{x})}{1+e^{+D(\mathbf{x})}} \Big\{ p(\mathbf{x}) - q_\theta(\mathbf{x}) e^{+D(\mathbf{x})} \Big\} d\mathbf{x} \quad \Rightarrow e^{D^*(\mathbf{x})} = \frac{p(\mathbf{x})}{q_\theta(\mathbf{x})} \\
&\Downarrow \\
\max_D V(D) &= V(D^*) = \int_X \Big\{ p(\mathbf{x}) \log \frac{1}{1+\frac{q_\theta(\mathbf{x})}{p(\mathbf{x})}} + q_\theta(\mathbf{x}) \log \frac{1}{1+\frac{p(\mathbf{x})}{q_\theta(\mathbf{x})}} \Big\} d\mathbf{x} \\
&= D_{JS}(p, q_\theta) - 2\log 2
\end{aligned}
$$

↩ 汎関数表示に戻る

↩ コメントに戻る

■ **ランジュバンMCの平衡分布が $p(\mathbf{x})$ であること**

1. $p(\mathbf{x}) = \frac{e^{-E(\mathbf{x})}}{Z}$ と表現しても一般性は失いません。この時

$$\nabla_{\mathbf{x}} \log p(\mathbf{x}) = -\nabla_{\mathbf{x}} E(\mathbf{x})$$

2. ランジュバンダイナミクスの式は

$$\mathbf{x}_t = \underbrace{\mathbf{x}_{t-1} - \epsilon \nabla_{\mathbf{x}} E(\mathbf{x})}_{\text{平均}} + \sqrt{\underbrace{2\epsilon}_{\text{分散}}}\, \mathbf{z}_t, \quad \mathbf{z}_t \sim \mathcal{N}(0, I)$$

$$\Leftrightarrow \mathbf{x}_t \sim p_\epsilon(\mathbf{x}_t | \mathbf{x}_{t-1}) \propto \exp\left(-\frac{1}{2(\underbrace{2\epsilon}_{\text{分散}})} \left(\mathbf{x}_t - (\underbrace{\mathbf{x}_{t-1} - \epsilon \nabla_{\mathbf{x}} E(\mathbf{x})}_{\text{平均}})\right)^2 \right)$$

3. p_ϵ は $\epsilon \to 0$ で 平衡分布が $\frac{e^{-E(\mathbf{x})}}{Z}$ の時の詳細釣り合い条件を満たす

（次ページに続く）

■ ランジュバンMCの平衡分布が $p(\mathbf{x})$ であること

$$
\begin{aligned}
&\frac{p_\epsilon(\mathbf{x}|\mathbf{y})}{p_\epsilon(\mathbf{y}|\mathbf{x})}\\
&= \exp\left[\frac{-1}{2(2\epsilon)}\left\{(\mathbf{x}-\mathbf{y}+\epsilon\nabla_{\mathbf{y}}E(\mathbf{y}))^2-(\mathbf{y}-\mathbf{x}+\epsilon\nabla_{\mathbf{x}}E(\mathbf{x}))^2\right\}\right]\\
&= \exp\left[\frac{-1}{2(2\epsilon)}\left\{(\mathbf{x}-\mathbf{y})^2-(\mathbf{y}-\mathbf{x})^2+2\epsilon\underbrace{(\mathbf{x}-\mathbf{y})(\nabla_{\mathbf{x}}E(\mathbf{x})+\nabla_{\mathbf{y}}E(\mathbf{y}))}_{(\text{cross term})}+O(\epsilon^2)\right\}\right]\\
&= \exp\left[\frac{-1}{2}\left\{(\mathbf{x}-\mathbf{y})(\nabla_{\mathbf{x}}E(\mathbf{x})+\underbrace{\nabla_{\mathbf{y}}E(\mathbf{y})}_{\nabla_{\mathbf{x}}E(\mathbf{x})+O(\mathbf{x}-\mathbf{y})})+O(\epsilon)\right\}\right]\\
&= \exp\left[-1\left\{\underbrace{(\mathbf{x}-\mathbf{y})\Big(\nabla_{\mathbf{x}}E(\mathbf{x})}_{E(\mathbf{x})-E(\mathbf{y})+O((\mathbf{x}-\mathbf{y})^2)}+O(\mathbf{x}-\mathbf{y})\Big)+O(\epsilon)\right\}\right]\\
&= \frac{e^{-E(\mathbf{x})}}{e^{-E(\mathbf{y})}}\exp\left(O(\epsilon)+\underbrace{O((\mathbf{x}-\mathbf{y})^2)}_{\text{分散程度なので } O(\epsilon)}\right)
\end{aligned}
$$

