

# Note "A fast approach to optimal transport: the back-and-forth method"

坂井幸人

2023 年 5 月 9 日

## 1 Introduction

狭凸関数の最適輸送問題を効率よく解く解法。

$n$  個の点を持つ離散グリッド上の最適輸送マップを  $O(n \log(n))$  の操作と  $O(n)$  のストレージで計算。

一連のソースからシンクへの質量輸送の最も費用効率の良い方法を見つけることを目的としている。

この理論は、1942 年に Kantorovich によって近代化され、最適輸送と線形計画法の間に重要な関連性を見出しました。近年、最適輸送への関心は爆発的に高まっています。

これは、2 次コスト最適輸送問題と、統計力学や流体力学で発生する偏微分方程式 (PDE) の多様なクラスとの深い関連性が発見されたことに一因があります。

最適マップの計算は非常に困難な課題でした。私たちの知る限りでは、これまでに最適輸送問題を解決するための既知のすべての方法は、問題サイズに関して線形スケーリングしない [3,5]、最適マップを正確に計算できない [10]、または限られた確率密度のクラスにしか適用できない [4] という問題があります。

本論文の目的は、効率的かつ正確なアルゴリズムを提供し、最適輸送マップを計算することです。

新しい最適輸送問題の解法である「the back-and-forth method」を提案します。確定的凸コスト関数を持つ二つの確率密度関数  $\mu$  と  $\nu$  を、 $n$  点のグリッドに離散化した場合、前後法は一回の繰り返しあたり  $O(n \log(n))$  の計算時間と  $O(n)$  のストレージスペースを使用して最適マップを計算します。全ての実験で  $\epsilon$  の精度を達成するために必要なイテレーション数は、 $O(\max(\|\mu\|_\infty, \|\nu\|_\infty) \log(1/\epsilon))$  のように成長します。

この方法は非常に速く収束します。特筆すべきことに、他の多くの方法とは異なり、前後法は確率密度の正の下限を必要としません。そのため、大部分の領域でゼロになる密度であっても最適輸送問題を解くことができます。前後法の効率性と柔軟性により、以前の手法では扱えなかった最適輸送問題を解くことができます。

0

### 1.1 Overall approach

**Theorem 1.1** (カントロヴィッチ双対問題). 二つの確率測度  $\nu, \mu$  と、 $x$  から  $y$  への輸送にかかるコスト関数  $c(x, y)$  が与えられたとする。 $\nu$  を  $\mu$  に輸送する最適なコスト  $C$  は、Kantorovich 双対問題の値によって与えられる:

$$C = \sup_{\phi, \psi} \int \phi d\nu + \int \psi d\mu$$

ただし, *Kantorovich* ポテンシャル  $\phi(y), \psi(x)$  は, 以下の不等式制約を満たす 2 つのスカラー関数である。

$$\phi(y) + \psi(x) \leq c(x, y). \quad (1)$$

最適なコスト  $C$  は輸送する砂の体積を最大化させる問題として扱うことができる。

$\mu \rightarrow \nu$  への最適マップが存在すると、双対問題の最大値  $\phi_*(y), \psi_*(x)$  が復元できる。

そのとき,

$$\phi_*(y) + \psi_*(x) = c(x, y).$$

したがって、最適なマップを計算するには、双対問題を解くだけで十分。

$\phi_*$  と  $\psi_*$  が双対問題の最大値をとるとき、以下の関係が成立することに基づいている。

$$\begin{aligned} \phi_*(y) &= \psi_*^c(y) = \inf\{c(x, y) - \psi_*(x)\} \\ \psi_*(x) &= \phi_*^c(x) = \inf\{c(x, y) - \phi_*(y)\} \end{aligned}$$

ここで、連続関数  $\phi: \Omega \rightarrow \mathbb{R}$  に対し、その c-変換  $\phi^c: \Omega \rightarrow \mathbb{R}$  を次のように定義する。

$$\phi^c(x) := \inf_{y \in \Omega} \{c(x, y) - \phi(y)\}$$

従って、式 (1) の制約を取り除くためには、 $\phi$  を  $\phi^c$  に置き換えるか、 $\psi$  を  $\psi^c$  に置き換えるかのどちらかができます。

このことから、以下の 2 つの関数が得られます。

$$J(\phi) = \int \phi d\nu + \int \phi^c d\mu, \quad I(\psi) = \int \psi^c d\nu + \int \psi d\mu,$$

カントロヴィッチ双対問題  $C = \sup_{\phi, \psi} \int \phi d\nu + \int \psi d\mu$  は c-変換を用いて、 $J$  と  $I$  それぞれの sup と表すことができる。

なお、 $J$  は「 $\phi$  空間」で問題を定式化し、 $I$  は「 $\psi$  空間」で問題を定式化している。

The back-and-forth メソッドは、 $\phi$ -空間での  $J$  の勾配上昇の更新と  $\psi$ -空間での  $I$  の勾配上昇の更新を行うことで、*Kantorovich* の双対問題を解決します。勾配は  $\dot{H}^1$  距離 (式 (2) を参照) に関して取られます。勾配ステップの間に、一方の空間の情報は c-transform を取ることによって、他の空間に逆伝播されます (アルゴリズム 1 を参照)。

