

Note "THE USE OF THE BACK-AND-FORTH METHOD FOR WASSERSTEIN GRADIENT FLOWS TO SOLVE PDES"

坂井幸人

2023 年 6 月 4 日

1 Dual Problem

Condition

$\mathcal{P}(\mathbb{R}^n)$: (\mathbb{R}^n) 上の確率測度, 非負の測度で質量 (mass) 1

U : lsc(lower semi continuous) on $\mathcal{P}(\mathbb{R}^n)$

簡単のため、積分が 1 となる非負の $L^1(\mathbb{R}^n)$ で考える。

JKO スキーム

JKO スキームについて考える:

$$\min_{\rho \in \mathcal{P}} U(\rho) + \frac{1}{2\tau} W_2^2(\rho, \mu),$$

古い論文 [JL] からの記号を使用すると、2-Wasserstein 距離の Kantorovich の双対公式は次のように表されます:

$$\frac{1}{2\tau} W_2^2(\rho, \mu) = \sup_{(\phi, \psi) \in \mathcal{C}} \left(\int \phi d\rho + \int \psi d\mu \right),$$

ここで、 \mathcal{C} は制約

$$\mathcal{C} := \{(\phi, \psi) \in C(\Omega) \times C(\Omega) : \psi(x) + \phi(y) \leq \frac{1}{2\tau} |x - y|^2\}.$$

を満たす関数 (ϕ, ψ) の集合を表します。重要な点として、集合 \mathcal{C} は凸であることに注意してください。
[Exercise 1.1]

Exercise 1.1. \mathcal{C} が convex set であることを示せ。(Proof 4)

よって,

$$\begin{aligned} \min_{\rho \in \mathcal{P}} U(\rho) + \frac{1}{2\tau} W_2^2(\rho, \mu) &= \min_{\rho \in \mathcal{P}} \left(U(\rho) + \sup_{(\phi, \psi) \in \mathcal{C}} \left(\int \phi d\rho + \int \psi d\mu \right) \right) \\ &= \min_{\rho \in \mathcal{P}} \sup_{(\phi, \psi) \in \mathcal{C}} \left(U(\rho) + \int \phi d\rho + \int \psi d\mu \right) \end{aligned}$$

次に、 \mathcal{P} と \mathcal{C} が凸であり、 U が凸であるとする、関数

$$L(\rho, (\varphi, \psi)) := U(\rho) + \int \varphi d\rho + \int \psi d\mu$$

は ρ に関して凸関数 ((φ, ψ) が固定された場合) ([Proof](#))、および (φ, ψ) に関しては凹関数 (実際には線形) です ([Proof](#)) (さらにおそらく、いくつかの condition が必要である)。

このため、[ET, Ch. VI, Prop. 2.4 (p176)] のような最小最大の定理を適用して、 \min と \sup の順序を交換することができます。したがって、次のように結論付けることができます：

$$\begin{aligned} \min_{\rho \in \mathcal{P}} U(\rho) + \frac{1}{2\tau} W_2^2(\rho, \mu) &= \sup_{(\varphi, \psi) \in \mathcal{C}} \min_{\rho \in \mathcal{P}} \left(U(\rho) + \int \varphi d\rho + \int \psi d\mu \right), \\ &= \sup_{(\varphi, \psi) \in \mathcal{C}} \left(\min_{\rho \in \mathcal{P}} \left(U(\rho) + \int \varphi d\rho \right) + \int \psi d\mu \right), \\ &= \sup_{(\varphi, \psi) \in \mathcal{C}} \left(\int \psi d\mu - U^*(-\varphi) \right). \end{aligned}$$

