

# 有限集合上の最適輸送問題と Bregman ダイバージェンス

高津 飛鳥 (東京都立大学)

概要. 本稿では, 有限集合上の最適輸送問題の Bregman ダイバージェンスによる緩和問題, すなわち有界凸集合上の線形関数に摂動項として狭義凸関数を加えた最小化問題に対する [4] の結果を述べ, 結果に現れる仮定の妥当性に関して論じる.

## 1 有限集合の最適輸送問題

有限集合上の最適輸送問題の概略は, 例えば [6] を参考していただくことにして, ここでは記号の導入のみを行う.

### 1.1 記号

- $N, I, J$ : 固定された自然数.
- $n, i, j$ : 添字に現れる自然数.  
指定がない場合, 小文字アルファベットは 1 から大文字アルファベットを亘るとする.

- $\mathcal{P}_N := \left\{ z = (z_n)_n \in \mathbb{R}^N \mid z_n \geq 0 \quad \forall n, \quad \sum_{n=1}^N z_n = 1 \right\}$ .

ここで  $\mathcal{P}_{I \times J}$  は  $M_{I \times J}(\mathbb{R})$  の部分集合とみなす.

- ユークリッド空間の次元によらず標準内積を  $\langle \cdot, \cdot \rangle$  と記す.  
ここで  $M_{I \times J}(\mathbb{R})$  の Hilbert–Schmidt 内積に対しても  $\langle \cdot, \cdot \rangle$  を用いる.
- $(x, y) \in \mathcal{P}_I \times \mathcal{P}_J$  に対し

–  $x \otimes y \in \mathcal{P}_{I \times J}(\mathbb{R})$  を各  $i, j$  に対し  $(x \otimes y)_{ij} := x_i y_j$  と定める.

–  $\Pi(x, y) := \left\{ \Pi = (\pi_{ij})_{ij} \in \mathcal{P}_{I \times J} \mid \sum_{k=1}^J \pi_{ik} = x_i \quad \forall i, \quad \sum_{k=1}^I \pi_{kj} = y_j \quad \forall j \right\}$  は  
内点に  $x \otimes y$  を含む有界な閉凸多面体となる.

### 1.2 問題

各  $(C, x, y) \in M_{I \times J}(\mathbb{R}) \times \mathcal{P}_I \times \mathcal{P}_J$  に対し,

$$\inf_{\Pi \in \Pi(x, y)} \langle C, \Pi \rangle \tag{1.1}$$

の値を求めよ.

### 1.3 解の存在と性質

各  $(C, x, y) \in M_{I \times J}(\mathbb{R}) \times \mathcal{P}_I \times \mathcal{P}_J$  に対し,  $\langle C, \cdot \rangle$  は空でない有界閉集合  $\Pi(x, y)$  上の線形関数, とくに連続関数なので, 必ず最小化因子が存在する. しかし最小化因子は一意的であるとは限らない. また, 最小化因子が一意的であるときは必ず  $\Pi(x, y)$  の境界に属することが [8, Theorem 5.10] から従う (例えば,  $I = J = 2$  の場合を考えてみると分かりやすい).

最小値を求める手法は色々あるが, 一部の手法は解の一意性がない場合や, 最小化因子が定義域の境界に出てくる場合は適用しがたい. そこで凸関数で緩和することを考える.

## 2 Bregman ダイバージェンス

有界閉凸集合上の線形関数に摂動項として狭義凸関数を加えた最小化問題では, 最小化因子が必ず一意的に存在する. 有限集合上の最適輸送問題の凸緩和問題において, 最小化因子を  $\Pi(x, y)$  の内点にすべく凸関数の影響を用いて  $x \otimes y$  の方に引き寄せられることを考えると, ダイバージェンスによる摂動が候補にあがる. ここで  $\mathcal{P}_N$  上のダイバージェンス  $D$  とは,  $D(z, w) = 0$  であるための必要十分条件が  $z = w$  である関数  $D: \mathcal{P}_N \times \mathcal{P}_N \rightarrow [0, \infty]$  のことである. ダイバージェンスについては, 例えば [1, Section 2] を参照にいただきたい. 我々はダイバージェンスの中でも, Bregman ダイバージェンスを用いる.

### 2.1 記号

- $U \in C([0, 1]) \cap C^1((0, 1))$ : 狭義凸関数.
- $d_U: [0, 1] \times [0, 1] \rightarrow \mathbb{R}$  を

$$d_U(r, s) := U(r) - U(s) - (r - s)U'(s)$$

と定義する. すると  $U$  の狭義凸性から  $d_U$  は必ず非負になる. さらに自然な拡張  $d_U: [0, 1] \times [0, 1] \rightarrow [0, \infty]$  が存在する.

