Schrödinger Bridge の数理

安田 勇輝

2026 年 1 月 30 日
Ver 1.0.1

目次

1 問題設定

『拡散モデルの数理』に引き続き，確率微分方程式 (Stochastic Differential Equation; SDE) を用いたデータ変換について考える．問題の本質を掴むため，できるだけ簡単な問題設定で議論を進める．具体的に，状態変数 $x \in \mathbb{R}$ を考え，この x に対して確率密度 $\rho_0(x)$ から $\rho_1(x)$ への SDE による変換を考える．対応する例として画像の有色化タスクを考えると，x はピクセル値の集まりからなる画像 1 枚を表し，$\rho_0(x)$ は白黒画像の集合，$\rho_1(x)$ はカラー画像の集合を表す．この場合に求めたいのは，白黒からカラーへの変換方法である．問題を簡単にするため x は実数値，つまり 1 ピクセルの画素値に対応する状況を扱うが，x をベクトルにした場合と本質は変わらない．変換を行う SDE として以下を導入する．

$$dx = u(x,t)dt + \sigma dW \tag{1}$$

ここで $t = 0$ から $t = 1$ まで SDE を積分し，確率密度の変換を行う．何も工夫をしなければ，この SDE の持つノイズによって無秩序な状態へと遷移してしまう．そこで，変換の速度 $u(x,t)$ を適切に決定することで，確率密度 $\rho_0(x)$ から $\rho_1(x)$ への変換を実現させる．つまり，我々の目的は与えられた $\rho_0(x)$, $\rho_1(x)$, そしてパラメータ σ のもとで，$u(x,t)$ の関数形を決定することになる．この問題は Schrödinger Bridge と呼ばれる．

速度 u の関数形を決定するための方針として，変換ができるだけランダムとなるようにする．このような方針は，最大エントロピー法など理学や工学の様々な問題で現れる．直感的には，データの多様性を効率的に

表現するために，変換をできるだけランダムにするのは妥当なように見える．ここで，ランダムな変換を表すSDE として

$$dx = \sigma dW \tag{2}$$

を考える．これはただノイズ σdW を加え続ける変換であり，十分 t が大きいときに x の分布は正規分布となる．

2　経路積分形式の導入

つぎに「変換ができるだけランダムとなる」という指針を数式で表す．この数式を目的関数として u の最適化を行えば，求める変換が得られる．ランダムさを測るためには，確率密度があればよさそうである．そこで，式 (1) と (2) のパスの確率密度をそれぞれ $p\left(\{x\}_t\right)$ および $q\left(\{x\}_t\right)$ とする．このノートでは波括弧 $\{\cdot\}$ はパスを表す意味で用いる．パスとは $t=0$ から $t=1$ までの x の時系列であり，$\{x\}_t$ は $\{x(t)\}_{t=0}^{t=1}$ の省略形である．確率解析の慣習にならい，時系列ではなくパスという用語を用いる．

まず簡単そうな式 (2) から考える．時刻を刻み幅 Δt で離散化すれば，式 (2) は

$$x_{t+\Delta t} - x_t = \sigma \Delta W_t \tag{3}$$

とかける．この ΔW_t は標準正規乱数 $\epsilon \sim \mathcal{N}(0,1)$ を用いて $\Delta W_t = \epsilon\sqrt{\Delta t}$ とかけるのであった (下付き添字 t は基準時刻を表す)．つまり，ΔW_t 自体は分散が Δt の正規分布に従う．

$$\frac{d\Delta W_t}{\sqrt{2\pi\Delta t}} \exp\left[-\frac{1}{2}\frac{(\Delta W_t)^2}{\Delta t}\right] \tag{4}$$

式 (3) を ΔW_t (と x_t) が与えられた時，$x_{t+\Delta t}$ を計算する式とみなせば，$x_{t+\Delta t}$ の遷移確率分布が得られる．

$$q\left(x_{t+\Delta t} \mid x_t\right) = \frac{dx_{t+\Delta t}}{\sqrt{2\pi\sigma^2\Delta t}} \exp\left[-\frac{1}{2}\frac{\left(x_{t+\Delta t} - x_t\right)^2}{\sigma^2\Delta t}\right] \tag{5}$$

ここで確率分布の変換の Jacobian が $1/\sigma$ であることを用いた．別の言い方をすれば，式 (3) を解いて得られる $\Delta W_t = (x_{t+\Delta t} - x_t)/\sigma$ を式 (4) に代入している．Jacobian により，規格化定数の中に σ が現れる．これにより確率分布が 1 に規格化される．この遷移確率分布に従い，x_t が与えられた時，次の時刻の $x_{t+\Delta t}$ がランダムに決定される．