ここで、 $U^*(\varphi)$ は以下のように定義している。

$$U^*(\varphi) := \sup_{\rho \in \mathcal{P}} \left(\int \varphi d\rho - U(\rho) \right).$$

detail :

$$\begin{aligned} U^*(\varphi) &= \sup_{\rho \in \mathcal{P}} \left(\int \varphi d\rho - U(\rho) \right), \\ U^*(-\varphi) &= \sup_{\rho \in \mathcal{P}} \left(\int -\varphi d\rho - U(\rho) \right), \\ &= \sup_{\rho \in \mathcal{P}} \left(- \int \varphi d\rho - U(\rho) \right), \\ &= \sup_{\rho \in \mathcal{P}} \left\{ - \left(\int \varphi d\rho + U(\rho) \right) \right\}, \\ &= - \inf_{\rho \in \mathcal{P}} \left\{ \left(\int \varphi d\rho + U(\rho) \right) \right\}. \end{aligned}$$

よって、

$$\min_{\rho \in \mathcal{P}} \left(U(\rho) + \int \varphi d\rho \right) + \int \psi d\mu = -U^*(-\varphi)$$

注意しておきますが、与えられた $(\varphi, \psi) \in \mathcal{C}$ に対して、 $\psi^c \geq \varphi$ が成り立ちます。ただし、

$$\psi^c(x) := \inf_y \left(\frac{1}{2\tau} |x - y|^2 - \psi(y) \right)$$

は ψ の c-変換 (c-transform) です。なぜなら、 $\mathcal{C} := \{(\phi, \psi) \in C(\Omega) \times C(\Omega) : \psi(x) + \phi(y) \leq \frac{1}{2\tau} |x - y|^2\}$ であり、 $\varphi(x) \leq \frac{1}{2\tau} |x - y|^2 - \psi(y)$ なので、 $\varphi(x)$ の中での \sup が ψ^c になるためである。

また、 $\rho \geq 0$ の場合、 $-U^*(-\varphi)$ は φ に関して増加する関数です。よって、 $-U^*(-\varphi) \leq -U^*(-\psi^c)$ である。

したがって、以下ようになります。

$$\sup_{(\varphi, \psi) \in \mathcal{C}} \left(\int \psi d\mu - U^*(-\varphi) \right) \leq \sup_{\psi} \left(\int \psi d\mu - U^*(-\psi^c) \right)$$

また、 $(\varphi, \psi) \in \mathcal{C} \implies (\psi^c, \psi) \in \mathcal{C}$ であるので、

$$\sup_{(\varphi, \psi) \in \mathcal{C}} \left(\int \psi d\mu - U^*(-\varphi) \right) \geq \sup_{\psi} \left(\int \psi d\mu - U^*(-\psi^c) \right)$$

が成立する。よって、

$$\sup_{(\varphi, \psi) \in \mathcal{C}} \left(\int \psi d\mu - U^*(-\varphi) \right) = \sup_{\psi} \left(\int \psi d\mu - U^*(-\psi^c) \right) \quad (1.1)$$

同様に、 $\mu \geq 0$ であるため、

$$\sup_{(\varphi, \psi) \in \mathcal{C}} \left(\int \psi d\mu - U^*(-\varphi) \right) = \sup_{\varphi} \left(\int \varphi^c d\mu - U^*(-\varphi) \right) \quad (1.2)$$

となります。

正規化アルゴリズムの主要なアイデアは、上記の右辺の ϕ と ψ の 2 つの関数に対して交互に勾配上昇ステップを実行することであり、 c -変換を使用して ϕ と ψ を変換します。勾配は適切な重み付き Sobolev 空間で計算されます。