- Bregman ダイバージェンスを  $D_U: \mathcal{P}_N \times \mathcal{P}_N \rightarrow [0, \infty]$  を

$$D_U(z, w) := \sum_{n=1}^N d_U(z_n, w_n)$$

と定義する. すると  $D_U(z, w) = 0$  が成り立つのは  $z = w$  の場合に限る.

注記 2.1 より一般に  $\mathcal{P}_N$  ではなくユークリッド空間上の確率測度のなす空間を考える場合は,  $U$  の定義域は  $[0, 1]$  ではなく  $[0, \infty)$  である必要がある. Bregman ダイバージェンスについては, 例えば [5, Section 8.7] を参照にいただきたい.

### 2.2 緩和された問題

狭義凸関数  $U \in C([0, 1]) \cap C^1((0, 1))$  を固定する. 各  $(C, x, y) \in M_{I \times J}(\mathbb{R}) \times \mathcal{P}_I \times \mathcal{P}_J$  および  $\varepsilon \in (0, \infty)$  に対し,

$$\inf_{\Pi \in \Pi(x, y)} (\langle C, \Pi \rangle + \varepsilon D_U(\Pi, x \otimes y)) \quad (2.1)$$

の最小化因子を  $\Pi^U(C, x, y, \varepsilon)$  と記す. このとき,  $\langle C, \Pi^U(C, x, y, \varepsilon) \rangle$  の値を求めよ.

### 3 定量的評価

以下,  $(C, x, y) \in M_{I \times J}(\mathbb{R}) \times \mathcal{P}_I \times \mathcal{P}_J$  を固定する.

狭義凸関数  $U \in C([0, 1]) \cap C^1((0, 1))$  に対し, 緩和問題の収束, すなわち

$$\lim_{\varepsilon \downarrow 0} \langle C, \Pi^U(C, x, y, \varepsilon) \rangle = \inf_{\Pi \in \Pi(x, y)} \langle C, \Pi \rangle$$

が期待され, 実際に成り立つ. このことについては, 例えば [4, Subsection 2.4] を参照にしていただきたい. 本稿では連続講演において説明した  $\langle C, \Pi^U(C, x, y, \varepsilon) \rangle - \inf_{\Pi \in \Pi(x, y)} \langle C, \Pi \rangle$  の上からの評価を述べるための仮定を2つ与える.

以下,  $V(x, y)$  を  $\Pi(x, y)$  の頂点のなす集合とする. また,

$$\mathfrak{D}_U(x, y) := \sup_{\Pi \in \Pi(x, y)} D_U(\Pi, x \otimes y)$$

とおく. すると  $\mathfrak{D}_U(x, y) \in [0, \infty)$  が成り立つ.

**条件 3.1** 問題 (1.1) の最小化因子がなす集合は  $\Pi(x, y)$  の真部分集合である.

条件 3.1 を仮定すると

$$\inf_{V \in V(x, y)} \langle C, V \rangle < \sup_{V \in V(x, y)} \langle C, V \rangle, \quad \mathfrak{D}_U(x, y) \in (0, \infty)$$

が成り立つ. そして

$$\Delta_C(x, y) := \inf_{V' \in V(x, y) \& \langle C, V' \rangle \neq \inf_{V \in V(x, y)} \langle C, V \rangle} \langle C, V' \rangle - \inf_{V \in V(x, y)} \langle C, V \rangle \in (0, \infty)$$

となる.

**条件 3.2**  $U \in C([0, 1]) \cap C^1((0, 1]) \cap C^2((0, 1))$  は以下の3条件を満たす.

- (1)  $\lim_{h \downarrow 0} U'(h) = -\infty$ .
- (2)  $(0, 1)$  上で  $U'' > 0$ .
- (3)  $(0, 1)$  上で  $r \mapsto rU''(r)$  は非減少.

条件 3.2(1),(2) より,  $U' : (0, 1] \rightarrow (-\infty, U'(1)]$  の逆関数が存在する. この逆関数を  $e_U$  と書く. また, 条件 3.2(2) より  $r \mapsto U'(r) - U'(1 - r)$  は  $(0, 1)$  上狭義単調増加であり, さらに条件 3.2(1) よりある  $R_U(x, y) \in [1/2, 1)$  が一意的に存在し

$$\mathfrak{D}_U(x, y) = U'(R_U(x, y)) - U'(1 - R_U(x, y))$$

が成り立つ. そして条件 3.2(2),(3) より

$$\nu_U(x, y) := \sup_{r \in (0, R_U(x, y))} (U'(1 - r) + rU''(r))$$

は有界である. さらに条件 3.1 を仮定すると  $\mathfrak{D}_U(x, y) + \nu_U(x, y) - U'(1) \in (0, \infty)$  が成り立つ.