↩ ランジュバンMCの説明に戻る

■ スコアマッチング目的関数の変形

$$
\begin{aligned}
&\left\langle \left(s_\theta(\mathbf{x}) - \nabla_{\mathbf{x}} \log p(\mathbf{x}) \right)^2 \right\rangle_{p(\mathbf{x})} \\
&= \int d\mathbf{x}\, p(\mathbf{x}) \left(s_\theta(\mathbf{x}) - \nabla_{\mathbf{x}} \log p(\mathbf{x}) \right)^2 \\
&= \int d\mathbf{x}\, p(\mathbf{x}) s_\theta(\mathbf{x})^2 - 2 \int d\mathbf{x}\, \underbrace{p(\mathbf{x}) s_\theta(\mathbf{x}) \cdot \underbrace{\nabla_{\mathbf{x}} \log p(\mathbf{x})}_{\nabla_{\mathbf{x}} p(\mathbf{x})/p(\mathbf{x})}}_{\nabla_{\mathbf{x}} \cdot (s_\theta(\mathbf{x}) p(\mathbf{x})) - p(\mathbf{x}) \nabla_{\mathbf{x}} \cdot s_\theta(\mathbf{x})} + \underbrace{\int d\mathbf{x}\, p(\mathbf{x}) (\nabla_{\mathbf{x}} \log p(\mathbf{x}))^2}_{\mathrm{const}} \\
&= \int d\mathbf{x}\, p(\mathbf{x}) \left[s_\theta(\mathbf{x})^2 + 2 \nabla_{\mathbf{x}} \cdot s_\theta(\mathbf{x}) \right] - 2 \underbrace{\int d\mathbf{x}\, \nabla_{\mathbf{x}} \cdot (s_\theta(\mathbf{x}) p(\mathbf{x}))}_{\text{表面項で } 0} + \mathrm{const} \\
&= \left\langle s_\theta(\mathbf{x})^2 + 2 \nabla_{\mathbf{x}} \cdot s_\theta(\mathbf{x}) \right\rangle_{p(\mathbf{x})} + \mathrm{const}
\end{aligned}
$$

↩ スコアマッチング目的に戻る

■ デノイジング スコアマッチング目的関数の変形

$$\left\langle \left(s_\theta(\tilde{\mathbf{x}}) - \nabla_{\tilde{\mathbf{x}}} \log p_\sigma(\tilde{\mathbf{x}}) \right)^2 \right\rangle_{p_\sigma(\tilde{\mathbf{x}})}$$

$$= \int d\tilde{\mathbf{x}}\, p_\sigma(\tilde{\mathbf{x}}) \left(s_\theta(\tilde{\mathbf{x}}) - \nabla_{\tilde{\mathbf{x}}} \log p_\sigma(\tilde{\mathbf{x}}) \right)^2$$

$$= \int d\tilde{\mathbf{x}}\, p_\sigma(\tilde{\mathbf{x}}) s_\theta(\tilde{\mathbf{x}})^2 - 2 \underbrace{\int d\tilde{\mathbf{x}}\, p_\sigma(\tilde{\mathbf{x}}) s_\theta(\tilde{\mathbf{x}}) \cdot \underbrace{\nabla_{\tilde{\mathbf{x}}} \log p_\sigma(\tilde{\mathbf{x}})}_{\nabla_{\tilde{\mathbf{x}}} p_\sigma(\tilde{\mathbf{x}})/p_\sigma(\tilde{\mathbf{x}})}}_{s_\theta(\tilde{\mathbf{x}}) \cdot \nabla_{\tilde{\mathbf{x}}} p_\sigma(\tilde{\mathbf{x}})} + \underbrace{\int d\tilde{\mathbf{x}}\, p_\sigma(\tilde{\mathbf{x}}) (\nabla_{\tilde{\mathbf{x}}} \log p_\sigma(\tilde{\mathbf{x}}))^2}_{\text{const}}$$

$$= \int d\tilde{\mathbf{x}} \underbrace{p_\sigma(\tilde{\mathbf{x}})}_{\int d\mathbf{x}\, p_\sigma(\tilde{\mathbf{x}}|\mathbf{x}) p(\mathbf{x})} s_\theta(\tilde{\mathbf{x}})^2 - 2 \int d\tilde{\mathbf{x}}\, s_\theta(\tilde{\mathbf{x}}) \cdot \nabla_{\tilde{\mathbf{x}}} \underbrace{p_\sigma(\tilde{\mathbf{x}})}_{\int d\mathbf{x}\, p_\sigma(\tilde{\mathbf{x}}|\mathbf{x}) p(\mathbf{x})} + \text{const}$$

$$= \iint d\tilde{\mathbf{x}} d\mathbf{x} \left[p_\sigma(\tilde{\mathbf{x}}|\mathbf{x}) p(\mathbf{x}) s_\theta(\tilde{\mathbf{x}})^2 - 2 s_\theta(\tilde{\mathbf{x}}) \cdot \underbrace{(\nabla_{\tilde{\mathbf{x}}} p_\sigma(\tilde{\mathbf{x}}|\mathbf{x}))}_{p_\sigma(\tilde{\mathbf{x}}|\mathbf{x}) \nabla_{\tilde{\mathbf{x}}} \log p_\sigma(\tilde{\mathbf{x}}|\mathbf{x})} p(\mathbf{x}) \right] + \text{const}$$

$$= \iint d\tilde{\mathbf{x}} d\mathbf{x} p_\sigma(\tilde{\mathbf{x}}|\mathbf{x}) p(\mathbf{x}) \underbrace{\left[s_\theta(\tilde{\mathbf{x}})^2 - 2 s_\theta(\tilde{\mathbf{x}}) \cdot \nabla_{\tilde{\mathbf{x}}} \log p_\sigma(\tilde{\mathbf{x}}|\mathbf{x}) \right]}_{(s_\theta(\tilde{\mathbf{x}}) - \nabla_{\tilde{\mathbf{x}}} \log p_\sigma(\tilde{\mathbf{x}}|\mathbf{x}))^2 - \text{const}_2} + \text{const}$$

$$= \left\langle \left(s_\theta(\tilde{\mathbf{x}}) - \nabla_{\tilde{\mathbf{x}}} \log p_\sigma(\tilde{\mathbf{x}}|\mathbf{x}) \right)^2 \right\rangle_{p_\sigma(\tilde{\mathbf{x}}|\mathbf{x}) p(\mathbf{x})} + \text{const}'$$

