arXiv:2305.02758v1 [math.PR] 4 May 2023

线性分布敏感性检验对偶的极小极大方法

Gusti van Zyl 比勒陀利亚大学数学与应用数学系

2023年5月5日

关键词 — 对偶；优化、极小极大；敏感性检验；分布鲁棒性 s

摘要

我们考虑求取 $\mathbb{E}_\nu[f(X)]$ 的最大值这一问题，其中允许 ν 在波兰空间 S 上所有满足 $d_c(\mu,\nu) \leq \theta$ 的概率测度中变化，其中 d_c 是最优传输距离， f 是定义在 S 上并满足某些正则性条件的实值函数， μ 是一个“基准”测度，而 $\theta \geq 0$。尽管这一优化问题的对偶形式的某些推导依赖于 Fenchel对偶，但我们对 S 施加紧性条件，以便改用 K. Fan极小极大定理，该定理不要求具有向量空间结构。这样便可以避免使用测度向量空间，或除拉格朗日乘子之外的其他对偶变量。

1 引言

设 \((S,d)\) 为一个波兰度量空间（即完备且可分的度量空间），且 \(\mu \) 为一个测度，称为基准测度或分布。考虑以下优化问题。

$$\text{maximize} \quad \int_S f(x)\, d\nu(x) \tag{1}$$

$$\text{subject to} \quad \nu \in \mathcal{B}_r(\mu), \tag{2}$$

其中， $\mathcal{B}_r(\mu)$（称为不确定性集合）是一个测度集合，其运输成本为 $\leq r$ ，相对于 μ而言。运输成本由成本函数 $c : S \times S \to \mathbb{R}_+$导出。假设 f 具有足够的正则性，使得最大值存在，并将该最大值记为 v_P。对偶性结果的不同版本

$$v_P = v_D := \inf_{\lambda \geq 0} \left\{ \lambda r + \mathbb{E}_\mu[\sup_{y \in S}\{f(y) - \lambda c(X,y)\}] \right\} \tag{3}$$

在不同假设下，Esfahani 和 Kuhn [9]、Gao 和 Kleywegt [8]、Blanchet 和 Murthy [3]、Bartl、Drapeau 和 Tangpi [1]，以及 Feng 和 Schl¨ogl [7]等人都得到了此类结果。这种对偶性将可行集从无限维问题缩减为有限维问题。如果（如常见情形那样）公式 (3) 中的内层上确界在解析上可处理，那么问题便转化为一个一维优化问题，而该问题本身也可能有解析解；上述参考文献中的若干示例便说明了这一点。仅举两个例子：利用此类对偶性结果，研究者计算了对风险中性分布发生一定幅度变化具有稳健性的期权价格 [2] ，以及稳健破产概率 [3] 。

Esfahani 和 Kuhn [9] 假设基准测度 μ 是一个经验分布，仅考虑 $p = 1$ 的 Wasserstein 距离，并假设 f 具有一种特殊结构，即由有限个最大值构成。

凸函数。Gao 和 Kleywegt 针对 $\inf_{y\in S}\{\lambda c(x,y) - f(y)\}$ 的极小值点集，使用了相关的 lim inf 和 lim sup 不等式，适用于 $c(x,y) = d(x,y)^p$ 且满足 $p \geq 1$ 的情形。他们的对偶性证明主要基于若干结果，其中包括与 Moreau-Yosida正则化类似的结果。Feng 和 Schl¨ogl 概述了一种使用 Karush-Kuhn-Tucker论证的证明方法。Blanchet 和 Murthy 采用一般成本函数 $c(x,y)$，并以 Fenchel对偶为其结果提供理论基础。他们的对偶变量为拉格朗日参数 λ 和可测函数 ϕ。为了应用 Fenchel对偶，他们考虑了定义在 $S\times S$ 上的有界连续函数向量空间，以及定义在 $S\times S$ 上、配备全变差范数的有限符号博雷尔测度向量空间。