定理 3.3 条件 3.1, 3.2 を仮定する. 任意の  $(C, x, y) \in M_{I \times J}(\mathbb{R}) \times \mathcal{P}_I \times \mathcal{P}_J$  および

$$\varepsilon \in \left(0, \frac{\Delta_C(x, y) R_U(x, y)}{\mathfrak{D}_U(x, y)}\right] \cap \left(0, \frac{\Delta_C(x, y)}{\mathfrak{D}_U(x, y) + \nu_U(x, y) - U'(1)}\right]$$

に対し

$$\langle C, \Pi^U(C, x, y, \varepsilon) \rangle - \inf_{\Pi \in \Pi(x, y)} \langle C, \Pi \rangle \leq \Delta_C(x, y) \cdot e_U \left( -\frac{\Delta_C(x, y)}{\varepsilon} + \mathfrak{D}_U(x, y) + \nu_U(x, y) \right)$$

が成り立つ.

### 3.1 条件 3.1, 3.2 についての考察

条件 3.1 が成り立たない場合, 問題 (1.1) は容易に解けるので緩和問題を考える必要はない. そこで条件 3.1 は自然な仮定である.

狭義凸関数  $U \in C([0, 1]) \cap C^1((0, 1])$  を考える. すると [7, Lemma 3.7&Remark 3.9] より, 条件 3.2(1) は, 各  $(C, x, y) \in M_{I \times J}(\mathbb{R}) \times \mathcal{P}_I \times \mathcal{P}_J$  と  $\varepsilon \in (0, \infty)$  に対し  $\Pi^U(C, x, y, \varepsilon)$  が  $\Pi(x, y)$  の内点になることに同値である. そして Aleksandrov の定理 [2, pp.241–245] より,  $U$  は  $(0, 1)$  上殆ど至る所 2 階微分可能である. そこで  $U \in C([0, 1]) \cap C^1((0, 1]) \cap C^2((0, 1))$  としたとき,  $U$  が狭義凸関数ならば  $U'' > 0$  が  $(0, 1)$  上殆ど至る所成り立つ. よって条件 3.2(2) は  $U$  の正則性をやや強めたものと言える.

最後に, 条件 3.2(3) を考える. そこで  $U$  は条件 3.2 を満たすとし, 実数  $\mu_0, \mu_1$  および正数  $\lambda$  に対し, 関数  $U_{\lambda, \mu_0, \mu_1} : [0, 1] \rightarrow \mathbb{R}$  を

$$U_{\lambda, \mu_0, \mu_1}(r) := \lambda U(r) + \mu_1 r + \mu_0$$

と定めれば,  $U_{\lambda, \mu_0, \mu_1}$  も条件 3.2 を満たす. このとき  $D_{U_{\lambda, \mu_0, \mu_1}} = \lambda D_U$  が  $\mathcal{P}_N \times \mathcal{P}_N$  上で成り立つので, 任意の  $\varepsilon \in (0, \infty)$  に対し

$$\mathfrak{D}_{U_{\lambda, \mu_0, \mu_1}}(x, y) = \lambda \mathfrak{D}_U(x, y), \quad \Pi^{U_{\lambda, \mu_0, \mu_1}}(C, x, y, \varepsilon) = \Pi^U(C, x, y, \lambda \varepsilon)$$

となる. さらに  $U'_{\lambda, \mu_0, \mu_1}(r) - U'_{\lambda, \mu_0, \mu_1}(1-r) = \lambda(U'(r) - U'(1-r))$  が  $(0, 1)$  上で成り立つので

$$R_{U_{\lambda, \mu_0, \mu_1}}(x, y) = R_U(x, y)$$

であり,  $U'_{\lambda, \mu_0, \mu_1}(1-r) + rU''_{\lambda, \mu_0, \mu_1}(r) = \lambda(U'(1-r) + rU''(r)) + \mu_1$  が  $(0, 1)$  上で成り立つので

$$\nu_{U_{\lambda, \mu_0, \mu_1}}(x, y) = \lambda \nu_U(x, y) + \mu_1$$

となる. また,  $\tau \in (-\infty, U'_{\lambda, \mu_0, \mu_1}(1)] = (-\infty, \lambda U'(1) + \mu_1]$  に対し