The back-and-forth method の利点は、最適解のペア  $(\phi_*, \psi_*)$  の特定の特徴が、片方の空間で他方の空間よりも簡単に構築できる場合があるということです。たとえば、 $\psi = \phi^c$  の Hessian は、 $\phi$  の Hessian の逆行列と非常に密接に関連しています。したがって、一方の空間で大きな Hessian 固有値は、他方の空間で小さな Hessian 固有値に対応し、より小さな特徴をより少ない勾配ステップで構築することができます。2 つの空間を行ったり来たりすることで、より良い状態になります。つまり、より小さな特徴を構築するのに最適な空間でそれを行う機会が常にあります。その結果、バックアンドフォース法は、 $\phi$  空間だけまたは  $\psi$  空間だけに作用する vanilla の勾配上昇法よりもはるかに迅速に収束します。実際、特定の例では、追加の 2~4 回の反復で

誤差が 10,000 倍に減少することがあります (表 1、2 を参照)。

$\dot{H}^1$  距離における勾配上昇ステップの選択がアルゴリズムの安定性を維持するために重要である。実際、勾配上昇ステップを  $\dot{H}^1$  よりも弱い Hilbert 空間で取ると、双対問題の値の増加を保証することはできないと思われます (詳細については 3.1 節および命題 1 を参照)。

計算効率に関しては、the back-and-forth method は  $J$  と  $I$  (つまり  $\phi$  空間と  $\psi$  空間) で交互に  $\dot{H}^1$  勾配上昇イテレーションを実行し、c-transforms を計算することに相当します。重要なことに、 $J$  の導関数 (および対称性から  $I$  の導関数) は単純な形式を取ることができ、効率的に計算できます [12]。さらに、多くのコスト関数に対して、c-transforms を非常に効率的に計算できます。離散グリッド上では、正確な c-transforms は  $O(n \log(n))$  の操作で計算できます [9,19] (詳細については Sect. 4.1 を参照)。

## 1.2 Future work and paper outline

この研究は最適輸送問題に焦点を当てていますが、the back-and-forth method は多くの他の問題にも応用できると予想されます。たとえば、the back-and-forth method は簡単に Wasserstein 勾配流を計算するために適応できます。これにより、大規模な重要で興味深い PDE シミュレーションが可能になります。[6、16、17、21] など、いくつかの例を挙げることができます。また、この方法は平均場ゲームと呼ばれる急速に発展する分野から生じる計算問題を解決するために有用である可能性があります [15、18]。今後の研究でこれらの応用を探究することを期待しています。

本論文の残りの部分は次のように構成されています。第 2 節では、最適輸送と勾配ベースの最適化スキームに必要な背景情報を再確認します。第 3 節では、the back-and-forth method を紹介し、その安定性と効率性の理由を示します。第 4 節では、アルゴリズムの数値実装について説明し、そのパフォーマンスを示すために様々な実験を行います。

## 2 Background

### 2.1 Optimal transport and the c-transform

$\mathbb{R}^d$  次元空間の凸でコンパクトな部分集合を  $\omega$  とする。 $\omega$  上のコストとは、連続関数  $c: \omega \times \omega \rightarrow \mathbb{R}$  のことである。最適輸送理論では、かなり一般的なコストが考慮されるが、本論文では、以下の場合に焦点を当てる。

$$c(x, y) = h(y - x),$$

ただし、 $h: \mathbb{R}^d \rightarrow \mathbb{R}$  は厳密に凸で偶関数である。 $\omega$  上に支持される 2 つの確率測度  $\mu$  と  $\nu$  が与えられた場合、最適輸送問題のモンジュ形式は、次のように定義される。

$$C(\mu, \nu) = \inf_T \int_{\Omega} c(x, T(x)) d\mu(x),$$

ただし、infimum は  $\mu$  を  $\nu$  に輸送するような写像  $T: \omega \rightarrow \omega$  に対して実行される。すなわち、 $T_{\#}\mu = \nu$  となるものである。

**最適輸送問題 (Monge の問題 (1871))** ある砂山から砂山 (測度  $\mu$ ) と同じ体積の穴 (測度  $\nu$ ) に砂を運ぶ (写像  $T$ ). 輸送にかかるコストは重さと移動距離に依存する時, コストを最小にする方法を求めよ.

写像  $T$  による押し出し測度  $T_{\#}\mu$  は、可測部分集合  $A \subset \omega$  に対して  $T_{\#}\mu(A) = \mu(T^{-1}(A))$  で定義されます。また、連続なテスト関数  $f : \Omega \rightarrow \mathbb{R}$  に対する押し出し測度の積分を定義することで、押し出し測度の特徴付けることもできます：

$$\int_{\Omega} f(y) d(T_{\#}\mu)(y) = \int_{\Omega} f(T(x)) d\mu(x).$$

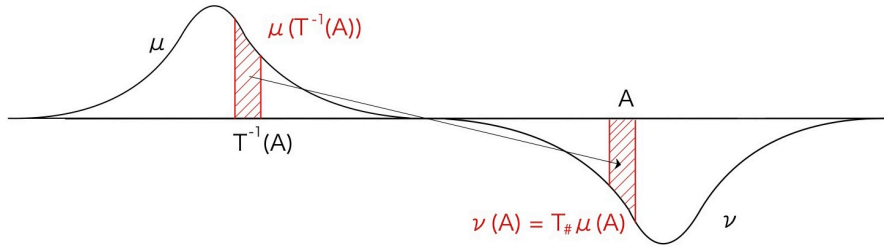


図 1: transport map

本論文では、確率測度  $\mu$  と  $\nu$  がルベグ測度に関して絶対連続である特別な場合に焦点を当てます。そのため、しばしば測度とその密度関数を同一視します。この仮定の下では、 $\mu$  を  $\nu$  に移す最適写像  $T_*$  が唯一存在し、その逆写像  $T^{-1}$  は  $\nu$  を  $\mu$  に移す最適写像です [7,12]。