本文中，我们使用 K. Fan 极小极大定理而非 Fenchel对偶来证明 (3)。这样做的一个优点是，向量空间结构被较弱的凸凹性要求所取代。因此，极小极大定理使我们无需引入向量空间结构、符号测度或不同的拓扑即可继续进行。我们仅使用拉格朗日参数 λ 作为对偶变量。为此付出的代价是，我们假设 S 是紧致的。尽管存在这一限制，我们希望这种相对简单的证明方式能够促进进一步的推广，同时也有助于使分布敏感性检验这一主题更容易被已经熟悉极小极大论证的研究人员所理解。类似于 [3] 中从紧空间推广到一般情形的步骤，这一论证也有可能扩展到非紧空间。

关于该问题的解释、最优传输距离与 Kulback-Leibler散度的比较，以及各种应用，请参阅上述文献。本文仅限于讨论方程 (3) 的极小极大证明。

本文的结构如下：我们定义记号，并回顾用运输计划表述问题的形式；推导拓扑预备知识，以便将极小极大方法应用于拉格朗日函数；然后应用极小极大方法。

2 用运输计划表述

为方便记号起见，我们考虑将（1）中的最大化替换为最小化后得到的问题。由于对于 f 的最大化问题可以通过求解对于 $-f$ 的最小化问题来解决，因此这样做并无损失。

为了定义将作为优化问题可行集的测度球，我们回顾一些关于最优传输的事实。设 $\mathcal{X}$ 和 $\mathcal{Y}$ 为波兰空间。若 μ 是 $\mathcal{X}$ 上的任意博雷尔测度，且 $T\colon \mathcal{X}\to\mathcal{Y}$ 是一个博雷尔映射，则 $T_{\#}\mu$ 表示由 $(T_{\#}\mu)(A) = \mu(T^{-1}(A))$ 定义的像测度，其中 $A\subseteq\mathcal{Y}$ 为博雷尔集。若 γ 是 $\mathcal{X}\times\mathcal{Y}$ 上的概率测度，则其边缘测度，或到 $\mathcal{X}$ 的投影，是测度 $(\mathrm{proj}_1)_{\#}\gamma$，其中 proj_1 是坐标投影 $\mathcal{X}\times\mathcal{Y}\to\mathcal{X}\colon (x,y)\mapsto x.$。等价地，对于每个博雷尔集 $A\subseteq\mathcal{X}.$，有 $((\mathrm{proj}_1)_{\#}\gamma)(A) = \gamma(A\times S)$。到 $\mathcal{Y}$ 的边缘测度，即 $(\mathrm{proj}_2)_{\#}\gamma$，定义类似。回忆 [10, 定义 1。1]：在 $\mathcal{X}$ 上的测度与在 $\mathcal{Y}$ 上的测度 ν 之间的传输计划，或耦合，是 $\mathcal{X}\times\mathcal{Y}$ 上满足 $(\mathrm{proj}_1)_{\#}\pi = \mu$ 且 $(\mathrm{proj}_2)_{\#}\pi = \nu.$ 的测度 π。我们用 $P(\mathcal{X})$ 表示波兰空间 $\mathcal{X}$ 上所有博雷尔概率测度的集合，并赋予其概率测度的弱收敛拓扑。若 $\mathcal{X}$ 是波兰空间，则 $P(\mathcal{X})$ 也是波兰空间。特别地，$P(\mathcal{X})$ 中测度的弱收敛是可度量化的。若 $\mathcal{P}$ 和 $\mathcal{Q}$ 是满足 $\mathcal{P}\subseteq P(\mathcal{X})$ 和 $\mathcal{Q}\subseteq P(\mathcal{Y})$ 的测度集合，则从任意 $\mu\in\mathcal{P}$ 到任意 $\nu\in\mathcal{Q}$ 的所有传输计划的集合记为 $\Pi(\mathcal{P},\mathcal{Q})$。成本函数是任意下半连续（l.s.c.）$c\colon \mathcal{X}\times\mathcal{Y}\to\mathbb{R}_{+\infty}$。一个典型选择是 $c(x,y) = d(x,y)^p$，其中 $p\geq 1$。测度 μ,ν 在 $\mathcal{X}$ 上的最优传输成本定义为 $d_c(\mu,\nu) := \inf\{\int_{\mathcal{X}\times\mathcal{X}} c(x,y)d\pi(x,y)\colon \pi\in\Pi(\{\mu\},\{\nu\})\}$.