$$e_{U_{\lambda, \mu_0, \mu_1}}(\tau) = e_U(\lambda^{-1}(\tau - \mu_1))$$

が成り立つ. 以上より,

$$\begin{aligned} \varepsilon \in \left(0, \frac{\Delta_C(x, y) R_{U_{\lambda, \mu_0, \mu_1}}(x, y)}{\mathfrak{D}_{\lambda, \mu_0, \mu_1}(x, y)}\right] \cap \left(0, \frac{\Delta_C(x, y)}{\mathfrak{D}_{U_{\lambda, \mu_0, \mu_1}}(x, y) + \nu_{U_{\lambda, \mu_0, \mu_1}}(x, y) - U'_{\lambda, \mu_0, \mu_1}(1)}\right] \\ \iff \lambda \varepsilon \in \left(0, \frac{\Delta_C(x, y) R_U(x, y)}{\mathfrak{D}_U(x, y)}\right] \cap \left(0, \frac{\Delta_C(x, y)}{\mathfrak{D}_U(x, y) + \nu_U(x, y) - U'(1)}\right] \end{aligned}$$

であり, このとき定理 3.3 から

$$\begin{aligned} & \langle C, \Pi^{U, \mu_0, \mu_1}(C, x, y, \varepsilon) \rangle - \inf_{\Pi \in \Pi(x, y)} \langle C, \Pi \rangle \\ & \leq \Delta_C(x, y) e_{U, \mu_0, \mu_1} \left( -\frac{\Delta_C(x, y)}{\varepsilon} + \mathfrak{D}_{U, \mu_0, \mu_1}(x, y) + \nu_{U, \mu_0, \mu_1}(x, y) \right) \end{aligned}$$

が従い, これは

$$\langle C, \Pi^U(C, x, y, \lambda \varepsilon) \rangle - \inf_{\Pi \in \Pi(x, y)} \langle C, \Pi \rangle \leq \Delta_C(x, y) \cdot e_U \left( -\frac{\Delta_C(x, y)}{\lambda \varepsilon} + \mathfrak{D}_U(x, y) + \nu_U(x, y) \right)$$

と同値である. 以上より,  $u'_1 > u_1 - u_0$  を満たす実数  $u_0, u_1, u'_1$  に対し

$$U(0) = u_0, \quad U(1) = u_1, \quad U'(1) = u'_1 \quad (3.1)$$

を仮定しても, 定理 3.3 の評価は本質的には変わらない.

さらに  $(0, 1)$  上で  $U'' > 0$  を満たす  $U \in C([0, 1]) \cap C^1((0, 1]) \cap C^2((0, 1))$  に対し,  $q_U : (0, 1) \rightarrow [-\infty, \infty]$  を

$$q_U(r) := rU''(r) \cdot \limsup_{h \downarrow 0} \frac{1}{h} \left( \frac{1}{U''(r+h)} - \frac{1}{U''(r)} \right)$$

と定める. すると  $U, W \in C([0, 1]) \cap C^1((0, 1]) \cap C^2((0, 1))$  が  $(0, 1)$  上で  $U'', W'' > 0$  を満たすならば

ある実数  $\mu_0, \mu_1$  および正数  $\lambda$  が存在し,  $W = U_{\lambda, \mu_0, \mu_1} \implies (0, 1)$  上で  $q_U = q_W$

が成り立つ. さらに,  $q_U, q_W$  が  $(0, 1)$  上で殆ど至る所有限のときは上記の命題の逆が成り立つことが [3, Theorem 2.4] から従う. よって,  $q_U, q_W$  が  $(0, 1)$  上で殆ど至る所有限のとき, 正規化条件 (3.1) 下では  $U$  そのものではなく  $q_U$  を扱う方が本質的である.

そこで,  $U, W \in C([0, 1]) \cap C^1((0, 1]) \cap C^2((0, 1))$  は  $(0, 1)$  上で  $U'', W'' > 0$  を満たし, そして  $U, W$  はともに正規化条件 (3.1) を満たすとす. さらに  $q_U, q_W$  が  $(0, 1)$  上で殆ど至る所有限だとす. このとき,  $(0, 1)$  上で  $q_U \leq q_W$  が成り立つならば  $W'(1) \leq U'(1)$  かつ任意の  $\tau \in (-\infty, W'(1)]$  上で

$$e_U(\tau) \leq e_W(\tau) \quad (3.2)$$

が成り立つことが [3, Theorem 2.4] から従う. よって, 定理 3.3 の評価においては  $q_U$  は小さい方がよい. さらに  $U$  が条件 3.2(1) を満たすならば

$$Q_U := \sup_{r \in (0, 1)} q_U(r) \geq 1$$

となることが [3, Corollary 2.7] から従うので  $Q_U = 1$  は最適な条件である. そして条件  $Q_U = 1$  が条件 3.2(3) と同値であることが [3, Corollary 2.6] から従う.