ある 1 本の時系列 x_t の従う確率分布は，この遷移確率のかけ合わせとして得られる．

$$\begin{aligned}
q(\{x\}_t \mid x_0) &= \prod_n q\left(x_{n+1} \mid x_n\right) && (6)\\
&= \left(\prod_n \frac{dx_{n+1}}{\sqrt{2\pi\sigma^2\Delta t}}\right) \exp\left[-\frac{1}{2}\sum_n \frac{\left(x_{n+1} - x_n\right)^2}{\sigma^2\Delta t}\right] && (7)\\
&= \left(\prod_n \frac{dx_{n+1}}{\sqrt{2\pi\sigma^2\Delta t}}\right) \exp\left[-\frac{1}{2\sigma^2}\sum_n \Delta t\left(\frac{x_{n+1} - x_n}{\Delta t}\right)^2\right] && (8)\\
&\propto \exp\left[-\frac{1}{2\sigma^2}\int dt\,(\dot{x})^2\right] && (9)
\end{aligned}$$

時刻を $t = t_0, \ldots, t_N$ (i.e., $t_n = n\Delta t$) のように離散化し，添字 n で時刻を表した (e.g., $x_n = x(t_n)$)，さらに，指数関数の性質を使って遷移確率の積を指数の和に直した．この時，差分近似を用いて $\dot{x} = (x_{n+1} - x_n)/\Delta t$

と表現している．最後の表式では，Δt が十分小さいとして，和を積分記号で置き換えている．これはあくまで形式的な表式であり，実際に $\Delta t \to 0$ とはしていない．時間方向の増分の和により，ある時系列 1 本の確率を表現できた．これを「経路積分によるパスの確率密度の表現」などと呼ぶ．初期条件 x_0 は外から与える形となっており，それを表すために最左辺を条件付き確率密度 $q(\{x\}_t \mid x_0)$ としている．パスの確率分布を求める手順は以下のようにまとめられる．例えば，(株価などの) ある 1 本の時系列が得られたとする $(x_0, \ldots, x_N)$．この時系列の出現頻度 (つまり確率密度) を求めるには，各時間ステップの時間変化率 $\dot{x} = (x_{n+1} - x_n)/\Delta t$ を計算し，各変化率に対する正規分布をかけ合わせればよい．

同様の議論を行うことで，式 (1) に x が従う場合のパスの確率密度も求められる

$$p(\{x\}_t \mid x_0) = \left(\prod_n \frac{dx_{n+1}}{\sqrt{2\pi\sigma^2\Delta t}}\right) \exp\left[-\frac{1}{2\sigma^2}\sum_n \Delta t \left(\frac{x_{n+1} - x_n - u(x_n, t_n)\Delta t}{\Delta t}\right)^2\right] \tag{10}$$

$$\propto \exp\left[-\frac{1}{2\sigma^2}\int dt\,(\dot{x} - u(x,t))^2\right] \tag{11}$$

確率密度 $q(\{x\}_t \mid x_0)$ と同様に，初期条件 x_0 を与える必要があり，それは確率密度の条件として $p(\{x\}_t \mid x_0)$ のように明記している．今，初期分布 ($t = 0$ の確率密度分布) は問題設定により $\rho_0(x_0)$ に従う．つまり，結合確率密度は $p\left(\{x\}_t\right) = p(\{x\}_t \mid x_0)\rho_0(x_0)$ と書ける ($q\left(\{x\}_t\right)$ も同様).

3 パスを用いた最適化問題の定式化

準備が完了したので「変換ができるだけランダムとなる」という条件を数式で書き下す．得られる数式が最適化問題の目的関数となる．SDE (2) による変換をランダムにするためには，SDE の与えるパスの確率密度 $p\left(\{x\}_t\right)$ が，参照となる確率密度 $q\left(\{x\}_t\right)$ にできるだけ近ければよい．この近さの指標として，Kullback–Leibler ダイバージェンス (KL ダイバージェンス) D_{KL} を選ぶ.

$$D_{\mathrm{KL}}(p\left(\{x\}_t\right) \| q\left(\{x\}_t\right)) := \mathbb{E}_p\left[\ln\left(\frac{p\left(\{x\}_t\right)}{q\left(\{x\}_t\right)}\right)\right] \tag{12}$$

$$= \int \mathcal{D}x\, p\left(\{x\}_t\right) \ln\left(\frac{p\left(\{x\}_t\right)}{q\left(\{x\}_t\right)}\right) \tag{13}$$

ここで $\mathcal{D}x := \prod_n dx_n$ とした.