↩ 説明に戻る

■ 拡散バウンドの証明

$$
\begin{aligned}
&\log q_\theta(\mathbf{x}_0) \\
&= \langle \log q_\theta(\mathbf{x}_0) \rangle_{p(\mathbf{x}_1,\ldots,\mathbf{x}_L|\mathbf{x}_0)} \\
&= \left\langle \log \frac{q_\theta(\mathbf{x}_0,\mathbf{x}_1,\ldots,\mathbf{x}_L)}{q_\theta(\mathbf{x}_1,\ldots,\mathbf{x}_L|\mathbf{x}_0)} \right\rangle_{p(\mathbf{x}_1,\ldots,\mathbf{x}_L|\mathbf{x}_0)} \\
&= \left\langle \log \frac{q_\theta(\mathbf{x}_0,\mathbf{x}_1,\ldots,\mathbf{x}_L)}{q_\theta(\mathbf{x}_1,\ldots,\mathbf{x}_L|\mathbf{x}_0)} \frac{p(\mathbf{x}_1,\ldots,\mathbf{x}_L|\mathbf{x}_0)}{p(\mathbf{x}_1,\ldots,\mathbf{x}_L|\mathbf{x}_0)} \right\rangle_{p(\mathbf{x}_1,\ldots,\mathbf{x}_L|\mathbf{x}_0)} \\
&= \left\langle \log \frac{q_\theta(\mathbf{x}_0,\mathbf{x}_1,\ldots,\mathbf{x}_L)}{p(\mathbf{x}_1,\ldots,\mathbf{x}_L|\mathbf{x}_0)} \frac{p(\mathbf{x}_1,\ldots,\mathbf{x}_L|\mathbf{x}_0)}{q_\theta(\mathbf{x}_1,\ldots,\mathbf{x}_L|\mathbf{x}_0)} \right\rangle_{p(\mathbf{x}_1,\ldots,\mathbf{x}_L|\mathbf{x}_0)} \\
&= \left\langle \log \frac{q_\theta(\mathbf{x}_0,\mathbf{x}_1,\ldots,\mathbf{x}_L)}{p(\mathbf{x}_1,\ldots,\mathbf{x}_L|\mathbf{x}_0)} \right\rangle_{p(\mathbf{x}_1,\ldots,\mathbf{x}_L|\mathbf{x}_0)} + \underbrace{D_{KL}(p(\mathbf{x}_1,\ldots,\mathbf{x}_L|\mathbf{x}_0) \| q_\theta(\mathbf{x}_1,\ldots,\mathbf{x}_L|\mathbf{x}_0))}_{\geq 0} \\
&\geq \left\langle \log \frac{q_\theta(\mathbf{x}_0,\mathbf{x}_1,\ldots,\mathbf{x}_L)}{p(\mathbf{x}_1,\ldots,\mathbf{x}_L|\mathbf{x}_0)} \right\rangle_{p(\mathbf{x}_1,\ldots,\mathbf{x}_L|\mathbf{x}_0)}
\end{aligned}
$$

あとはこれを $p(\mathbf{x}_0)$ で期待値を取る。

↩ 説明に戻る

■ Gaussian の マルコフ過程における補題

補題1

$$\mathbf{x} \quad \xleftarrow{\overbrace{p(\mathbf{x}|\mathbf{y})}^{\mathcal{N}(\mathbf{x};\sqrt{\alpha_x}\mathbf{y},\beta_x I)}} \quad \mathbf{y} \quad \xleftarrow{\overbrace{p(\mathbf{y}|\mathbf{z})}^{\mathcal{N}(\mathbf{y};\sqrt{\alpha_y}\mathbf{z},\beta_y I)}} \quad \mathbf{z}$$

$$\Downarrow$$

$$p(\mathbf{x}|\mathbf{z}) = \int p(\mathbf{x}|\mathbf{y})p(\mathbf{y}|\mathbf{z})d\mathbf{y} = \mathcal{N}(\mathbf{x};\sqrt{\alpha_x\alpha_y}\mathbf{z},(\alpha_x\beta_y+\beta_x)I)$$

素朴に積分を取れば平方完成で示せるが、直感的には

$$\begin{aligned} \mathbf{x} &= \sqrt{\alpha_x}\mathbf{y} + \sqrt{\beta_x}\epsilon_x, \quad \epsilon_x \sim \mathcal{N}(0,I) \\ \mathbf{y} &= \sqrt{\alpha_y}\mathbf{z} + \sqrt{\beta_y}\epsilon_y, \quad \epsilon_y \sim \mathcal{N}(0,I) \end{aligned} \Rightarrow \mathbf{x} = \sqrt{\alpha_x\alpha_y}\mathbf{z} + \underbrace{\sqrt{\alpha_x\beta_y}\epsilon_y + \sqrt{\beta_x}\epsilon_x}_{(*)}$$