2 Porous medium equation(多孔質媒体方程式)

多孔質媒体方程式 (PME) は、固定された $m > 1$ に対して以下の偏微分方程式 (PDE) のことをいう：

$$\rho_t = \frac{\partial \rho}{\partial t} = \Delta(\rho^m)$$

ここで、非負の解 $\rho \geq 0$ に興味があります。この PDE は、エネルギー関数

$$U(\rho) := \frac{1}{m-1} \int \rho^m dx$$

に基づく Wasserstein 勾配フローとして表現することができます。

$\rho \in \mathcal{P}(\mathbb{R}^n) \setminus L^m(\mathbb{R}^n)$ のとき、 $U(\rho)$ は $+\infty$ と定義されているとする。ただし、これは $s \mapsto s^m$ が $[0, \infty)$ 上で凸関数であるため、 $U(\rho)$ は $\mathcal{P}(\mathbb{R}^n)$ 上で凸な汎関数です。[JLL] のアルゴリズムでは、 δU^* を計算する必要があります。ここで、

$$\begin{aligned} U^*(\varphi) &= \sup_{\rho \in \mathcal{P}} \left(\int \varphi d\rho - U(\rho) \right). \\ &= \sup_{\rho \in \mathcal{P}} \int \left(-\frac{1}{m-1} \rho^m + \rho \varphi \right) dx \end{aligned}$$

となります。 φ の前の符号が [JLL] とは異なることに注意してください、しかしこれは問題ありません。

U^* は、実質的には U の Lagrange-Fenchel 変換を $-\varphi$ に対して行ったものですが、重要な違いとして、 \mathcal{P} がヒルベルト空間の部分集合ではなく、測度と連続関数の間の双対性を使って内積の代わりに $\int \varphi d\rho$ の積分を扱っています。

どちらにせよ、 $\delta U^*(\varphi)$ は通常の設定における $\partial U^*(\varphi)$ に類似しており、その場合、下半連続（つまり閉じている）凸関数 f について、次の関係が成り立ちます。

$$x \in \partial f^*(y) \iff z \cdot y - f(z) \text{ が } z = x \text{ において最大値を取る}$$

言い換えると、

$$\partial f^*(y) = \operatorname{argmax}_x (x \cdot y - f(x))$$

よって、 δU^* を見つけるために、以下の最大値を求める必要がある。

$$V(\rho) := \int \left(-\frac{1}{m-1} \rho^m + \rho \varphi \right) dx.$$

Lemma 2.1. $\varphi \in \mathcal{C}$ と仮定し、以下のように定義されたとする。

$$\rho_*(x) := \left(\frac{m-1}{m} (C + \varphi)_+ \right)^{\frac{1}{m-1}}$$

ただし、 $C \in \mathbb{R}$ は $\int \rho_* = 1$ となる。 $(s)_+ := \max(s, 0)$ と定義している。

この時、 ρ_* は $\mathcal{P}(\mathbb{R}^n)$ 上の関数 V の最大化関数。

Proof. C の選び方により、 $\rho_* \in \mathcal{P}(\mathbb{R}^n)$ であることがわかる。次に、以下を示す。

$$V(\rho) \leq V(\rho_*) \quad \text{for all } \rho \in \mathcal{P}(\mathbb{R}^n) \cap L^1(\mathbb{R}^n).$$

ρ を固定し、

$$\mu(x) := \rho(x) - \rho_*(x)$$

とします。注意点として、

$$\int \mu dx = 0 \tag{2.1}$$

であり、また $\mu(x) \geq 0$ は $\rho_*(x) = 0$ の場所、つまり $\varphi(x) \geq C$ の場所で成り立ちます。

よって、

$$\begin{aligned} V(\rho) - V(\rho_*) &= V(\rho_* + \mu) - V(\rho_*) \\ &= \int \left(-\frac{1}{m-1} ((\rho_* + \mu)^m - \rho_*^m) + \mu \varphi \right) dx \end{aligned}$$

ここで、関数 $s \mapsto s^m$ は $[0, \infty)$ 上で凸であるため、 $s, t \geq 0$ に対して

$$(s+t)^m \geq s^m + ms^{m-1}t, \quad s+t \geq 0$$

が成り立ちます。この不等式を適用することで、さらに次のように簡略化できます：

$$\begin{aligned} V(\rho) - V(\rho_*) &\leq \int \left(-\frac{1}{m-1} ((\rho_*^m + m\rho_*^{m-1}\mu) - \rho_*^m) + \mu\varphi \right) dx \\ &\leq \int \left(-\frac{m}{m-1} \rho_*^{m-1}\mu + \mu\varphi \right) dx \end{aligned}$$