以下ではこの KL ダイバージェンスの具体形を求める．条件より $t = 0$ と $t = 1$ では x は $\rho_0(x_0)$ と $\rho_1(x_1)$ に従うから，経路積分表現により

$$p\left(\{x\}_t\right) = p\left(\{x\}_t \mid x_0\right)\rho_0\left(x_0\right) \tag{14}$$

$$q\left(\{x\}_t\right) = q\left(\{x\}_t \mid x_0\right)\rho_0\left(x_0\right) \tag{15}$$

となる．これを代入すると対数関数を以下のように変形できる.

$$\ln\left(\frac{p\left(\{x\}_t\right)}{q\left(\{x\}_t\right)}\right) = \ln\left(\frac{p\left(\{x\}_t \mid x_0\right)}{q\left(\{x\}_t \mid x_0\right)}\right) \tag{16}$$

$$= -\frac{1}{2\sigma^2}\sum_n \Delta t\left[\left(\frac{x_{n+1} - x_n - u\left(x_n, t_n\right)\Delta t}{\Delta t}\right)^2 - \left(\frac{x_{n+1} - x_n}{\Delta t}\right)^2\right] \tag{17}$$

ここで，確率密度の規格化定数が同一であることと，式 (9) と (11) を用いて，指数部分のみを取り出した.

ここで $\Delta x_n := x_{n+1} - x_n$ および $u_n := u\left(x_n, t_n\right)$ とすれば，式展開の続きは以下のようになる．

$$\ln\left(\frac{p\left(\{x\}_t \mid x_0\right)}{q\left(\{x\}_t \mid x_0\right)}\right) = -\frac{1}{2\sigma^2}\sum_n \Delta t\left[\left(\frac{\Delta x_n}{\Delta t} - u_n\right)^2 - \left(\frac{\Delta x_n}{\Delta t}\right)^2\right] \tag{18}$$

$$= -\frac{1}{2\sigma^2}\sum_n \Delta t\left(u_n^2 - 2u_n\frac{\Delta x_n}{\Delta t}\right) \tag{19}$$

$$= \frac{1}{\sigma^2}\sum_n u_n \Delta x_n - \frac{1}{2\sigma^2}\sum_n u_n^2 \Delta t \tag{20}$$

以上より KL ダイバージェンスは以下のように変形される．

$$D_{\mathrm{KL}}\left(p\left(\{x\}_t\right) \| q\left(\{x\}_t\right)\right) = \mathbb{E}_p\left[\frac{1}{\sigma^2}\sum_n u_n \Delta x_n - \frac{1}{2\sigma^2}\sum_n u_n^2 \Delta t\right], \tag{21}$$

さらに変形を続けるには，Δx の具体形が必要である．D_{KL} を計算する際の期待値はパスの確率密度 $p\left(\{x\}_t\right)$ に対して取っている．つまり，この期待値の中では x は SDE (1) に従うとみなされる．そこで，この離散化表現 (3) を利用すれば，

$$\Delta x_n = u_n \Delta t + \sigma \Delta W_n \tag{22}$$

となる．これを代入することで，式 (21) の第一項の期待値は以下のように変形できる．

$$\mathbb{E}_p\left[u_n \Delta x_n\right] = \mathbb{E}_p\left[u_n^2 \Delta t\right] + \sigma \mathbb{E}_p\left[u_n \Delta W_n\right] \tag{23}$$

$$= \mathbb{E}_p\left[u_n^2 \Delta t\right] \tag{24}$$

ここで，ΔW_n は平均 0 かつ分散 Δt の正規乱数であるから，その期待値は 0 となることを用いた．この結果を式 (21) に代入すれば

$$D_{\mathrm{KL}}\left(p\left(\{x\}_t\right) \| q\left(\{x\}_t\right)\right) = \mathbb{E}_p\left[\frac{1}{\sigma^2}\sum_n u_n^2 \Delta t - \frac{1}{2\sigma^2}\sum_n u_n^2 \Delta t\right] = \frac{1}{2\sigma^2}\mathbb{E}_p\left[\sum_n u_n^2 \Delta t\right] \tag{25}$$

となる．最後に Δt が十分小さいとして，形式的に和を積分で表すと

$$D_{\mathrm{KL}}\left(p\left(\{x\}_t\right) \| q\left(\{x\}_t\right)\right) = \frac{1}{2\sigma^2}\mathbb{E}_p\left[\int_0^1 u(x(t), t)^2 dt\right] \tag{26}$$

となる．期待値は確率密度 $p\left(\{x\}_t\right) = p\left(\{x\}_t \mid x_0\right)\rho_0\left(x_0\right)$ に対してとる．ここでパスの条件付き確率密度 $p\left(\{x\}_t \mid x_0\right)$ は式 (11) で与えられる．