以上の議論より, 条件 3.2 が妥当であると言える.

## 4 条件 3.2 を満たす $U$ の例

関数  $U_o : [0, 1] \rightarrow \mathbb{R}$  を

$$U_o(r) := \begin{cases} r \log r & (r \in (0, 1]) \\ 0 & (r = 0) \end{cases}$$

と定めれば,  $U_o$  は条件 3.2 と正規化条件 (3.1) を満たす. また,  $(0, 1)$  上で  $q_{U_o}$  は常に 1 である. そして任意の  $\tau \in (-\infty, U_o'(1)] = (-\infty, 1]$  に対し

$$e_{U_o}(\tau) = \exp(\tau - 1)$$

である. また,

$$R_{U_o}(x, y) = \frac{\exp(\mathfrak{D}_{U_o}(x, y))}{1 + \exp(\mathfrak{D}_{U_o}(x, y))}, \quad \nu_{U_o}(x, y) = 2$$

であり,

$$\left(0, \frac{\Delta_C(x, y) R_{U_o}(x, y)}{\mathfrak{D}_{U_o}(x, y)}\right] \cap \left(0, \frac{\Delta_C(x, y)}{\mathfrak{D}_{U_o}(x, y) + \nu_{U_o}(x, y) - U_o'(1)}\right] = \left(0, \frac{\Delta_C(x, y)}{1 + \mathfrak{D}_{U_o}(x, y)}\right]$$

が従う. このことは, 定理 3.3 の  $U = U_o$  の場合の評価と Weed の評価 [9, Theorem5] が一致することを保証する.

また, 条件 3.2, とくに条件 3.2(3) を満たす  $U$  に対し, もし  $q_U$  が  $(0, 1)$  上で殆ど至る所有限ならば

$$q_{U_o}(r) = 1 = Q_U = \sup_{r' \in (0, 1)} q_U(r') \geq q_U(r)$$

が  $(0, 1)$  上の殆ど全ての  $r$  に対し成り立つので, 不等式 (3.2) より, 我々の結果は [9, Theorem5] の結果を改善すると言える.

そして  $\alpha \in (0, 1]$  に対し

$$U_{\ell_\alpha}(r) := \begin{cases} -\frac{1}{\Gamma(\alpha + 1)} \Gamma(\alpha + 1, -\log r) + r & (r \in (0, 1]) \\ 0 & (r = 0) \end{cases}$$

と定めれば, これは条件 3.2 を満たす. ここで

$$\Gamma(p, \tau) = \int_\tau^\infty t^{p-1} \exp(-t) dt \quad (p > 0, \tau \geq 0)$$

は第 2 種不完全ガンマ関数である. とくに  $[0, 1]$  上で  $U_{\ell_1} = U_o$  である.

以上の例は, [4, Subsections 5.1, 5.3] で論じている.

## 参考文献

- [1] S. Eguchi, *Geometry of minimum contrast*, Hiroshima Math. J. **22** (1992), 631–647.
- [2] L. C. Evans and R. F. Gariepy, *Measure theory and fine properties of functions*, CRC Press, Boca Raton, FL, 2015.
- [3] K. Ishige, P. Salani, and A. Takatsu, *Hierarchy of deformations in concavity*, Inf. Geom., DOI 10.1007/s41884-022-00088-4.
- [4] K. Morikuni, K. Sakakibara, and A. Takatsu, *Error estimate for regularized optimal transport problems via Bregman divergence*, arXiv:2309.11666.
- [5] J. Naudts, *Generalised thermostatics*, Springer-Verlag London, Ltd., London, 2011.
- [6] 高津飛鳥, 輸送計画, 輸送写像, 輸送経路, 応用数理. **32** (2022), 69–79.
- [7] A. Takatsu, *Relaxation of optimal transport problem via strictly convex functions*, arXiv:2102.07336.
- [8] C. Villani, *Optimal transport*, Vol. 338, Springer-Verlag, Berlin, 2009.
- [9] J. Weed, *An explicit analysis of the entropic penalty in linear programming*, Proceedings of the 31st Conference on Learning Theory, 2018, pp. 1841–1855.