さらに、最適輸送問題の Kantorovich 双対定式化を解くことで、最適写像を見つけることができます。双対形式は、pushforward 制約条件のためのラグランジュ乗数を導入することで導出することができます。pushforward 制約条件  $T_{\#}\mu = \nu$  はすべての連続関数  $\phi$  について以下が成り立つ場合にのみ成り立ちます。

$$\int_{\Omega} \phi(T(x)) d\mu(x) = \int_{\Omega} \phi(y) d\nu(y) \quad \left( \Leftrightarrow \int_{\Omega} \phi(T(x)) \mu(x) dx = \int_{\Omega} \phi(y) \nu(y) dy \right).$$

従って

$$\begin{aligned} C(\mu, \nu) &= \inf_{T_{\#}\mu=\nu} \int_{\omega} c(x, T(x)) d\mu(x) \\ &= \inf_{T_{\#}\mu=\nu} \sup_{\phi} \int_{\omega} c(x, T(x)) d\mu(x) - \phi(T(x)) d\mu(x) + \int_{\omega} \phi(y) d\nu(y). \end{aligned}$$

$\mu$  が絶対連続であるとき、infimum と supremum は入れ替え可能である。 $T$  に関する項でまとめると、

$$C(\mu, \nu) = \sup_{\phi} \inf_{T_{\#}\mu=\nu} \int_{\omega} (c(x, T(x)) - \phi(T(x))) d\mu(x) + \int_{\omega} \phi(y) d\nu(y).$$

このように、ある点  $x_0$  での  $T$  の最適な選択は、他の任意の点  $x$  での選択と完全に切り離されていることが明確になった。したがって、infimum を積分の内側に移動でき、 $\inf_{T(x)} c(x, T(x)) - \psi(T(x))$  という操作が双対問題に重要な役割を果たすことが明らかである。この操作は  $c$ -変換と呼ばれ、最適輸送の中心にあります。 $c$ -変換は、関数  $\phi$  を別の関数  $\phi^c$  にマップするものであり、凸解析のルジャンドル変換 (凸共役) の一般化と見ることができます。

**Definition 2.1.** 連続関数  $\phi : \omega \rightarrow \mathbb{R}$  に対して、その  $c$ -transform  $\phi^c : \omega \rightarrow \mathbb{R}$  を以下のように定義する。

$$\phi^c(x) = \inf_{y \in \omega} c(x, y) - \phi(y)$$

また、 $\phi$  が  $c$ -凹関数とは、 $\phi = \psi^c$  となる連続関数  $\psi : \Omega \rightarrow \mathbb{R}$  が存在することをいう。さらに関数の組  $(\phi, \psi)$  が  $c$ -共役であるとは、 $\phi = \psi^c$  かつ  $\psi = \phi^c$  のときをいう。

ここで、Kantorovich の双対汎関数を導入する準備が整いました。

$$J(\phi) = \int \phi d\nu + \int \phi^c d\mu, \quad I(\psi) = \int \psi^c d\nu + \int \psi d\mu,$$

以上の議論から、以下が成り立ちます。

$$C(\mu, \nu) = \sup_{\phi} J(\phi) = \sup_{\psi} I(\psi)$$

**Lemma 2.1.** [12, 13] 連続関数  $\phi : \Omega \rightarrow \mathbb{R}$  について、 $c$ -transform は次の特性を持ちます：

- (i) 任意の  $x \in \Omega$  について、 $\phi(x) \leq \phi^{cc}(x)$  が成り立ち、 $\phi(x) = \phi^{cc}(x)$  であるための必要十分条件は、 $\phi$  が  $c$ -concave であることである。その結果、任意の連続関数  $\phi$  について、 $\phi^{ccc}(x) = \phi^c(x)$  が成立する。
- (ii)  $\phi$  が  $c$ -concave の場合、最小化問題  $\inf_{y \in \Omega} c(x, y) - \phi(y)$  はほとんど全ての (*almost every*)  $x$  について唯一の最小値  $T_\phi(x)$  を持つ。さらに、次の明示的な式が成り立ちます：

$$T_\phi(x) = x - (\nabla h)^{-1}(\nabla \phi^c(x)),$$

ただし  $c(x, y) = h(y - x)$  である。

- (iii)  $\phi$  が  $c$ -concave で  $u$  が  $\omega$  上の連続関数である場合、ほとんど全ての  $x \in \Omega$  について、以下が成り立ちます。

$$\lim_{\epsilon \rightarrow 0} \frac{(\phi + \epsilon u)^c(x) - \phi^c(x)}{\epsilon} = -u(T_\phi(x))$$

以下は改めて  $J$  を定義している。

**Theorem 2.2.** [7, 12, 13] 二重双対問題

$$\sup_{\phi} \int_{\omega} \phi d\nu + \int_{\omega} \phi^c d\mu =: J(\phi),$$

は、以下の性質を持ちます。

- (i)  $\phi$  に関して  $J$  は concave です。
- (ii)  $\mu$  がルベーグ測度に関して絶対連続である場合、 $J$  は  $c$ -concave な関数  $\phi_*$  によって最大化され、 $T_{\phi_*}$  は  $\mu$  を  $\nu$  にプッシュする一意な最適マップです。
- (iii)  $\nu$  もまた絶対連続である場合、 $T_{\phi_*}$  はほとんどどこでも (*almost everywhere*) 逆写像を持ち、 $T^{-1}$  は  $\nu$  を  $\mu$  にプッシュする最適マップです。