以上より，我々が解くべき最適化問題が得られた．

$$\min D_{\mathrm{KL}}\left(p\left(\{x\}_t\right) \| q\left(\{x\}_t\right)\right) = \min \frac{1}{2\sigma^2}\mathbb{E}_p\left[\int_0^1 u(x(t), t)^2 dt\right] \tag{27}$$

$$\text{s.t.} \quad p(x, 0) = \rho_0(x) \quad \text{and} \quad p(x, 1) = \rho_1(x) \tag{28}$$

である．つまり，$t = 0$ と $t = 1$ の端点で確率密度が $\rho_0(x)$ と $\rho_1(x)$ になるという条件のもとで，目的関数 D_{KL} をできるだけ小さくするような関数 $u(x, t)$ を求める．この $u(x, t)$ は x の SDE の決定論的速度を与え，SDE (1) 自体は

$$dx = u(x, t)dt + \sigma dW \tag{29}$$

と与えられるのだった．最もランダムな参照確率密度 $q\left(\{x\}_t\right)$ にできるだけ近いという条件は，速度の二乗の期待値 (つまりエネルギー) をできるだけ小さくするという条件に書き換えられた．ただし，端点条件 $p(x,0)=\rho_0(x)$ と $p(x,1)=\rho_1(x)$ があるため， $u(x,t)$ を恒等的に 0 とするような (つまりエネルギーをゼロにするような) 結果は解として認められない．この最適化問題を解くことで，SDE の速度 $u(x,t)$ が適切に調節され，分布 $\rho_0(x)$ を $\rho_1(x)$ に変換する SDE が得られる．

4　パスを用いない最適化問題の定式化

式 (28) を変形し，より解きやすい形にする．この形にした後で，随伴方程式を導出し，「大域」最適解 $u^*(x,t)$ が得られることを示す．その変形の鍵になるのは，式 (28) の目的関数は，陽にパスに依存していないことである．D_{KL} の期待値はパスの確率密度 $p\left(\{x\}_t\right)$ 関して取るが，目的関数の被積分関数 $u(x,t)$ はある時刻 t の x さえ知っていれば計算できる．つまり，時系列全体を知る必要はない．そこで，パスの確率密度 $p\left(\{x\}_t\right)$ を時刻 t に関して周辺化すると，時刻 t における状態変数 x の確率密度 $p(x,t)$ が得られる．この密度は，SDE (1) に対応する Fokker–Planck 方程式 (FPE) の解となる．FPE は下で与えられる．

$$\frac{\partial}{\partial t}p(x,t)=-\frac{\partial}{\partial x}\left[u(x,t)p(x,t)\right]+\frac{\sigma^2}{2}\frac{\partial^2}{\partial x^2}p(x,t) \tag{30}$$

実際に $p\left(\{x\}_t\right)$ を周辺化した確率密度が FPE の解に一致することを示すこともできるが，ここでは直感的な説明を行う．パス $\{x\}_t$ は時間変化に関するすべての情報を持ち，$p\left(\{x\}_t\right)$ は各パスの相対頻度を表す．一方で，ある時刻 t とある時刻 x の周辺化された確率密度 $p(x,t)$ はあらゆるパスの中で (x,t) を通るパスの頻度 $p\left(\{x\}_t\right)$ を足し合わせたものとなる．この足し合わせにより，パスの情報は失われる．つまり，どういう経路を通り (x,t) に到達したかは不明となり，また (x,t) を通った後にどのような未来の状態に到達するかも不明となる．一方，FPE が記述するのは，全ての有り得る時間発展のパターンを網羅し (つまり SDE により生成されたアンサンブルを考え)，時刻 t において x をとる頻度を考えている．すべてを網羅し，ある点 (x,t) に着目している点でパスを介した記述と，直接 FPE で $p(x,t)$ を記述する方法は一致する．

この FPE を用いると最適化問題 (28) は以下となる．

$$\min D_{\mathrm{KL}}\left(p\left(\{x\}_t\right)\|q\left(\{x\}_t\right)\right)=\min_{u,p}\frac{1}{2\sigma^2}\int_0^1 u(x,t)^2p(x,t)\,dxdt \tag{31}$$

$$\text{s.t.}\quad p(x,0)=\rho_0(x),\quad p(x,1)=\rho_1(x),\quad \text{and}\quad \text{Eq. (30)} \tag{32}$$