设 (S,d) 为波兰空间，且 $f\colon S\to\mathbb{R}$ 下半连续。对于任意 $\mu\in P(S)$ 和 $r>0$，我们定义

$$\mathcal{B}_r(\mu) := \{\nu\in P(S)\colon d_c(\mu,\nu)\leq r\}.$$

我们考虑以下优化问题

$$\text{minimize } \int_S f(x)\,d\nu(x) \tag{4a}$$

$$\text{subject to } \nu\in\mathcal{B}_r(\mu). \tag{4b}$$

（使用“最小化”一词并不意味着最小值能够取到。）传输计划集合 $X := \Pi(\{\mu\}, \mathcal{B}_r(\mu))$ 继承了来自 $P(S \times S)$ 上概率测度的弱拓扑。

将条件 $\nu \in \mathcal{B}_r(\mu)$ 转换为关于传输计划的条件后，类似于 [7, 第 3 节1], 中的论证，可以得到问题（4）等价于以下关于运输计划的问题，其中二者的下确界相同：

$$\text{minimize} \int_{S\times S} f(y)\, d\pi(x,y), \tag{5a}$$

$$\text{subject to } \pi \in X \text{ and} \tag{5b}$$

$$\int_{S\times S} c(x,y)\, d\pi(x,y) \le r. \tag{5c}$$

我们将其称为原始问题，并将相应的下确界值记为 v_P。

3 Comp S 紧致时运输计划的紧致性 t

现在假设 S 也紧致。那么，根据 Prokhorov定理或[10, 注 6.19], $P(S)$ 是紧致的。由于 $\mathcal{B}_r(\mu)$ 是 S 的闭子集，因此它也紧致。

我们的主要工具 将是 K. Fan极小极大定理 [6] ，其表述见 Borwein和Zhuang [4] .

定义 1. *[4]* 设 X 和 Y 为集合，不一定具有向量空间结构。若函数 $K : X \times Y \to \mathbb{R}$ 在 $X \times Y$上满足类凸凹性质，则称其为类凸凹函数；如果对所有 $t, 0 \le t \le 1$，都有

(a) 对所有 $x_1, x_2 \in X$ ，存在 $x_3 \in X$ ，使得对所有 $y \in Y$

$$K(x_3, y) \le tK(x_1, y) + (1-t)K(x_2, y);\ \textit{and}$$

(b) 对所有 $y_1, y_2 \in Y$，存在 $y_3 \in Y$，使得对所有 $x \in X$

$$K(x, y_3) \ge tK(x, y_1) + (1-t)K(x, y_2).$$

定理 1. *[4,* 定理 A*]*假设 X 和 Y 是非空集合，且 K 在$X \times Y$上是类凸凹的。假设 X 是紧致的，并且对于 Y中的每个 y， $K(\cdot, y)$ 在 X 上是下半连续的。则

$$\min_{x\in X} \sup_{y\in Y} K(x,y) = \sup_{y\in Y} \min_{x\in X} K(x,y).$$

在我们的应用中， $X = \Pi(\{\mu\}, \mathcal{B}_r(\mu))$ 是传输计划（有时称为转移计划）的集合；根据下述定理，该集合在 $P(S \times S)$ 中是紧致的。

引理 1. *[10,* 推论 5.*21]* 设 $\mathcal{X}$ 和 $\mathcal{Y}$ 是波兰空间，且 $c(x, y)$ 是实值连续成本函数，$\inf c > -\infty$。设 $\mathcal{K}$ 和 $\mathcal{L}$ 分别是 $P(\mathcal{X})$ 和 $P(\mathcal{Y})$ 的两个紧致集合。那么，其边缘分布分别属于 $\mathcal{K}$ 和 $\mathcal{L}$ 的最优转移计划集合 π 本身在$P(\mathcal{X} \times \mathcal{Y})$ 中是紧致的。