ρ_* の式を利用することで、

$$\begin{aligned} \int \left(\frac{m}{m-1} \rho_*^{m-1}\mu + \mu\varphi \right) dx &= \int (-(C+\varphi)_+\mu + \mu\varphi) dx \\ &= \int (-(C+\varphi)_+\mu + \mu\varphi + C\mu) dx \\ &= \int (-(C+\varphi)_+ + \varphi + C)\mu dx \end{aligned}$$

ただし、(2.1) を利用する。

ここで、以下のような観察をします：

$$-(C+\varphi)_+ + \varphi + C = -(C+\varphi)_- \begin{cases} = 0 & \text{if } \varphi > -C \\ \leq 0 & \text{if } \varphi \leq -C \end{cases}$$

また、 $\mu(x) \geq 0$ は、 $\varphi(x) \geq C$ のとき成り立ちます。したがって、

$$V(\rho) - V(\rho_*) \leq \int (-(C+\varphi)_+ + \varphi + C)\mu dx \leq 0$$

したがって、 ρ_* は V の最大化点であることがわかります。 □

3 GRADIENT ASCENT

関数 (1.1), (1.2) の交互に勾配上昇を行う。

$$J(\varphi) = \int \varphi^c d\mu - U^*(-\varphi),$$

$$I(\psi) = \int \psi d\mu - U^*(-\psi^c).$$

φ, ψ が c -凹の場合、の時、 H 空間での勾配は以下ようになる。

$$\nabla_H J(\varphi) = (\theta_1 I - \theta_2 \Delta)^{-1} (\delta U^*(-\varphi) - T_{\varphi\#}\mu),$$

$$\nabla_H I(\psi) = (\theta_1 I - \theta_2 \Delta)^{-1} (\mu - T_{\psi\#}\delta U^*(-\psi^c)).$$

また、以下が成り立つ。

$$\begin{aligned} T_\varphi(x) &= x - \tau \nabla \varphi^c(x), \\ T_\varphi^{-1}(x) &= x - \tau \nabla \varphi(x) \end{aligned}$$

[JLL] の Proposition 2.4. を参照すること。

もし μ と φ が十分に滑らかであり、 φ が c -凹関数である場合、写像の変数変換の公式を用いて、押し出し密度 (pushforward density) は以下のように求められる。

$$\begin{aligned} T_{\varphi\#}\mu(x) &= \mu(T_\varphi^{-1}(x)) |\det \nabla T(T_\varphi^{-1}(x))|^{-1}, \\ &= \mu(x - \tau \nabla \varphi(x)) |\det(I - \tau D^2 \varphi(x))|. \end{aligned}$$

4 appendix

Proof of Exercise 1.1. 集合 C が凸であることを示すためには、 C 内の任意の 2 点 (φ_1, ψ_1) と (φ_2, ψ_2) を結ぶ線分、すなわち

$$[(t\varphi_1 + (1-t)\varphi_2, t\psi_1 + (1-t)\psi_2)]$$

が C に含まれることを示せばよい。

C 内の 2 点 (φ_1, ψ_1) と (φ_2, ψ_2) を考え、制約条件 $\varphi_1(x) + \psi_1(y) \leq \frac{1}{2\tau}|x-y|^2$ および $\varphi_2(x) + \psi_2(y) \leq \frac{1}{2\tau}|x-y|^2$ をすべての $x, y \in \mathbb{R}^n$ に対して満たすとする。

線分 $[(t\varphi_1 + (1-t)\varphi_2, t\psi_1 + (1-t)\psi_2)]$ 上の点 (φ, ψ) を考える。この点は $t \in [0, 1]$ に対して $(\varphi, \psi) = (t\varphi_1 + (1-t)\varphi_2, t\psi_1 + (1-t)\psi_2)$ とパラメータ化できる。