式 (28) と異なり，パスではなく FPE の解 $p(x,t)$ を用いて，陽に期待値を計算している．最適化を達成するためには u と p の関数形を動かす．その際に拘束条件を満たす必要があり，端点条件により $t=0$ と $t=1$ の確率密度は ρ_0 と ρ_1 に固定され，またある与えられた $u(x,t)$ に対して $p(x,t)$ は FPE (30) の解となる．これは一見すると簡単ではない．なぜなら FPE は初期条件 ρ_0 を与えると，その後の確率密度 $p(x,t)$ の時間発展は決定論的に定まる．この決定論的な解が $t=1$ で $p(x,1)=\rho_1(x)$ となる保証はどこにもない．この端点条件を満たすために速度 $u(x,t)$ (またはドリフトとも呼ばれる) の関数形をうまく調節し，FPE による自然な時間発展が $p(x,1)=\rho_1(x)$ を満たすようにする．

この一見すると難しい最適化問題 (32) に対し，大域最適解のみが存在する (つまり解は目的関数の大域最小値)．言い換えると，適当な勾配法で最適化を行えば，得られるのは局所解ではなく，大域解であることが保証されている．この事実を示すために，補助変数 $r(x,t):=p(x,t)u(x,t)$ を導入する．$p(x,t)$ を質量密度

とみなせば，$r(x,t)$ は運動量フラックスに対応する．この時，式 (32) は以下となる．

$$\min_{r,p} \frac{1}{2\sigma^2} \int_0^1 \frac{r(x,t)^2}{p(x,t)}\, dxdt \quad \text{s.t.} \quad p(x,0) = \rho_0(x), \quad p(x,1) = \rho_1(x), \tag{33}$$

$$\text{and} \quad \frac{\partial}{\partial t} p(x,t) = -\frac{\partial}{\partial x} r(x,t) + \frac{\sigma^2}{2} \frac{\partial^2}{\partial x^2} p(x,t) \tag{34}$$

となる．関数 $r(x,t)^2/p(x,t)$ は $p(x,t) > 0$ とすれば 凸関数となるため，目的関数は凸となる．この時，拘束条件である偏微分方程式は線形であるため，この問題は線形拘束条件付きの凸最適化問題となる．このような凸問題を解くと，必ず大域最適解が得られる．この事実を理解するために，例えば (x,t) に関して適当な離散化を行い，得られた格子点上の値 p と r を一列に並べてベクトルにする．時間方向の格子点もまとめてベクトルにしていることに注意する．この時，拘束条件は適当な差分近似を用いれば一次式でかける．つまり，拘束条件を満たすベクトル p と r は，ある超平面上にある．この最適化問題は，凸関数である (つまり，お椀型である) 目的関数 $\sum r_i^2/p_i$ (i はベクトルの添字) をこの超平面で輪切りにして，輪切りにした超平面上で最小値を探索する問題となる．

5　随伴方程式による最適化問題の解法

連続座標値 x, t を仮定し，最適化問題 (32) を解く．連続座標を用いることで，式展開が容易になる．もし解を求めることができれば，それは大域最適解であり，目的関数を最小化する．

解くためには Lagrange の未定乗数法を適用する．未定乗数 (あるいは随伴変数とも呼べる) $\lambda(x,t)$ とすれば，最適化問題 (32) は拘束条件を外すことが可能であり，以下の目的関数の最小化問題となる．ここで目的関数のスケール因子 $1/2\sigma^2$ は単に $1/2$ と表した (スケール倍しても一般性を失わない)．

$$\mathcal{L} = \int dxdt \left[\frac{1}{2} u^2 p + \lambda \left(\frac{\partial p}{\partial t} + \frac{\partial up}{\partial x} - \frac{\sigma^2}{2} \frac{\partial^2 p}{\partial x^2} \right) + \mu_0 (p_0 - \rho_0) + \mu_1 (p_1 - \rho_1) \right] \tag{35}$$

ここで端点条件を満たすための追加の未定乗数 $\mu_0(x)$ および $\mu_1(x)$ を導入し，また $p_0 := p(x,0)$ と $p_1 := p(x,1)$ とした．式を短くするために，u, p, λ などの関数に対して引数 (x,t) は省略している．$\mathcal{L}$ を減少させるために，その勾配を求める．そのためには，微小変分を考えれば良い．