我们将需要以下内容。下面的引理是 [10, 引理 4.3] 的一个变体，适用于我们的目的。

引理 2. 设 $\mathcal{X}$ 和 $\mathcal{Y}$ 为波兰空间。(1) 如果 g 是 $\mathcal{X}$ 上的非负下半连续实值函数，那么映射 $P(\mathcal{X}) \to \mathbb{R}$: $\nu \mapsto \int_{\mathcal{X}} g(x)\ d\nu(x)$ 是下半连续的。

(2) 如果 g 是 $\mathcal{X} \times \mathcal{Y}$ 上的非负下半连续实值函数，那么映射 $P(\mathcal{X} \times \mathcal{Y}) \to \mathbb{R}$: $\gamma \mapsto \int_{\mathcal{X}\times\mathcal{Y}} g(x, y)\ d\gamma(x, y)$ 是下半连续的。

证明。（1）设 $\nu_k \to \nu$ 弱收敛。由于 g 下半连续且非负，我们可以利用贝尔定理得到一个连续实值函数序列 $(g_n)_{n\in\mathbb{N}}$，使得 $0 \le g_n \uparrow g$。如有必要，用 $\min\{g_n, n\}$ 替换 g_n 后，我们可以假设 g_n 有界。根据单调收敛定理，

$$\int_{\mathcal{X}} g\ d\nu = \lim_{n\to\infty} \int_{\mathcal{X}} g_n\ d\nu = \lim_{n\to\infty} \lim_{k\to\infty} \int_{\mathcal{X}} g_n\ d\nu_k \le \liminf_{k\to\infty} \int_{\mathcal{X}} g\ d\nu_k. \tag{6}$$

(2) 的证明类似。 □

4 极小极大方法的应用

定义拉格朗日函数

$$L(\pi,\lambda) := \int_{S\times S} f(x,y)\,d\pi(x,y) + \lambda\left(\int_{S\times S} c(x,y)\,d\pi(x,y) - r\right) \tag{7}$$

$$= \int_{S\times S} f(x,y) + \lambda c(x,y)\,d\pi(x,y) - \lambda r. \tag{8}$$

与拉格朗日函数方法中的常见情形一样，如果不满足约束条件 (5c)，则项 $\sup_{\lambda\geq 0}\lambda\left(\int_{S\times S} c(x,y)\,d\pi(x,y) - r\right)$ 为无穷大；如果满足，则为 0。由此可得 $v_P = \min_{\pi\in X}\sup_{\lambda\geq 0} L(\pi,\lambda)$。

定理 2. 对于上述定义的 X 、 $Y=\mathbb{R}_+$ ，以及选为方程 (7) 的拉格朗日函数 L 的函数 K ，极小极大定理 1 的条件均得到满足。具体而言：

1. X 在弱拓扑下是紧 Hausdorff 的

2. 对每个 $\lambda\in Y$， $L\colon X\times Y\to\mathbb{R}$ 在 X 上下半连续。

3. L 是类凸凹的。

证明。（1） X 的弱紧性由引理 1 得出。由于 $P(S\times S)$ 上的弱拓扑可度量化， $P(S\times S)$，因此 X 是豪斯多夫的。

（2）固定 $\lambda\geq 0$.由方程（7）和引理 2 可知，映射 $x\mapsto L(x,\lambda)$ 是下半连续的 .c.