定理 1 から、最適マップは凸最大化問題を解くことで計算できることがわかります。したがって、勾配上昇法などの最適化の標準的なテクニックを用いて、最適輸送問題を解くことができます。ただし、**勾配上昇法を効果的に使用するためには、適切な距離とステップサイズを選択する必要があります。** そのため、以下では Hilbert 空間における勾配上昇法を解説します。

## 2.2 Gradient ascent

凸関数に対する一定ステップサイズ勾配上昇法の初等的な結果について説明。 $(\mathcal{H}, \|\cdot\|_{\mathcal{H}})$  を可分なヒルベルト空間とし、 $F$  が滑らかな凸汎関数

$$F : \mathcal{H} \rightarrow \mathbb{R}$$

であると仮定します。

**Definition 2.2.** 点  $\phi \in \mathcal{H}$  において、有界線型写像  $\delta F_{\phi} : \mathcal{H} \rightarrow \mathbb{R}$  が  $F$  のフレシェ微分（第1変分）であるとは、以下の条件を満たすこととする。

$$\lim_{\|h\|_{\mathcal{H}} \rightarrow 0} \frac{\|F(\phi + h) - F(\phi) - \delta F_{\phi}(h)\|_{\mathcal{H}}}{\|h\|_{\mathcal{H}}} = 0.$$

**Definition 2.3.**  $\langle \cdot, \cdot \rangle_{\mathcal{H}}$  を Hilbert 空間  $\mathcal{H}$  上で定義された内積とする。写像  $\nabla_{\mathcal{H}} F : \mathcal{H} \rightarrow \mathcal{H}$  が  $F$  の  $\mathcal{H}$ -勾配であるとは、任意の  $(\phi, h) \in \mathcal{H} \times \mathcal{H}$  に対して、以下の式が成立することを言う。

$$\langle \nabla_{\mathcal{H}} F(\phi), h \rangle_{\mathcal{H}} = \delta F_{\phi}(h)$$

The back-and-forth method では、 $\dot{H}^1$ -勾配を使用します。ここで、 $\dot{H}^1(\Omega)$  は以下のように定義されます。

$$\dot{H}^1(\Omega) = \left\{ \varphi : \Omega \rightarrow \mathbb{R} : \int_{\Omega} \varphi(x) dx = 0 \quad \text{and} \quad \int_{\Omega} |\nabla \varphi|^2 dx < \infty \right\}, \quad (2)$$

$\dot{H}^1(\Omega)$  上の内積は、以下のように定義されます。

$$\langle \varphi_1, \varphi_2 \rangle_{\dot{H}^1(\Omega)} = \int_{\Omega} \nabla \varphi_1(x) \cdot \nabla \varphi_2(x) dx.$$

双対空間  $\dot{H}^{-1}$  も述べます。2つの確率密度関数  $\rho_1$  と  $\rho_2$  に対して、次のように定義します。

$$\|\rho_2 - \rho_1\|_{\dot{H}^{-1}}^2 = \int_{\Omega} |\nabla \phi(x)|^2 dx \quad (3)$$

ただし、 $\phi$  はラプラス方程式  $-\Delta \phi = \rho_2 - \rho_1$  かつ zero Neumann 境界条件を満たす  $\dot{H}^1(\omega)$  の唯一の解である。(詳細は [22, Section 5.5.2] を参照)。

以下の補題は、 $\dot{H}^1$  勾配が特に単純な形式を持つことを示している。証明は積分部分法の直接的な応用である。

**Lemma 2.3.**  $F : \dot{H}^1(\omega) \rightarrow \mathbb{R}$  が、 $\forall \phi \in \dot{H}^1(\omega)$  に対して、任意の  $h \in \dot{H}^1(\omega)$  で評価される一次変分  $\delta F_{\phi}$  が関数  $f_{\phi}$  に対する積分として表される、つまり

$$\delta F_{\phi}(h) = \int_{\Omega} h(x) f_{\phi}(x) dx$$

このとき、 $F$  の  $\dot{H}$ -勾配は次の形式を持ちます。

$$\nabla_{\dot{H}^1} F(\phi) = (-\Delta)^{-1} \bar{f}_{\phi},$$

ここで、 $(-\Delta)^{-1}$  は zero Neumann 境界条件を持つ負ラプラシアン逆演算子を表し、 $\bar{f}_{\phi} = f_{\phi} - \frac{1}{|\Omega|} \int_{\Omega} f_{\phi}$ 。

以下は、一般的な Hilbert 空間上での勾配法について説明したものです。勾配法は、Fréchet-微分可能な汎関数  $F : \mathcal{H} \rightarrow \mathbb{R}$  を最大化することを目的として、以下の反復式で表されます。

$$\phi_{n+1} = \phi_n + \sigma \nabla_{\mathcal{H}} F(\phi_n),$$

ここで、ステップサイズ  $\sigma > 0$  は定数です。最大値に収束するためには、つまり

$$F(\phi_n) \rightarrow \sup F,$$

収束するためには、勾配写像  $\nabla_{\mathcal{H}} F$  の連続性を制御する必要があります。次の勾配法の収束定理は、最適化の基本的な考え方の 1 つです。

**Theorem 2.4.** ([20] を参照)  $F : H \rightarrow \mathbb{R}$  を Frechet 微分可能な凸関数とし、ある定数  $\sigma > 0$  が存在して、

$$F(\phi) \geq F(\hat{\phi}) + \delta F_{\hat{\phi}}(\phi - \hat{\phi}) - \frac{1}{2\sigma} \|\phi - \hat{\phi}\|_{\mathcal{H}}^2$$