$$\begin{aligned}
\delta\mathcal{L} = \int dxdt \Bigg[& up\delta u + \frac{1}{2} u^2 \delta p + \delta\lambda \left(\frac{\partial p}{\partial t} + \frac{\partial up}{\partial x} - \frac{\sigma^2}{2} \frac{\partial^2 p}{\partial x^2} \right) && (36)\\
& + \delta\mu_0 (p_0 - \rho_0) + \mu_0 \delta p_0 + \delta\mu_1 (p_1 - \rho_1) + \mu_1 \delta p_1 && (37)\\
& + \lambda \left(\frac{\partial}{\partial t} \delta p + \frac{\partial (u\delta p + p\delta u)}{\partial x} - \frac{\sigma^2}{2} \frac{\partial^2 \delta p}{\partial x^2} \right) \Bigg] && (38)\\
= \int dxdt \Bigg[& \delta u \left(up - p\frac{\partial \lambda}{\partial x} \right) + \delta p \left(\frac{1}{2} u^2 - \frac{\partial \lambda}{\partial t} - u \frac{\partial \lambda}{\partial x} - \frac{\sigma^2}{2} \frac{\partial^2 \lambda}{\partial x^2} \right) && (39)\\
& + \delta\lambda \left(\frac{\partial p}{\partial t} + \frac{\partial up}{\partial x} - \frac{\sigma^2}{2} \frac{\partial^2 p}{\partial x^2} \right) + \delta\mu_0 (p_0 - \rho_0) + \delta\mu_1 (p_1 - \rho_1) && (40)\\
& + \delta p_0 (\mu_0 - \lambda_0) + \delta p_1 (\mu_1 + \lambda_1) \Bigg] && (41)
\end{aligned}$$

ここで部分積分を実行し，$t = 0$ と $t = 1$ の端点の寄与が $[\lambda p]_0^1$ として出てくることを用いた．一方で，x 方向の部分積分は，適当な端点条件により端点の寄与を無視した (例えば，十分大きい x で確率が 0 になるな

どの条件を課す)．これにより目的関数 $\mathcal{L}$ の一次変分は以下で与えられる

$$\frac{\delta \mathcal{L}}{\delta u} = up - p\frac{\partial \lambda}{\partial x} \tag{42}$$
$$\frac{\delta \mathcal{L}}{\delta p} = \frac{1}{2}u^2 - \frac{\partial \lambda}{\partial t} - u\frac{\partial \lambda}{\partial x} - \frac{\sigma^2}{2}\frac{\partial^2 \lambda}{\partial x^2} \tag{43}$$
$$\frac{\delta \mathcal{L}}{\delta \lambda} = \frac{\partial p}{\partial t} + \frac{\partial up}{\partial x} - \frac{\sigma^2}{2}\frac{\partial^2 p}{\partial x^2} \tag{44}$$
$$\frac{\delta \mathcal{L}}{\delta \mu_0} = p_0 - \rho_0 \tag{45}$$
$$\frac{\delta \mathcal{L}}{\delta \mu_1} = p_1 - \rho_1 \tag{46}$$
$$\frac{\delta \mathcal{L}}{\delta p_0} = \mu_0 - \lambda_0 \tag{47}$$
$$\frac{\delta \mathcal{L}}{\delta p_1} = \mu_1 + \lambda_1 \tag{48}$$

この一次変分をうまく組み合わせることで $\mathcal{L}$ を減少させる．目標は，全ての変分を 0 にすることである．この時，目的関数 $\mathcal{L}$ の停留点が得られるが，これはもとの最適問題の凸性により，大域最適解である．まず，式 (42) を 0 にすると

$$u = \frac{\partial \lambda}{\partial x} \tag{49}$$

となる．つまり λ は求めたい速度 u のポテンシャル関数とみなせる．すべての u を λ で表し，かつ式 (43) と (44) を 0 に設定すれば

$$\frac{\partial \lambda}{\partial t} = \frac{1}{2}u^2 - u\frac{\partial \lambda}{\partial x} - \frac{\sigma^2}{2}\frac{\partial^2 \lambda}{\partial x^2} \tag{50}$$
$$= \frac{1}{2}\left(\frac{\partial \lambda}{\partial x}\right)^2 - \frac{\partial \lambda}{\partial x}\frac{\partial \lambda}{\partial x} - \frac{\sigma^2}{2}\frac{\partial^2 \lambda}{\partial x^2} \tag{51}$$
$$= -\frac{1}{2}\left(\frac{\partial \lambda}{\partial x}\right)^2 - \frac{\sigma^2}{2}\frac{\partial^2 \lambda}{\partial x^2} \tag{52}$$

および

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}\left(p\frac{\partial \lambda}{\partial x}\right) + \frac{\sigma^2}{2}\frac{\partial^2 p}{\partial x^2} \tag{53}$$

が得られる．後者の式 (53) は FPE そのものである．一方，前者の方程式 (52) は FPE の随伴演算子を含むため，(FPE に伴う) 随伴方程式と呼ばれる．残りの課題は，式 (45) から式 (48) を 0 にすることである．

以上の導出した式を用いると，目的関数 $\mathcal{L}$ を減少させ，最終的にすべての一次変分を 0 にするアルゴリズムは以下となる．ここで全ての未知関数には，計算開始前に適当な値が入っていると仮定する．