ϵ_x, ϵ_y は独立なので、$(*)$ が与える分散はそれぞれの分散の和

$$\alpha_x\beta_y + \beta_x$$

となるのがわかることからも示せる。↩ 時刻 t での拡散が従う分布 の証明に戻る

■ 時刻 t での拡散が従う分布

時刻 t での分布

$$p(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}\left(\mathbf{x}_t; \sqrt{\prod_{\tau=1}^{t} \alpha_\tau \mathbf{x}_0}, \left(1 - \prod_{\tau=1}^{t} \alpha_\tau\right) I\right)$$

特に $\mathbf{x}_\infty \sim \mathcal{N}(\mathbf{0}, I)$ となる。

$t = 1$の時 明らかに成り立つ（定義そのもの）

t で成り立つとする。この時 $t+1$ の分布は

$$p(\mathbf{x}_{t+1}|\mathbf{x}_0) = \int \overbrace{p(\mathbf{x}_{t+1}|\mathbf{x}_t)}^{\mathcal{N}(\sqrt{\alpha_{t+1}}\mathbf{x}_t, \beta_{t+1} I)} \underbrace{p(\mathbf{x}_t|\mathbf{x}_0)}_{\mathcal{N}(\sqrt{\prod_{\tau=1}^{t} \alpha_\tau \mathbf{x}_0}, \left(1-\prod_{\tau=1}^{t} \alpha_\tau\right) I)} d\mathbf{x}_t$$

となるため、1ページ前の補題1を使うと $t+1$ の場合も $\beta_{t+1} = 1 - \alpha_{t+1}$ に注意すれば示せる。↩ 説明に戻る

■ Gauss分布の分散による展開

補題2

$$\mathcal{N}(\tilde{\mathbf{x}};\mathbf{x},\sigma^2 I) = \delta(\tilde{\mathbf{x}}-\mathbf{x}) + \frac{\sigma^2}{2}\nabla^2\delta(\tilde{\mathbf{x}}-\mathbf{x}) + O(\sigma^4)$$

形式的にテイラー展開すると

$$\mathcal{N}(\tilde{\mathbf{x}}|\mathbf{x},\sigma^2 I) = \underbrace{\mathcal{N}(\tilde{\mathbf{x}}|\mathbf{x},\sigma^2 I)|_{\sigma^2\downarrow 0}}_{\delta(\tilde{\mathbf{x}}-\mathbf{x})} + \sigma^2 \underbrace{\frac{\partial}{\partial\sigma^2}\mathcal{N}(\tilde{\mathbf{x}}|\mathbf{x},\sigma^2 I)|_{\sigma^2\downarrow 0}}_{\frac{1}{2\sigma^2}\left(-\dim(\tilde{\mathbf{x}})+\frac{(\tilde{\mathbf{x}}-\mathbf{x})^2}{\sigma^2}\right)\mathcal{N}(\tilde{\mathbf{x}}|\mathbf{x},\sigma^2 I)|_{\sigma^2\downarrow 0}} + O(\sigma^4)$$

ところで、$\mathcal{N}(\tilde{\mathbf{x}}|\mathbf{x},\sigma^2 I)$ のラプラシアンを計算すると

$$\nabla^2\mathcal{N}(\tilde{\mathbf{x}}|\mathbf{x},\sigma^2 I)|_{\sigma^2\downarrow 0} = \frac{1}{2\sigma^2}\left(-\dim(\tilde{\mathbf{x}}) + \frac{(\tilde{\mathbf{x}}-\mathbf{x})^2}{\sigma^2}\right)\mathcal{N}(\tilde{\mathbf{x}}|\mathbf{x},\sigma^2 I)|_{\sigma^2\downarrow 0}$$

となるので示せたことになる。 ↩ forwardの証明に戻る ↩ backwardの証明に戻る

■ デルタ関数の公式

補題3

$$\delta(\tilde{\mathbf{x}}_t - \tilde{\mathbf{x}} - \mathbf{f}(\tilde{\mathbf{x}}, t)\Delta t) = \frac{\delta(\tilde{\mathbf{x}}_t - \tilde{\mathbf{x}} - \mathbf{f}(\tilde{\mathbf{x}}_t, t)\Delta t)}{1 + \nabla_{\tilde{\mathbf{x}}} \cdot \mathbf{f}(\tilde{\mathbf{x}}_t, t)\Delta t} + O(\Delta t^2)$$

$$\delta(F(\tilde{\mathbf{x}})) = \frac{F(\tilde{\mathbf{x}})\text{のゼロ点まわり1次展開}}{\det|\nabla_{\tilde{\mathbf{x}}} F^\top(\tilde{\mathbf{x}})|}$$

を使う。まず分子は、ゼロ点を見つけないといけないが、再起的に

$$\tilde{\mathbf{x}} = \tilde{\mathbf{x}}_t - \mathbf{f}(\tilde{\mathbf{x}}, t)\Delta t = \tilde{\mathbf{x}}_t - \mathbf{f}(\tilde{\mathbf{x}}_t - \mathbf{f}(\tilde{\mathbf{x}}, t)\Delta t, t)\Delta t = \tilde{\mathbf{x}}_t - \mathbf{f}(\tilde{\mathbf{x}}_t, t)\Delta t + O(\Delta t^2)$$