全ての  $\phi, \hat{\phi} \in \mathcal{H}$  について成り立つと仮定する。このとき、勾配降下法の反復

$$\phi_{n+1} = \phi_n + \sigma \nabla_{\mathcal{H}} F(\phi_n)$$

は、以下の上昇性の性質を持つ:

$$F(\phi_{n+1}) - F(\phi_n) \geq \frac{\sigma}{2} \|\nabla_{\mathcal{H}} F(\phi_n)\|_{\mathcal{H}}^2$$

さらに、 $F$  が唯一の極大点  $\phi_*$  を持ち、 $\sup_n \|\phi_n\|_{\mathcal{H}} < \infty$  である場合、列  $\{\phi_n\}_{n=0}^{\infty}$  は  $\phi_*$  に弱収束する。

定理 2 の証明については、"Appendix" を参照。これらの道具を手に入れたので、我々は交互法を導入する準備が整った。

### 3 The back-and-forth method

$\omega \subset \mathbb{R}^d$  を凸かつコンパクトな領域とし、次の形式のコスト  $c : \omega \times \omega \rightarrow \mathbb{R}$  を考える。

$$c(x, y) = h(y - x).$$

ただし、 $h : \mathbb{R}^d \rightarrow \mathbb{R}$  は厳密に凸で偶関数である。確率密度関数  $\mu$  と  $\nu$  が  $\omega$  で支持されているとき、最適輸送問題の双対形式である次式を考えることを考えます。

$$C(\mu, \nu) = \sup_{\phi} \int_{\omega} \phi(y) \nu(y) dy + \int_{\omega} \phi^c(x) \mu(x) dx, \quad (4)$$

ここで、supremum は連続関数  $\phi : \omega \rightarrow \mathbb{R}$  について実行され、c-transform は  $\phi^c(x) = \inf_{y \in \omega} (c(x, y) - \phi(y))$  で定義されます。

双対問題の最大化解を計算する効率的なアルゴリズムを開発するのが目的。双対問題の最大化解を計算するための 2 つの等価な形式で双対関数を考える。

この節での目的は、デュアル問題の最大化関数を効率的に計算するアルゴリズムを開発することです。以下では、デュアル関数を 2 つの等価な形式で考えます。

$$J(\phi) = \int_{\omega} \phi d\nu + \int_{\omega} \phi^c d\mu \quad (5)$$

$$I(\psi) = \int_{\omega} \psi^c d\nu + \int_{\omega} \psi d\mu \quad (6)$$

$J$  と  $I$  の関数は基本的に同じであることに注意してください。異なる点は、 $\mu$  と  $\nu$  の役割が逆転することです。これ以上進めるには、 $J$  と  $I$  の変分の式が必要です。

**Lemma 3.1.** [11, 12] 連続関数  $\phi : \omega \rightarrow \mathbb{R}$  および  $\psi : \omega \rightarrow \mathbb{R}$  上で定義される関数  $J$  および  $I$  を考える。 $\phi$  が  $c$ -凹関数である場合、 $J$  の第一変分は、以下のように表される。

$$\delta J_{\phi} = \nu - T_{\phi\#}\mu,$$

同様に、 $\psi$  が  $c$ -凹関数である場合、 $I$  の第一変分は以下のように表される。

$$\delta I_{\psi} = \mu - T_{\psi\#}\nu.$$

ここで、任意の  $c$ -凹関数  $\phi : \omega \rightarrow \mathbb{R}$  に対して、

$$T_{\phi}(x) = x - (\nabla h)^{-1}(\nabla \psi^c(x)).$$

になることを思い出す。

最適輸送問題を解決するためのアプローチ、すなわち、the back-and-forth method を紹介する準備が整った。アルゴリズム 1 に概要を示す。アルゴリズムは、次の 2 つの主要なアイデアに基づいている：

(i)  $\dot{H}^1$  metric の勾配上昇ステップであり、

$$\begin{aligned} \nabla_{\dot{H}^1} J(\phi) &= (-\Delta)^{-1}(\nu - T_{\phi\#}\mu), \\ \nabla_{\dot{H}^1} I(\psi) &= (-\Delta)^{-1}(\mu - T_{\psi\#}\nu). \end{aligned} \quad (7)$$

(ii) back-and-forth update scheme は  $J$  と  $I$  の勾配上昇ステップを交互に行うものである。

ステップサイズ  $\sigma$  に関する情報については、セクション 4.2 を参照してください。

### 3.1 $\dot{H}^1$ -gradient ascent

The back-and-forth method の主要なステップは、 $\dot{H}^1$  メトリックにおける勾配上昇ステップであり、

$$\begin{aligned} \phi_{n+\frac{1}{2}} &= \phi_n + \sigma \nabla_{\dot{H}^1} J(\phi_n), \\ \psi_{n+1} &= \psi_{n+\frac{1}{2}} + \sigma \nabla_{\dot{H}^1} I(\psi_{n+\frac{1}{2}}). \end{aligned}$$

The back-and-forth method の収束性を得るためには、これらのステップがそれぞれ双対関数  $J$  および  $I$  の値を増加させるかどうかを知る必要があります。以下では、対称性により、 $J$  について成立する性質は  $I$  についても成立するため、 $J$  の更新ステップに焦点を当てます。