1. $\lambda_1(x) \leftarrow -\mu_1(x)$ とする．この操作は式 (48) を 0 にする．
2. これを終端条件として ($\lambda(x,1) \leftarrow \lambda_1(x)$)，$t = 1$ から $t = 0$ へと逆方向に式 (52) を解き，$\lambda(x,0)$ を得る．これにより $\lambda_0(x) \leftarrow \lambda(x,0)$ と更新される．
3. $\mu_0(x) \leftarrow \lambda_0(x)$ とする．この操作は式 (47) を 0 にする．μ_0 は反復計算で特に利用されないが，全ての変分を 0 にできることを示すため，更新する．
4. $p(x,0) \leftarrow \rho_0(x)$ とする．この操作は式 (45) を 0 にする．

5. これを初期条件として，$t=0$ から $t=1$ へと順方向に式 (53) を解く．このとき，ステップ 2 から得た各 (x,t) における $\lambda(x,t)$ を用いる．積分の結果，$p(x,1)$ を得る．これにより $p_1(x) \leftarrow p(x,1)$ と更新される．
6. $\delta\mathcal{L}/\delta\mu_1 = p_1 - \rho_1$ が十分小さく，0 と見なせる場合，計算を終了する．この時，$\mathcal{L}$ の全ての一次変分は 0 と見なせる
7. そうでない場合，$\mu_1(x) \leftarrow \mu_1(x) - \eta\delta\mathcal{L}/\delta\mu_1$ を実行する (η: 正の微小数)．この操作により μ_1 は $\mathcal{L}$ を減少させる向きに変化する．1 に戻り，反復計算をくりかえす．

このように式 (52) を逆方向に解き，式 (53) を順方向に解くことを繰り返し，速度 $u(x,t)$ を更新する．この反復計算は，もとの最適化問題が凸問題であるため収束し，最終的に式 (49) から得られる速度 $u(x,t)$ は大域最適解となる．

6 指数変換による最適化問題の線形化

この章では，前章で導出した随伴方程式

$$\frac{\partial\lambda}{\partial t} = -\frac{1}{2}\left(\frac{\partial\lambda}{\partial x}\right)^2 - \frac{\sigma^2}{2}\frac{\partial^2\lambda}{\partial x^2} \tag{54}$$

を出発点として，変数変換

$$\lambda(x,t) = \sigma^2 \ln\psi(x,t) \tag{55}$$

を導入する．この変換により，非線形であった随伴方程式が線形な偏微分方程式に書き換えられる．さらに確率密度 $p(x,t)$ が

$$p(x,t) = \phi(x,t)\psi(x,t) \tag{56}$$

と補助変数 ϕ と ψ に因子分解されることを導く．方程式が線形化されることで，最適化問題が解きやすくなる．

随伴方程式 (54) が非線形である原因は，右辺第一項に $(\partial\lambda/\partial x)^2$ という二乗項が含まれている点にある．この二乗項を直接扱うことは一般に難しい．そこで，λ を対数の形で表すことを考える．対数関数は，微分すると商の形になり，さらにもう一度微分を取ると必ず二乗項を生むという特徴を持つ．この性質を利用すれば，もともと存在していた二乗項と，二階微分から新たに生じる二乗項とを打ち消し合わせることができる．この発想に基づき，式 (55) を導入する．ln を取るため，$\psi(x,t)$ は正の関数と仮定する．

この変換 (55) により，λ の各微分は以下のように書き換えられる．

$$\frac{\partial\lambda}{\partial t} = \sigma^2\frac{1}{\psi}\frac{\partial\psi}{\partial t}, \tag{57}$$

$$\frac{\partial\lambda}{\partial x} = \sigma^2\frac{1}{\psi}\frac{\partial\psi}{\partial x}, \tag{58}$$

$$\frac{\partial^2\lambda}{\partial x^2} = \sigma^2\left(\frac{1}{\psi}\frac{\partial^2\psi}{\partial x^2} - \frac{1}{\psi^2}\left(\frac{\partial\psi}{\partial x}\right)^2\right). \tag{59}$$

特に重要なのは，二階微分 (59) の中に $(\partial\psi/\partial x)^2$ に比例する二乗項が現れている点である．これは対数関数の微分に由来するものであり，(任意の) 滑らかな関数に対して成立する．