となって、右辺のデルタ関数での指定になる。分母は

$$\begin{aligned}
&\det|-I - \nabla_{\tilde{\mathbf{x}}}\mathbf{f}^\top(\tilde{\mathbf{x}}, t)\Delta t| = \det(I + \nabla_{\tilde{\mathbf{x}}}\mathbf{f}^\top(\tilde{\mathbf{x}}, t)\Delta t) \\
&= \sum_{\sigma\in S_N} |\sigma| \prod_{i=1}^{N} \left(\delta_{i,\sigma(i)} + \partial_i f^{\sigma(i)}(\tilde{\mathbf{x}}, t)\Delta t\right) = \sum_{\sigma\in S_N} |\sigma| \Bigg(\underbrace{\prod_{i=1}^{N} \delta_{i,\sigma(i)} + \sum_{j=1}^{N} \Bigg[\prod_{i=1, i\neq j}^{N} \delta_{i,\sigma(i)}\Bigg] \partial_j f^{\sigma(j)}(\tilde{\mathbf{x}}, t)\Delta t}_{\sigma=1\text{のみ非ゼロ}} + o(\Delta t)\Bigg) \\
&= 1 + \nabla_{\tilde{\mathbf{x}}} \cdot \mathbf{f}(\tilde{\mathbf{x}}, t) + o(\Delta t)
\end{aligned}$$

↩ forwardの証明に戻る

■ **Kolmogorov forward 方程式 (Fokker-Planck方程式)**

離散版（Gauss分布によるMarkov連鎖）を考え、時刻 t での分布の表現を Δt 一次までで考える。ここでは $\mathbf{x}_t = \tilde{\mathbf{x}}, \mathbf{x}_{t-1} = \mathbf{x}$ としている。

$$
\begin{aligned}
p_t(\tilde{\mathbf{x}}) &= \int p_t(\tilde{\mathbf{x}}|\mathbf{x}) p_{t-1}(\mathbf{x}) d\mathbf{x} \\
&= \int \underbrace{\mathcal{N}(\tilde{\mathbf{x}}|\mathbf{x} + \mathbf{f}(\mathbf{x}, t)\Delta t, g(t)^2 \Delta t I)}_{\delta(\tilde{\mathbf{x}}-\mathbf{x}-\mathbf{f}(\mathbf{x},t)\Delta t)+\frac{g(t)^2\Delta t}{2}\nabla^2_{\tilde{\mathbf{x}}}\delta(\tilde{\mathbf{x}}-\mathbf{x}-\mathbf{f}(\mathbf{x},t)\Delta t)} p_{t-1}(\mathbf{x}) d\mathbf{x}
\end{aligned}
$$

[補題2]

[補題3]

$$
\begin{aligned}
&= p_{t-1}(\tilde{\mathbf{x}} - \mathbf{f}(\tilde{\mathbf{x}}, t)\Delta t)/(1 + \nabla_{\tilde{\mathbf{x}}} \cdot \mathbf{f}(\tilde{\mathbf{x}}, t)\Delta t) + \frac{g(t)^2 \Delta t}{2} \nabla^2_{\tilde{\mathbf{x}}} p_{t-1}(\tilde{\mathbf{x}}) \\
&= p_{t-1}(\tilde{\mathbf{x}}) \underbrace{- \mathbf{f}(\tilde{\mathbf{x}}, t)\Delta t \cdot \nabla_{\tilde{\mathbf{x}}} p_{t-1}(\tilde{\mathbf{x}}) - \nabla_{\tilde{\mathbf{x}}} \cdot \mathbf{f}(\tilde{\mathbf{x}}, t)\Delta t p_{t-1}(\tilde{\mathbf{x}})}_{-\Delta t \nabla_{\tilde{\mathbf{x}}} \cdot (\mathbf{f}(\tilde{\mathbf{x}},t) p_{t-1}(\tilde{\mathbf{x}}))} + \frac{g(t)^2 \Delta t}{2} \nabla^2_{\tilde{\mathbf{x}}} p_{t-1}(\tilde{\mathbf{x}})
\end{aligned}
$$

あとは 一項目を移行し、Δt で割れば

$$
\frac{p_t(\tilde{\mathbf{x}}) - p_{t-1}(\tilde{\mathbf{x}})}{\Delta t} = -\nabla_{\tilde{\mathbf{x}}} \cdot (\mathbf{f}(\tilde{\mathbf{x}}, t) p_{t-1}(\tilde{\mathbf{x}})) + \frac{g(t)^2}{2} \nabla^2_{\tilde{\mathbf{x}}} p_{t-1}(\tilde{\mathbf{x}})
$$

最後に $p_{t-1} = p_t + O(\Delta t)$ を使い $\Delta t \to +0$。 ↩ 説明に戻る

■ Kolmogorov backward 方程式

離散版（Gauss分布によるMarkov連鎖）を考え、時刻 t での分布の表現を Δt 一次までで考える。ここでは $\mathbf{x}_t = \tilde{\mathbf{x}}, \mathbf{x}_{t+1} = \mathbf{x}$ としている。