前のセクションの定理 2.4 を思い出すと、一般的な Hilbert 空間  $\mathcal{H}$  において、 $\mathcal{H}$ -勾配上昇ステップ

$$\phi_{n+1} = \phi_n + \sigma \nabla_{\mathcal{H}} J(\phi_n)$$



は、不等式

$$J(\phi) \geq J(\hat{\phi}) + \delta J_{\hat{\phi}}(\phi - \hat{\phi}) - \frac{1}{2\sigma} \|\phi - \hat{\phi}\|_{\mathcal{H}}^2 \quad (8)$$

が  $\sigma > 0$  と任意の  $\phi, \hat{\phi} \in H$  について成立するならば、勾配上昇ステップは**上昇性**を持ちます。すなわち、

$$J(\phi_{n+1}) \geq J(\phi_n) + \frac{\sigma}{2} \|\nabla_{\mathcal{H}} J(\phi_n)\|_{\mathcal{H}}^2$$

が成立します。注意すべきは、この不等式は、 $\mathcal{H}$  に関連するノルムが強くなると（つまり、右辺がより負になると）、満たしやすくなることです。

不等式を確立するために、次の式の下限を求める必要があります。

$$J(\phi) - J(\hat{\phi}) - \delta J_{\hat{\phi}}(\phi - \hat{\phi})$$

この式は、 $\hat{\phi}$  周りで 1 次の Taylor 展開で  $J$  を近似することによる誤差と認識できます。

進展するための最も簡単な方法は、凸不等式  $J(\hat{\phi}) \leq J(\phi) + \delta J_{\phi}(\hat{\phi} - \phi)$  を使って、以下の下限を得ることです。

$$J(\phi) - J(\hat{\phi}) - \delta J_{\hat{\phi}}(\phi - \hat{\phi}) \geq (\delta J_{\phi} - \delta J_{\hat{\phi}})(\phi - \hat{\phi})$$

Lemma 3 の一次変分の明示的な式を使って、以下のように書くことができます。

$$(\delta J_{\phi} - \delta J_{\hat{\phi}})(\phi - \hat{\phi}) = \int_{\omega} (\phi - \hat{\phi})(T_{\hat{\phi}\#}\mu - T_{\phi\#}\mu)$$

ここで、不等式 (8) は以下の式を示すことができれば成立します。

$$\int_{\Omega} (\phi - \hat{\phi})(T_{\hat{\phi}\#}\mu - T_{\phi\#}\mu) \geq -\frac{1}{2\sigma} \|\phi - \hat{\phi}\|_{\mathcal{H}}^2$$

*Proof.* 不等式 (8) を示すためには、以下の式変形が必要です。

$$\begin{aligned} \int_{\Omega} (\phi - \hat{\phi})(T_{\hat{\phi}\#}\mu - T_{\phi\#}\mu) &= \int_{\Omega} T_{\phi\#}(\phi - \hat{\phi})d\mu - \int_{\Omega} T_{\hat{\phi}\#}(\phi - \hat{\phi})d\mu \\ &= \langle T_{\phi\#}(\phi - \hat{\phi}), \phi - \hat{\phi} \rangle_{\mathcal{H}} - \langle T_{\hat{\phi}\#}(\phi - \hat{\phi}), \phi - \hat{\phi} \rangle_{\mathcal{H}} \\ &= \langle T_{\phi\#}(\phi - \hat{\phi}) - T_{\hat{\phi}\#}(\phi - \hat{\phi}), \phi - \hat{\phi} \rangle_{\mathcal{H}} \\ &= \langle T_{\phi\#}^*(T_{\hat{\phi}\#}\mu - T_{\phi\#}\mu), \phi - \hat{\phi} \rangle_{\mathcal{H}} \\ &= \langle (T_{\hat{\phi}\#} - I)^*(T_{\hat{\phi}\#}\mu - T_{\phi\#}\mu), \phi - \hat{\phi} \rangle_{\mathcal{H}} \\ &= \langle (T_{\hat{\phi}\#} - I)(T_{\hat{\phi}\#}\mu - T_{\phi\#}\mu), \phi - \hat{\phi} \rangle_{\mathcal{H}} \\ &\geq -\|(T_{\hat{\phi}\#} - I)(T_{\hat{\phi}\#}\mu - T_{\phi\#}\mu)\|_{\mathcal{H}} \|\phi - \hat{\phi}\|_{\mathcal{H}} \\ &= -\|T_{\hat{\phi}\#}\mu - T_{\phi\#}\mu\|_{\mathcal{H}} \|\phi - \hat{\phi}\|_{\mathcal{H}} \\ &= -\frac{1}{\sigma} \|\phi - \hat{\phi}\|_{\mathcal{H}}^2, \end{aligned}$$

ここで、2 行目から 3 行目への変形で、ヒルベルト空間  $\mathcal{H}$  上で内積をとりました。また、3 行目から 4 行目への変形では、 $T_{\hat{\phi}\#} - I$  が  $\mathcal{H}$  上の自己共役な線形作用素であることを用いました。最後の行では、前節での定義に従って、 $\sigma = \|T_{\hat{\phi}\#}\mu - T_{\phi\#}\mu\|_{\mathcal{H}}$  とおきました。

したがって、不等式 (8) は以下のように得られます。

$$\int_{\Omega} (\phi - \hat{\phi})(T_{\hat{\phi}\#}\mu - T_{\phi\#}\mu) \geq -\frac{1}{\sigma}\|\phi - \hat{\phi}\|_{\mathcal{H}}^2 \geq -\frac{1}{2\sigma}\|\phi - \hat{\phi}\|_{\mathcal{H}}^2,$$