（3）尽管 X 和 Y 不是向量空间，且 L 不是线性的，但 L 确实保持凸组合：若 π_1 和 π_2 是分别属于 X 和 $0\leq t\leq 1$ 的输运测度，那么显然，对于任意 $\lambda\in Y$，

$$L(t\pi_1+(1-t)\pi_2,\lambda) = tL(\pi_1,\lambda) + (1-t)L(\pi_2,\lambda). \tag{9}$$

因此， $\pi_3 := t\pi_1+(1-t)\pi_2$ 满足定义 1 第一部分中的等式。定义 1 第二部分中的等式证明类似。

□

这使我们能够推导出对偶。关于测度的优化子问题由一个集中于 $\{(x,y)$ 的测度求解这一事实 $\in S\times S\colon\ y\in\arg\min_{z\in S}\{f(z)+\lambda c(x,z)\}\}$ 是显然的；只需做一些显而易见的调整，以便在最小化问题和最大化问题之间转换，其中包括对 [3]和 [7]所作的调整。

推论1。 $v_P = v_D$，其中

$$v_D := \sup_{\lambda\geq 0}\left\{\int_S \psi_\lambda(x)\,d\mu(x) - \lambda r\right\}, \tag{10}$$

且 $\phi_\lambda(x)\colon = \text{最小化}_{y\in S}\{f(y)+\lambda c(x,y)\}.$

证明。根据定理 2，我们有

$$\min_{\pi\in X}\sup_{\lambda\geq 0} L(\pi,\lambda) = \sup_{\lambda\geq 0}\min_{\pi\in X} L(\pi,\lambda). \tag{11}$$

我们已经注意到， v_P 等于方程（11）的左端。对于右端，固定 $\lambda\geq 0$ 并考虑内层最小化问题

$$\min_{\pi\in X} L(\pi,\lambda) = \min_{\pi\in X}\left\{\int_{S\times S} f(x,y)+\lambda c(x,y)\,d\pi(x,y)\right\} - \lambda r. \tag{12}$$

由于 X 是紧致的，且对于每个 $\lambda\geq 0$， L 在 X 上是下半连续的，因此 L 在某个点（记为 $\pi^\star\in X$）处取得最小值。事实上，我们可以构造这个极小化元。由于 S是紧致的，且映射 $y\mapsto f(y)+\lambda c(x,y)$ 是下半连续的，因此对于每个$x\in S$，被积函数的极小化元集合 $m(x) := \arg\min_{y\in S}\{f(y)+\lambda c(x,y)\}$ 非空。令 $T\colon\ S\to S$为可测映射，并将 x 映射到 $m(x)$ 中的任意元素。此类映射的存在性可由一个可测

选择定理推出，相关细节见附录 A。这使我们能够定义确定性[10, 定义 1.2] 耦合 $\pi^* := (Id, T)\#\mu$· 根据该定义， π^* 的支集是 $\{(x, T(x)) : \ x \in S\} \subseteq S \times S$. 的子集

考虑任意一个传输计划 $\pi \in X$。结合上述事实以及传输计划 π^* 和 π的边缘化性质，可得

$$\begin{aligned}
&\int_{S\times S} f(y) + \lambda c(x,y)\, d\pi^*(x,y) \\
&= \int_{S\times S} \min_{y\in S}\{f(y) + \lambda c(x,y)\}\, d\pi^*(x,y) \\
&= \int_{S} \min_{y\in S}\{f(y) + \lambda c(x,y)\}\, d\mu(x) \\
&= \int_{S\times S} \min_{y\in S}\{f(y) + \lambda c(x,y)\}\, d\pi(x,y) \\
&\leq \int_{S\times S} f(y) + \lambda c(x,y)\, d\pi(x,y).
\end{aligned}$$

由于 π 是任意的， π^* 是一个极小化元，并且

$$\min_{\pi\in X} L(\pi,\lambda) = L(\pi^*,\lambda) = \int_S \phi_\lambda(x)\ d\mu(x) - \lambda r.$$