$$
\begin{aligned}
p_t(\mathbf{x}_s|\tilde{\mathbf{x}}) &= \int p_{t+1}(\mathbf{x}_s|\mathbf{x}) p_t(\mathbf{x}|\tilde{\mathbf{x}}) d\mathbf{x} \\
&= \int p_{t+1}(\mathbf{x}_s|\mathbf{x}) \quad \underbrace{\mathcal{N}(\mathbf{x}|\tilde{\mathbf{x}} + \mathbf{f}(\tilde{\mathbf{x}}, t)\Delta t, g(t)^2 \Delta t I)}_{\delta(\mathbf{x}-\tilde{\mathbf{x}}-\mathbf{f}(\tilde{\mathbf{x}},t)\Delta t) + \frac{g(t)^2 \Delta t}{2} \nabla_{\mathbf{x}}^2 \delta(\mathbf{x}-\tilde{\mathbf{x}}-\mathbf{f}(\tilde{\mathbf{x}},t)\Delta t)} \quad d\mathbf{x} \\
&= \underbrace{p_{t+1}(\mathbf{x}_s|\tilde{\mathbf{x}} + \mathbf{f}(\tilde{\mathbf{x}}, t)\Delta t)}_{p_{t+1}(\mathbf{x}_s|\tilde{\mathbf{x}}) + \mathbf{f}(\tilde{\mathbf{x}},t)\Delta t \cdot \nabla_{\tilde{\mathbf{x}}} p_{t+1}(\mathbf{x}_s|\tilde{\mathbf{x}})} + \frac{g(t)^2}{2} \Delta t \nabla_{\tilde{\mathbf{x}}}^2 p_{t+1}(\mathbf{x}_s|\tilde{\mathbf{x}})
\end{aligned}
$$

補題2

あとは 一項目を移行し、Δt で割れば

$$
\frac{p_t(\mathbf{x}_s|\tilde{\mathbf{x}}) - p_{t+1}(\mathbf{x}_s|\tilde{\mathbf{x}})}{\Delta t} = \mathbf{f}(\tilde{\mathbf{x}}, t) \cdot \nabla_{\tilde{\mathbf{x}}} p_{t+1}(\mathbf{x}_s|\tilde{\mathbf{x}}) + \frac{g(t)^2}{2} \nabla_{\tilde{\mathbf{x}}}^2 p_{t+1}(\mathbf{x}_s|\tilde{\mathbf{x}})
$$

最後に $p_{t+1} = p_t + O(\Delta t)$ を使い $\Delta t \to +0$。 ↩ 説明に戻る

■ 逆拡散

時刻 t で $\mathbf{x}$、時刻 s で $\mathbf{x}_s$ を見出す確率密度は、$p_t(\mathbf{x}_s,\mathbf{x}) = p_t(\mathbf{x}_s|\mathbf{x})p_t(\mathbf{x})$ だが、これを時間微分すると

$$
\begin{aligned}
&\frac{\partial}{\partial t}\underbrace{p_t(\mathbf{x}_s,\mathbf{x})}_{p_t(\mathbf{x}_s|\mathbf{x})p_t(\mathbf{x})} \\
&= \underbrace{\frac{\partial}{\partial t}p_t(\mathbf{x}_s|\mathbf{x})}_{-\mathbf{f}(\mathbf{x},t)\cdot\nabla_{\mathbf{x}}p_t(\mathbf{x}_s|\mathbf{x})-\frac{g(t)^2}{2}\nabla^2_{\mathbf{x}}p_t(\mathbf{x}_s|\mathbf{x})} p_t(\mathbf{x}) + p_t(\mathbf{x}_s|\mathbf{x})\underbrace{\frac{\partial}{\partial t}p_t(\mathbf{x})}_{-\nabla_{\mathbf{x}}\cdot(\mathbf{f}(\mathbf{x},t)p_t(\mathbf{x}))+\frac{g(t)^2}{2}\nabla^2_{\mathbf{x}}p_t(\mathbf{x})} \\
&= -\nabla_{\mathbf{x}}(\mathbf{f}(\mathbf{x},t)p_t(\mathbf{x}_s|\mathbf{x})p_t(\mathbf{x})) + \frac{g(t)^2}{2}\left(\underbrace{-\nabla^2_{\mathbf{x}}p_t(\mathbf{x}_s|\mathbf{x})\cdot p_t(\mathbf{x})}_{-\nabla^2_{\tilde{\mathbf{x}}}(p_t(\tilde{\mathbf{x}}_s|\tilde{\mathbf{x}})p_t(\tilde{\mathbf{x}}))\,+\,2\nabla_{\tilde{\mathbf{x}}}p_t(\tilde{\mathbf{x}}_s|\tilde{\mathbf{x}})\,\cdot\,\nabla_{\tilde{\mathbf{x}}}p_t(\tilde{\mathbf{x}})\,+\,p_t(\tilde{\mathbf{x}}_s|\tilde{\mathbf{x}})\nabla^2_{\tilde{\mathbf{x}}}p_t(\tilde{\mathbf{x}})} + p_t(\mathbf{x}_s|\mathbf{x})\nabla^2_{\mathbf{x}}p_t(\mathbf{x})\right) \\
&= -\nabla_{\tilde{\mathbf{x}}}\left\{[\mathbf{f}(\tilde{\mathbf{x}},t) - g(t)^2\nabla_{\tilde{\mathbf{x}}}\log p_t(\tilde{\mathbf{x}})]\underbrace{p_t(\tilde{\mathbf{x}}_s|\tilde{\mathbf{x}})p_t(\tilde{\mathbf{x}})}_{p_t(\tilde{\mathbf{x}}_s,\tilde{\mathbf{x}})}\right\} + \frac{g(t)^2}{2}\left(-\nabla^2_{\tilde{\mathbf{x}}}\left(\underbrace{p_t(\tilde{\mathbf{x}}_s|\tilde{\mathbf{x}})p_t(\tilde{\mathbf{x}})}_{p_t(\tilde{\mathbf{x}}_s,\tilde{\mathbf{x}})}\right)\right)
\end{aligned}
$$