これらを随伴方程式 (54) に代入すると，

$$\sigma^2 \frac{1}{\psi}\frac{\partial \psi}{\partial t} = -\frac{1}{2}\left(\sigma^2 \frac{1}{\psi}\frac{\partial \psi}{\partial x}\right)^2 - \frac{\sigma^2}{2}\left[\sigma^2\left(\frac{1}{\psi}\frac{\partial^2 \psi}{\partial x^2} - \frac{1}{\psi^2}\left(\frac{\partial \psi}{\partial x}\right)^2\right)\right] \tag{60}$$

$$= -\frac{\sigma^4}{2}\frac{1}{\psi^2}\left(\frac{\partial \psi}{\partial x}\right)^2 - \frac{\sigma^4}{2}\frac{1}{\psi}\frac{\partial^2 \psi}{\partial x^2} + \frac{\sigma^4}{2}\frac{1}{\psi^2}\left(\frac{\partial \psi}{\partial x}\right)^2. \tag{61}$$

ここで，$(\partial\psi/\partial x)^2/\psi^2$ に比例する項が，第一項と第三項に符号を変えて現れている．これらは完全に打ち消し合い，結果として二乗の非線形項は消滅する．残った項のみを整理すると，

$$\sigma^2 \frac{1}{\psi}\frac{\partial \psi}{\partial t} = -\frac{\sigma^4}{2}\frac{1}{\psi}\frac{\partial^2 \psi}{\partial x^2} \tag{62}$$

となり，両辺に ψ/σ^2 を掛けることで，

$$\frac{\partial \psi}{\partial t} = -\frac{\sigma^2}{2}\frac{\partial^2 \psi}{\partial x^2} \tag{63}$$

を得る．非線形であった随伴方程式 (54) は，ψ に関する線形な拡散型の方程式に変換された．この方程式は，最適化アルゴリズムで示したように，t が減少する向きに (つまり逆向きに) 解かれる．

次に，随伴方程式と対になる FPE (53) を書き換える．この FPE は λ の x 微分を含むため，式 (58) に見られるように，分母に ψ が現れる．そこでこの ψ をうまく消すために，別の補助変数 ϕ を導入する．

$$\phi(x,t) := \frac{p(x,t)}{\psi(x,t)} \tag{64}$$

この時，$p(x,t) = \phi(x,t)\psi(x,t)$ が成立する．

この表式と式 (58) を FPE (53) に代入する．まず，左辺の時間微分は

$$\frac{\partial p}{\partial t} = \frac{\partial}{\partial t}(\phi\psi) = \psi\frac{\partial \phi}{\partial t} + \phi\frac{\partial \psi}{\partial t} \tag{65}$$

となる．FPE (53) の右辺の第一項と第二項は

$$-\frac{\partial}{\partial x}\left(p\frac{\partial \lambda}{\partial x}\right) = -\sigma^2\left(\frac{\partial \phi}{\partial x}\frac{\partial \psi}{\partial x} + \phi\frac{\partial^2 \psi}{\partial x^2}\right) \tag{66}$$

$$\frac{\sigma^2}{2}\frac{\partial^2 p}{\partial x^2} = \frac{\sigma^2}{2}\left(\psi\frac{\partial^2 \phi}{\partial x^2} + 2\frac{\partial \phi}{\partial x}\frac{\partial \psi}{\partial x} + \phi\frac{\partial^2 \psi}{\partial x^2}\right) \tag{67}$$

となり，FPE (53) の右辺全体は

$$\frac{\sigma^2}{2}\psi\frac{\partial^2 \phi}{\partial x^2} - \frac{\sigma^2}{2}\phi\frac{\partial^2 \psi}{\partial x^2} \tag{68}$$

となる．得られた左辺と右辺を比較すると FPE は以下となる．

$$\psi\frac{\partial \phi}{\partial t} + \phi\frac{\partial \psi}{\partial t} = \frac{\sigma^2}{2}\psi\frac{\partial^2 \phi}{\partial x^2} - \frac{\sigma^2}{2}\phi\frac{\partial^2 \psi}{\partial x^2} \tag{69}$$

ここで ψ が式 (63) を満たすことを用いると，最終的に以下を得る．

$$\frac{\partial \phi}{\partial t} = \frac{\sigma^2}{2}\frac{\partial^2 \phi}{\partial x^2} \tag{70}$$

この方程式は，最適化アルゴリズムで FPE が t の増加方向 (つまり順方向) に解かれることから，同じく順方向の方程式である．

以上より，随伴方程式 (54) に現れていた非線形項は， $\lambda = \sigma^2 \ln \psi$ という変換によって消去され，Schrödinger Bridge 問題は二本の線形な拡散型の方程式 (63) と (70) を解くことに帰着される．その解の積として与えられる $p(x,t) = \phi(x,t)\psi(x,t)$ が確率密度を与え，式 (49) と (55) が速度 $u(x,t) = \sigma^2 \partial_x \ln \psi(x,t)$ を与える．順方向と逆方向の拡散型の方程式と，そこから得られる確率密度と速度が Schrödinger Bridge 問題の本質である．