结合方程（11），可得方程（10）。 □

注 1. 将结果推广到非紧致 S时，需要类似于 *[3]*或 *[8]*的增长条件。

参考文献

[1] Daniel Bartl、Samuel Drapeau 和 Ludovic Tangpi。鲁棒优化确定性等价物与期权定价的计算方面。Mathematical Finance，2019：1–23。[2] Daniel Bartl、Samuel Drapeau 和 Ludovic Tangpi。鲁棒优化确定性等价物与期权定价的计算方面。Math. Finance，30(1)：287–309，2020。[3] Jose Blanchet 和 Karthyek Murthy。通过最优传输量化分布模型风险。Math. Oper. Res.，44(2)：565–600，2019。[4] J. M. Borwein 和 D. Zhuang。关于 Fan 极小极大定理。Math. Programming，34(2)：232–234，1986。[5] L. D. Brown 和 R. Purves。极值的可测选择。Ann. Statist.，1：902–912，1973。[6] Ky Fan。极小极大定理。Proc. Nat. Acad. Sci. U.S.A.，39：42–47，1953。[7] Yu Feng 和 Erik Schl¨ogl。Wasserstein 距离下的模型风险度量。FIRN Research Paper，2018。[8] Rui Gao 和 Anton Kleywegt。基于 Wasserstein 距离的分布鲁棒随机优化。Mathematics of Operations Research，2022。[9] Peyman Mohajerin Esfahani 和 Daniel Kuhn。使用 Wasserstein 度量的数据驱动分布鲁棒优化：性能保证与可处理的重新表述。Mathematical Programming，171(1)：115–166，2018 年 9 月。

[10] C´edric Villani。《最优传输：旧与新》，《数学科学基础丛书》第338卷。施普林格出版社，柏林，2009年。

附录 A：推论 1 中极小化元的可测选择

对于 $x, y \in S$，设 $g(x,y) = f(y) + \lambda c(x,y)$。我们需要证明存在一个可测函数 $T\colon S \to S$，它将 x 映射到 $m(x) := \arg\min_{y\in S}\{f(y) + \lambda c(x,y)\}$。这类问题在统计学文献中曾与贝叶斯程序以及某些估计量的可测性相关联，并得到研究。

下面的结果适用于我们的目的。设 X、Y 为度量空间，且令 $D \subseteq X \times Y$。我们使用记号 $\text{proj}(D) = \{x \in X :$ 存在 $y \in Y$ 使得 $(x,y) \in D\}$。我们将截面 $\{y \in Y$ 记为 $D_x \subseteq Y\colon (x,y) \in D\}$，并将 $F_x(y)$ 记为：$= \inf_{y\in Y} F(x,y)$。

定理3。 *[5, 推论 1]* 设 X、Y 为完备可分度量空间，且 F 是定义在 $X \times Y$ 的博雷尔子集 D 上的实值博雷尔可测函数。

假设对于每个 $x \in \text{proj}(D)$，截面 D_x 是 σ-紧致的，并且 $F(x,\cdot)$ 关于 D_x 上的相对拓扑是下半连续的。那么：

(i) 这些集合

$$
\begin{aligned}
G &= \text{proj}(D), \\
I &= \{x \in G :\ \textit{for some } y \in D_x, F(x,y) = \inf F_x\},
\end{aligned}
$$

是博雷尔可测的。

(ii) 对于每个 $\epsilon > 0$，存在一个博雷尔可测函数 ϕ_ϵ，满足对于 $x \in G$，

$$
\begin{aligned}
f(x, \phi_\epsilon(x)) &= \inf F_x, && \textit{if } x \in I, \\
&\le \epsilon + \inf F_x, && \textit{if } x \notin I, \ \textit{and } \inf f_x \notin -\infty \\
&\le -\epsilon^{-1}, && \textit{if } x \notin I, \ \textit{and } \inf f_x = -\infty.
\end{aligned}
$$

应用此结果于 $X = Y = S$, $F = g$, $\epsilon = 1$, $D = \{(x,y) \in S \times S\colon\ x \in X,\ y \in m(x)\}$，可得一个可测函数 $T := \phi_\epsilon\colon S \to S$，使得 $F(x, T(x)) = \inf F_x$。σ的紧性由以下事实推出：g关于第二个变量是下半连续的，因此 $D_x = m(x) = F_x^{-1}((-\infty, \arg\min F_x])$ 是紧空间 S 的闭子集，故为紧集。在我们的应用中 $G = I = S$。