最後の段までくると、$p_t(\mathbf{x}_s,\mathbf{x})$ に関する方程式。これを $\mathbf{x}_s$ について積分する

■ 逆拡散

$\int p_t(\mathbf{x}_s, \mathbf{x}) d\mathbf{x}_s := p_t^{(b)}(\mathbf{x})$ とすると

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

さらに $\tau = T - t$ とすると

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

ドリフトと分散を読み取ると、これは 以下の確率微分方程式から誘導される Fokker-Plank方程式であるのがわかる：

$$d\mathbf{x}(\tau) = -[\mathbf{f}(\mathbf{x}, t) - g(t)^2 \nabla_{\mathbf{x}} \log p_t(\mathbf{x})] d\tau + g(t) d\bar{\mathbf{w}}(\tau)$$

↩ 説明に戻る

■ 逆拡散の確率フローでの尤度計算

解を $\mathbf{x}(t)$ とかくと、微分の連鎖率から

$$
\begin{aligned}
&\frac{d}{dt}p_t(\mathbf{x}(t)) \\
&= \underbrace{\dot{p}_t(\mathbf{x})}_{\text{Fokker-Planck方程式}} + \underbrace{\dot{\mathbf{x}}(t)}_{\text{逆拡散の確率フロー}} \cdot \nabla_{\mathbf{x}} p_t(\mathbf{x}(t)) \\
&= -\nabla_{\mathbf{x}} \cdot \left(\left[\mathbf{f}(\mathbf{x}, t) - \frac{g(t)^2}{2} s(\mathbf{x}, t)\right] p_t(\mathbf{x}) \right) + \left[\mathbf{f}(\mathbf{x}, t) - \frac{g(t)^2}{2} s(\mathbf{x}, t)\right] \cdot \nabla_{\mathbf{x}} p_t(\mathbf{x}(t)) \\
&= -\left(\nabla_{\mathbf{x}} \cdot \left[\mathbf{f}(\mathbf{x}, t) - \frac{g(t)^2}{2} s(\mathbf{x}, t)\right] \right) p_t(\mathbf{x}) \\
&\Downarrow \\
&\frac{d}{dt} \log p_t(\mathbf{x}(t)) = -\nabla_{\mathbf{x}} \cdot \left[\mathbf{f}(\mathbf{x}, t) - \frac{g(t)^2}{2} s(\mathbf{x}, t)\right] \\
&\Downarrow \\
&\log p_T(\mathbf{x}(T)) - \log p_0(\mathbf{x}(0)) = -\int_0^T \nabla_{\mathbf{x}} \cdot \left[\mathbf{f}(\mathbf{x}(t), t) - \frac{g(t)^2}{2} s(\mathbf{x}(t), t)\right] dt
\end{aligned}
$$

↩ 説明に戻る

■ 連続の方程式

まず、時刻 τ での分布は以下のようにかける：

$$p_\tau(\mathbf{x}) = \int \delta(\mathbf{x} - \mathbf{x}(\tau))p_0(\mathbf{x}(0))d\mathbf{x}(0)$$

これを素朴に時間微分すると示せる：

$$\begin{aligned}
\frac{\partial p_\tau(\mathbf{x})}{\partial \tau} &= \frac{\partial}{\partial \tau}\int \delta(\mathbf{x} - \mathbf{x}(\tau))p_0(\mathbf{x}(0))d\mathbf{x}(0) \\
&= \int (-\underbrace{\dot{\mathbf{x}}(\tau)}_{\mathbf{u}_\tau(\mathbf{x}(\tau))}) \cdot \nabla_{\mathbf{x}}\delta(\mathbf{x} - \mathbf{x}(\tau))p_0(\mathbf{x}(0))d\mathbf{x}(0) \\
&= -\nabla_{\mathbf{x}} \cdot \int (\mathbf{u}_\tau(\underbrace{\mathbf{x}(\tau)}_{\mathbf{x}\text{に置き換えて良い}}))\delta(\mathbf{x} - \mathbf{x}(t))p_0(\mathbf{x}(0))d\mathbf{x}(0) \\
&= -\nabla_{\mathbf{x}} \cdot \left(\mathbf{u}_\tau(\mathbf{x})\underbrace{\int \delta(\mathbf{x} - \mathbf{x}(\tau))p_0(\mathbf{x}(0))d\mathbf{x}(0)}_{p_\tau(\mathbf{x})}\right)
\end{aligned}$$

↩ 説明に戻る

■ **条件付きフロー**

$$\int \delta(\tilde{\mathbf{x}} - \tilde{\mathbf{x}}_\tau(\mathbf{x}))\underbrace{p_0(\tilde{\mathbf{x}}_0)}_{\mathcal{N}(\mathbf{0},I)}d\tilde{\mathbf{x}}_0 = \underbrace{p_\tau(\tilde{\mathbf{x}}|\mathbf{x})}_{\mathcal{N}(\mu_\tau(\mathbf{x}),\sigma_\tau(\mathbf{x})^2)}$$