最後の不等号は  $\sigma \leq \frac{1}{2}$  で成立するためです。従って、 $\sigma \leq \frac{1}{2}$  の場合には、不等式 (8) が成立することが示されました。

□

私たちの目標は、この不等式が成り立つようなヒルベルト空間  $\mathcal{H}$  とパラメータ  $\sigma$  を見つけることです。

任意のヒルベルト空間  $H$  に対して、以下のコーシー・シュワルツの不等式が成立します。

$$\int_{\Omega} (\phi - \hat{\phi})(T_{\hat{\phi}\#}\mu - T_{\phi\#}\mu) \geq -\|\phi - \hat{\phi}\|_{\mathcal{H}}\|T_{\hat{\phi}\#}\mu - T_{\phi\#}\mu\|_{\mathcal{H}^*},$$

ここで、 $\mathcal{H}^*$  は  $L^2$  内積に関する  $\mathcal{H}$  の双対空間です (すなわち、 $\mathcal{H}$  を  $L^2(\omega)$  といういわゆる "pivot 空間" に関して双対化しています)。

適切なヒルベルト空間  $\mathcal{H}$  を選ぶことで、

$$\|T_{\hat{\phi}\#}\mu - T_{\phi\#}\mu\|_{\mathcal{H}^*} \geq \frac{1}{2\sigma}\|\phi - \hat{\phi}\|_{\mathcal{H}}, \quad (9)$$

式 (9) が成立することを示せば、コーシー・シュワルツの不等式と組み合わせることで式 (8) が得られます (注意: より正確な議論により、式 (9) の右辺にある  $\frac{1}{2}$  の因子は除去できますが、ここでは気にしません)。再度強調しますが、 $\mathcal{H}$  に関連付けられたノルムが強くなるにつれて、不等式 (9) はより簡単に満たされるようになります。実際、 $\mathcal{H}$  が強くなると  $\mathcal{H}^*$  は弱くなり、左辺が小さくなる一方で、右辺が大きくなります。

不等式 (9) は、関数  $\phi \mapsto T_{\phi\#}$  が  $\mathcal{H} \rightarrow \mathcal{H}$  への Lipschitz 連続写像である必要があることを示しています。したがって、上昇特性を持つ勾配上昇スキームを得るために、 $\phi$  と  $\hat{\phi}$  が  $\mathcal{H}$  で近ければ、 $T_{\phi\#}\mu$  と  $T_{\hat{\phi}\#}\mu$  が  $\mathcal{H}^*$  で近くなる Hilbert 空間  $\mathcal{H}$  を選ぶ必要があります。これは容易な課題ではありません。たとえば、 $\phi$  が  $c$ -凸でない場合、写像  $\phi \mapsto T_{\phi}$  が well-defined ではない場合があります、必ずしも特定の連続性を持たない場合があります。

上記の複雑さに直接対処しようとする、どのように進めればいいのかは明確ではありません。進展をするために、二次コストかつある well-defined な  $c$ -凹関数に制限したとき、Hilbert 空間  $\dot{H}^1(\Omega) = \{\phi : \Omega \rightarrow \mathbb{R} \mid \int_{\Omega} \phi = 0 \text{ かつ } \|\nabla \phi\|_{L^2} < \infty\}$  が勾配上昇を保持できることを示します。

結果を述べる前に、 $\dot{H}^1$  の選択肢はおそらく弱めることができません。実際、 $T\phi$  の式は  $\nabla \phi^c$  に依存するため、 $\phi$  と  $\hat{\phi}$  の勾配を何らかの制御を持つ必要があり、(8) または (9) を示す可能性があります。

**Proposition 3.2.** 2 次コスト  $h(y - x) = \frac{1}{2}|x - y|^2$  を持つ Kantorovich 双対問題  $\sup_{\phi} J(\phi)$  を考える。 $\lambda > 0$  が存在して、

$$\phi_n, \phi_{n+\frac{1}{2}} \in S_{\lambda} := \{\phi : (1 - \lambda^{-1})I \leq D^2\phi(x) \leq (1 - \lambda)I\}$$

が任意の  $n$  に対して成り立つとき、ステップサイズ  $\sigma = \|\mu\|^{-1}\lambda^{-d+1}$  を持つ勾配上昇ステップ

$$\phi_{n+\frac{1}{2}} = \phi_n + \sigma \nabla_{\dot{H}^1} J(\phi_n)$$

は、以下の上昇性を満たします。

$$J(\phi_{n+\frac{1}{2}}) - J(\phi_n) \geq \frac{1}{2} \|\mu\|^{-1} \lambda^{d+1} \|\nabla_{\dot{H}^1} J(\phi_n)\|_{\dot{H}^1}^2.$$

さらに、質量密度はターゲット  $\nu$  に近づきます。

$$\|\rho_{n+\frac{1}{2}} - \nu\|_{\dot{H}^{-1}}^2 - \|\rho_n - \nu\|_{\dot{H}^{-1}}^2 \leq -\|\rho_{n+\frac{1}{2}} - \rho_n\|_{\dot{H}^{-1}}^2,$$

ここで、 $\rho_n = T_{\phi_n \# \mu}$  および  $\rho_{n+\frac{1}{2}} = T_{\phi_{n+\frac{1}{2}} \# \mu}$  を表し、 $\dot{H}^{-1}$  距離は (3) で定義されます。