次に、 (φ, ψ) が制約条件 $\varphi(x) + \psi(y) \leq \frac{1}{2\tau}|x-y|^2$ をすべての $x, y \in \mathbb{R}^n$ に対して満たすかどうかを確認する：

$$\begin{aligned} \varphi(x) + \psi(y) &= (t\varphi_1 + (1-t)\varphi_2)(x) + (t\psi_1 + (1-t)\psi_2)(y) \\ &= t(\varphi_1(x) + \psi_1(y)) + (1-t)(\varphi_2(x) + \psi_2(y)) \\ &\leq t \left(\frac{1}{2\tau}|x-y|^2 \right) + (1-t) \left(\frac{1}{2\tau}|x-y|^2 \right) \\ &= \frac{1}{2\tau}|x-y|^2 \end{aligned}$$

よって、 (φ, ψ) が制約条件を満たすことがわかり、 $(\varphi, \psi) \in C$ となります。したがって、集合 C が凸であることが示された。 \square

Proof of ρ が変数とした時 $L(\rho, (\varphi, \psi))$ が凸関数. 関数 $L(\rho, (\varphi, \psi)) := U(\rho) + \int \varphi d\rho + \int \psi d\mu$ が、 ρ を変数として固定された (φ, ψ) の関数として凸関数であることを証明します。

まず、 $L(\rho, (\varphi, \psi))$ が (φ, ψ) に関して線形であることを示します。つまり、

$$\begin{aligned} L(\rho, (t\varphi_1 + (1-t)\varphi_2, t\psi_1 + (1-t)\psi_2)) &= L(\rho, t(\varphi_1, \psi_1) + (1-t)(\varphi_2, \psi_2)) \\ &= tL(\rho, (\varphi_1, \psi_1)) + (1-t)L(\rho, (\varphi_2, \psi_2)) \end{aligned}$$

を示します。

$$\begin{aligned}
L(\rho, (t\varphi_1 + (1-t)\varphi_2, t\psi_1 + (1-t)\psi_2)) &= L(\rho, t(\varphi_1, \psi_1) + (1-t)(\varphi_2, \psi_2)) \\
&= U(\rho) + \int (t\varphi_1 + (1-t)\varphi_2) d\rho + \int (t\psi_1 + (1-t)\psi_2) d\mu \\
&= t(U(\rho) + \int \varphi_1 d\rho + \int \psi_1 d\mu) + (1-t)(U(\rho) + \int \varphi_2 d\rho + \int \psi_2 d\mu) \\
&= tL(\rho, (\varphi_1, \psi_1)) + (1-t)L(\rho, (\varphi_2, \psi_2))
\end{aligned}$$

次に、 $L(\rho, (\varphi, \psi))$ が凸関数であることを示します。つまり、 $\rho_1, \rho_2 \in \mathcal{P}$ および $0 \leq \lambda \leq 1$ に対して、以下の不等式が成り立つことを示します：

$$L(\lambda\rho_1 + (1-\lambda)\rho_2, (\varphi, \psi)) \leq \lambda L(\rho_1, (\varphi, \psi)) + (1-\lambda)L(\rho_2, (\varphi, \psi))$$

$$\begin{aligned}
L(\lambda\rho_1 + (1-\lambda)\rho_2, (\varphi, \psi)) &= U(\lambda\rho_1 + (1-\lambda)\rho_2) + \int \varphi d(\lambda\rho_1 + (1-\lambda)\rho_2) + \int \psi d\mu \\
&= \lambda(U(\rho_1) + \int \varphi d\rho_1 + \int \psi d\mu) + (1-\lambda)(U(\rho_2) + \int \varphi d\rho_2 + \int \psi d\mu) \\
&= \lambda L(\rho_1, (\varphi, \psi)) + (1-\lambda)L(\rho_2, (\varphi, \psi))
\end{aligned}$$