が条件なので $\dot{\tilde{\mathbf{x}}}_\tau(\mathbf{x}) = \mathbf{u}_\tau(\tilde{\mathbf{x}}_\tau \mid \mathbf{x})$ の解は

$$\tilde{\mathbf{x}}_\tau(\mathbf{x}) = \mu_\tau(\mathbf{x}) + \sigma_\tau(\mathbf{x})\tilde{\mathbf{x}}_0$$

となる。これを τ 微分すれば

$$\begin{aligned}\mathbf{u}_\tau(\tilde{\mathbf{x}}_\tau \mid \mathbf{x}) &= \dot{\mu}_\tau(\mathbf{x}) + \dot{\sigma}_\tau(\mathbf{x})\underbrace{\mathbf{x}_0}_{(\mathbf{x}_\tau(\mathbf{x})-\mu_\tau(\mathbf{x}))/\sigma_\tau(\mathbf{x})} \\ &= \frac{\dot{\sigma}_\tau(\mathbf{x})}{\sigma_\tau(\mathbf{x})}(\mathbf{x}_\tau(\mathbf{x}) - \mu_\tau(\mathbf{x})) + \dot{\mu}_\tau(\mathbf{x})\end{aligned}$$

となる。↩ 説明に戻る

■ フローと条件付きフローの関係式

補題4

$$\mathbf{u}_\tau(\tilde{\mathbf{x}}) = \frac{\int \mathbf{u}_\tau(\tilde{\mathbf{x}}|\mathbf{x})p_\tau(\tilde{\mathbf{x}}|\mathbf{x})p(\mathbf{x})d\mathbf{x}}{p_\tau(\tilde{\mathbf{x}})}$$

連続の方程式を比較する。

$$\begin{aligned}
-\nabla_{\tilde{\mathbf{x}}} \cdot \left(\underline{\mathbf{u}_\tau(\tilde{\mathbf{x}})p_\tau(\tilde{\mathbf{x}})}\right) &= \frac{\partial}{\partial \tau} p_\tau(\tilde{\mathbf{x}}) \\
&= \frac{\partial}{\partial \tau} \int p_\tau(\tilde{\mathbf{x}}|\mathbf{x})p(\mathbf{x})d\mathbf{x} \\
&= -\nabla_{\tilde{\mathbf{x}}} \cdot \underline{\int \mathbf{u}_\tau(\tilde{\mathbf{x}}|\mathbf{x})p_\tau(\tilde{\mathbf{x}}|\mathbf{x})p(\mathbf{x})d\mathbf{x}}
\end{aligned}$$

あとは移行すればOK。↩ ノイズ入りフローマッチング証明に戻る

■ ノイズ入りフローマッチング

ノルム二乗を展開して比較します：

$$\left\langle \left(\mathbf{v}_{\tau,\theta}(\tilde{\mathbf{x}}) - \mathbf{u}_\tau(\tilde{\mathbf{x}})\right)^2 \right\rangle_{\tau, p_\tau(\tilde{\mathbf{x}})}$$

$$= \left\langle \mathbf{v}_{\tau,\theta}(\tilde{\mathbf{x}})^2 - 2\mathbf{v}_{\tau,\theta}(\tilde{\mathbf{x}}) \cdot \underbrace{\mathbf{u}_\tau(\tilde{\mathbf{x}})}_{\int d\mathbf{x} p(\mathbf{x})[p_\tau(\tilde{\mathbf{x}}|\mathbf{x})\mathbf{u}_\tau(\tilde{\mathbf{x}}|\mathbf{x})]/p_\tau(\tilde{\mathbf{x}})} \right\rangle + \text{const.}$$

→補題4

$$= \left\langle \int \underbrace{p_\tau(\tilde{\mathbf{x}})}_{\int p_\tau(\tilde{\mathbf{x}}|\mathbf{x})p(\mathbf{x})d\mathbf{x}} \mathbf{v}_{\tau,\theta}(\tilde{\mathbf{x}})^2 d\tilde{\mathbf{x}} - 2\int \underline{p_\tau(\tilde{\mathbf{x}})} \mathbf{v}_{\tau,\theta}(\tilde{\mathbf{x}}) \cdot \frac{\int \mathbf{u}_\tau(\tilde{\mathbf{x}}|\mathbf{x}) p_\tau(\tilde{\mathbf{x}}|\mathbf{x}) p(\mathbf{x}) d\mathbf{x}}{\underline{p_\tau(\tilde{\mathbf{x}})}} d\tilde{\mathbf{x}} \right\rangle_\tau$$

$$+ \text{const.}$$

$$= \left\langle \mathbf{v}_{\tau,\theta}(\tilde{\mathbf{x}})^2 - 2\mathbf{v}_{\tau,\theta}(\tilde{\mathbf{x}}) \cdot \mathbf{u}_\tau(\tilde{\mathbf{x}}|\mathbf{x}) \right\rangle_{\tau, p_\tau(\tilde{\mathbf{x}}|\mathbf{x})p(\mathbf{x})} + \text{const.}$$

$$= \left\langle \left(\mathbf{v}_{\tau,\theta}(\tilde{\mathbf{x}}) - \mathbf{u}_\tau(\tilde{\mathbf{x}}|\mathbf{x})\right)^2 \right\rangle_{\tau, p_\tau(\tilde{\mathbf{x}}|\mathbf{x})p(\mathbf{x})} + \text{const'.}$$

↩ 説明に戻る