証明は Appendix を参照。

**Remark 3.1.** 命題 3.2 はすべての反復子  $\phi_n$  が上下ヘシアン境界を持つと仮定すると、 $\dot{H}^1$  上で下降特性が成立することを示しています。残念ながら、単純な上昇スキーム  $\phi_{n+1} = \phi_n + \sigma \nabla_{\dot{H}^1} J(\phi_n)$  を通じてこの性質が維持されるとは期待できません。加えて、勾配の更新によって、いずれはイテレーションが望ましい領域の外に押し出される可能性があります。実際、勾配更新ステップによって  $c$ -凹性が維持されることすら保証できません。

**Remark 3.2.** 前述の問題を回避する方法の 1 つは、反復解を集合

$$S_{\lambda} = \{\phi \mid (1 - \lambda^{-1})I \leq D^2\phi(x) \leq (1 - \lambda)I\}$$

に射影するステップを追加することです。 $S_{\lambda}$  は凸集合であるので、射影ステップは上昇特性に干渉しません。残念ながら、 $E_{\lambda}$  への射影を計算するのはコストがかかります（半定値計画問題を解く必要があるため）。さらに、双対問題の最大化は、任意の  $\lambda > 0$  に対して  $S_{\lambda}$  内に存在するわけではありません（ただし、二次コストの場合は  $c$ -凹性により  $S_0$  内に存在する必要があります）。このような場合、任意の正確な解を得るために  $\lambda \rightarrow 0$  として、 $\lambda$  を 0 に近づける必要があります。

上記の Remark で言及された問題のいくつかを回避するために、我々は従来の勾配上昇法の代わりに、 $J$  を最大化するステップと  $I$  を最大化するステップを交互に繰り返す前後法を用いることにします。

### 3.2 Back-and-forth updates

Kantorovich の双対問題は、以下の形式で書くことができます。

$$C(\mu, \nu) = \sup_{\phi, \psi} \left\{ \int_{\omega} \phi(y) \nu(y) d\nu(y) + \int_{\omega} \psi(x) \mu(x) d\mu(x) \right\}, \quad (10)$$

ここで、上限は、すべての  $x, y \in \omega$  に対して

$$\phi(y) + \psi(x) \leq h(y - x)$$

を満たす連続関数  $\phi$  と  $\psi$  の上で取られます。この式は、ポテンシャル  $\phi$  と  $\psi$  が対称的な役割を果たすことを強調しています。以下のように、双対汎関数を次の形式で表すことが便利です。

$$D(\phi, \psi) = \int_{\omega} \phi \nu dy + \int_{\omega} \psi \mu dx - \iota_C(\phi, \psi) \quad (11)$$

ここで、制約条件は凸指示関数  $\iota_C$  によって符号化されます。凸集合

$$C = \{(\phi, \psi) \mid \forall x, y \in \omega, \phi(y) + \psi(x) \leq h(y - x)\}$$

上では、 $\iota_C$  は 0 をとり、他のペア  $(\phi, \psi)$  に対しては  $+\infty$  をとります。

Kantorovich 問題 (10) は、双対汎関数  $D(11)$  を用いて、 $C(\mu, \nu) = \sup_{\phi, \psi} D(\phi, \psi)$  と簡潔に書くことができます。

また、対称的な表現 (11) から、 $\phi$  または  $\psi$  を排除することで、関数  $J$  と  $I$  を求めることができます。具体的には、

$$\phi^c = \arg \max_{\psi} D(\phi, \psi), \quad \psi^c = \arg \max_{\phi} D(\phi, \psi)$$

を見つけることで、次式を得る。

$$J(\phi) = D(\phi, \phi^c) = \sup_{\psi} D(\phi, \psi), \quad I(\psi) = D(\psi^c, \psi) = \sup_{\phi} D(\phi, \psi).$$

vanilla 勾配上昇法  $\phi_{n+1} = \phi_n + \sigma \nabla_{\dot{H}^1} J(\phi_n)$  は、(任意の)  $\phi$ -空間にフォーカスし、関数  $J$  の勾配ステップを反復します。また、代わりに  $\psi$ -空間で作業し、関数  $I$  上の勾配上昇スキームを書くこともできます。おそらくより良いアイデアは、 $\phi$ -空間と  $\psi$ -空間を交互に行き来することです。これは、アルゴリズム 1 で示されている back-and-forth 方法の中心的なアイデアです。現在の反復  $(\phi_n, \psi_n)$  は以下のように更新されます：

$$\begin{aligned} \phi_{n+\frac{1}{2}} &= \phi_n + \sigma \nabla_{\dot{H}^1} J(\phi_n), \\ \psi_{n+\frac{1}{2}} &= (\phi_{n+\frac{1}{2}})^c, \\ \psi_{n+1} &= \psi_{n+\frac{1}{2}} + \sigma \nabla_{\dot{H}^1} I(\psi_{n+\frac{1}{2}}), \\ \phi_{n+1} &= (\psi_{n+1})^c. \end{aligned}$$

## 4 Numerical implementation and results

### 4.1 The fast c-transform

### 4.2 Step sizes

### 4.3 Experiments

## 5 Details

ルベーク測度に関して絶対連続 確率測度  $\mu$  がルベーク積分に関して絶対連続であるとき、ルベーク積分 = リーマン積分になる。すなわち、 $d\mu(x) = \mu(x)dx$  であり、

$$\int_{\Omega} \phi(T(x)) d\mu(x) = \int_{\Omega} \phi(y) d\nu(y) \Leftrightarrow \int_{\Omega} \phi(T(x)) \mu(x) dx = \int_{\Omega} \phi(y) \nu(y) dy.$$

支持されている (supported)

$$\mu \text{ supported in } \omega \iff \text{supp } \mu = \overline{\{x \in \omega \mid \mu(x) \neq 0\}}$$