したがって、関数 $L(\rho, (\varphi, \psi))$ は (φ, ψ) を固定した場合に凸関数。 □

Proof of (φ, ψ) に関して $L(\rho, (\varphi, \psi))$ は凹関数. 関数 $L(\rho, (\varphi, \psi)) := U(\rho) + \int \varphi d\rho + \int \psi d\mu$ が (φ, ψ) を変数として固定された ρ の関数として凹関数（実際には線形関数）であることを示します。

まず、 $L(\rho, (\varphi, \psi))$ が ρ に関して線形であることを示します。つまり、

$$L(t\rho_1 + (1-t)\rho_2, (\varphi, \psi)) = tL(\rho_1, (\varphi, \psi)) + (1-t)L(\rho_2, (\varphi, \psi))$$

を示す。

$$\begin{aligned}
L(t\rho_1 + (1-t)\rho_2, (\varphi, \psi)) &= U(t\rho_1 + (1-t)\rho_2) + \int \varphi d(t\rho_1 + (1-t)\rho_2) + \int \psi d\mu \\
&= tU(\rho_1) + (1-t)U(\rho_2) + \int \varphi d(t\rho_1 + (1-t)\rho_2) + \int \psi d\mu \\
&= t(U(\rho_1) + \int \varphi d\rho_1 + \int \psi d\mu) + (1-t)(U(\rho_2) + \int \varphi d\rho_2 + \int \psi d\mu) \\
&= tL(\rho_1, (\varphi, \psi)) + (1-t)L(\rho_2, (\varphi, \psi))
\end{aligned}$$

次に、 $L(\rho, (\varphi, \psi))$ が凹関数であることを示します。つまり、 $\rho_1, \rho_2 \in \mathcal{P}$ および $0 \leq \lambda \leq 1$ に対して、以下の不等式が成り立つことを示す：

$$L(\lambda\rho_1 + (1-\lambda)\rho_2, (\varphi, \psi)) \geq \lambda L(\rho_1, (\varphi, \psi)) + (1-\lambda)L(\rho_2, (\varphi, \psi))$$

$$\begin{aligned}
L(\lambda\rho_1 + (1-\lambda)\rho_2, (\varphi, \psi)) &= U(\lambda\rho_1 + (1-\lambda)\rho_2) + \int \varphi d(\lambda\rho_1 + (1-\lambda)\rho_2) + \int \psi d\mu \\
&= \lambda(U(\rho_1) + \int \varphi d\rho_1 + \int \psi d\mu) + (1-\lambda)(U(\rho_2) + \int \varphi d\rho_2 + \int \psi d\mu) \\
&= \lambda L(\rho_1, (\varphi, \psi)) + (1-\lambda)L(\rho_2, (\varphi, \psi))
\end{aligned}$$

以上より、関数 $L(\rho, (\varphi, \psi))$ は (φ, ψ) を変数として固定された ρ の関数としては凹関数（実際には線形関数）であることが示されました。 \square

Definition 4.1. (lower limits) Let $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ and let x' is a limit point of f . Then the lower limit of function f is defined by

$$\begin{aligned} \liminf_{x \rightarrow x'} f(x) &= \lim_{\delta \searrow 0} \left[\inf_{x \in B(x', \delta)} f(x) \right] \\ &= \sup_{\delta > 0} \left[\inf_{x \in B(x', \delta)} f(x) \right] = \sup_{V \in \mathcal{N}(x')} \left[\inf_{x \in V} f(x) \right]. \end{aligned}$$

Definition 4.2. (lower semi-continuous) Let $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ and let x' is a limit point of f . Then f is lower semi-continuous at x' if and only if

$$\liminf_{x \rightarrow x'} f(x) \geq f(x'), \text{ or } \liminf_{x \rightarrow x'} f(x) = f(x')$$