弥合单目标与多目标博弈之间的鸿沟

Willem Röpke 人工智能实验室 布鲁塞尔自由大学 比利时
willem.ropke@vub.be

Carla Groenland 数学系 乌得勒支大学 荷兰
c.e.groenland@uu.nl

Roxana Rădulescu 人工智能实验室 布鲁塞尔自由大学 比利时
r
oxana.radulescu@vub.be

Ann Nowé 人工智能实验室 布鲁塞尔自由大学 比利时
ann.nowe@vub.be

Diederik M. Roijers 人工智能实验室 布鲁塞尔自由大学 比利时
diederik.roijers
@vub.be

arXiv:2301.05755v2 [cs.GT] 1 Mar 2023

摘要

研究多智能体系统中策略决策的经典模型是标准式博弈。该模型可以推广为允许存在无限多个纯策略，从而形成连续博弈。多目标标准式博弈则是另一种推广，用于刻画玩家在多个目标上分别获得收益的情形。我们通过提供一项理论保证，弥合这两种模型之间的差距：一种设定下的博弈总能转换为另一种设定下的博弈。我们将理论结果扩展至纳什均衡的等价性保证。这种映射使得我们可以将一种领域中的算法应用于另一种领域。为此，我们提出一种用于多目标博弈的虚拟博弈算法，并将其应用于两个著名的连续博弈。我们认为，这种等价关系有助于通过将一种形式体系中的理论保证转换到另一种形式体系中，获得新的见解。此外，当问题采用多目标博弈的简洁形式来求解更为自然时，这种关系或许能为连续博弈带来新的计算方法。

关键词

博弈论；连续博弈；多目标；纳什均衡

ACM 引用格式：
Willem Röpke、Carla Groenland、Roxana Rădulescu、Ann Nowé 和 Diederik M. Roijers。2023。《弥合单目标与多目标博弈之间的差距》。载于第*22*届自主智能体与多智能体系统国际会议论文集（*AAMAS 2023*），伦敦，英国，*2023*年*5*月*29*日至*6*月*2*日，IFAAMAS，共13页。

1 引言

连接看似无关的模型，可以搭建不同研究群体之间的桥梁，并补足任一设定中的缺失部分。为此，我们确定了标准式博弈的两种著名扩展，并指出它们存在可供利用的内在等价关系，从而为理论和算法研究作出贡献。一方面，我们考虑允许玩家拥有无限多个行动的连续博弈 [26]。这类博弈的一个例子见于经济模型，其中企业必须在一个连续范围内设定价格，以便在竞争环境中最大化利润 [14]。另一方面，我们考虑多目标标准式博弈（MONFGs） [2]，它通过返回向量收益而非标量收益来扩展标准式博弈。例如，多目标博弈可用于住宅建筑中的家用电器调度，其目标是在尽可能遵循预期运行时段的同时，尽量降低电力成本 [16]。

我们的理论结果确立了连续博弈与多目标标准式博弈之间一种新颖的等价关系。我们证明，这种等价关系不会改变底层的博弈动态，包括纳什均衡。所确立的联系使我们能够将任一模型中已有的结果和算法直接迁移到另一模型，从而弥合二者之间的差距。从理论角度看，关于连续博弈的已知成果远多于关于多目标标准式博弈的成果，因此能够推动后者迅速发展。反过来，由于形式简洁，多目标标准式博弈适用于通用算法求解，这些算法随后也可用于连续博弈。主要贡献总结如下[1]：

- 我们定义了连续博弈与多目标标准式博弈之间的纯策略等价。这种等价关系形式化地描述了一个双射函数的存在：该函数将连续博弈中的纯策略映射到多目标标准式博弈中的混合策略，并使效用保持相等。
- 我们证明，对于每个策略集为凸集的连续博弈，都可以构造一个与之纯策略等价的多目标标准式博弈。我们还证明了反方向的结论。
- 我们为多目标标准式博弈引入层级策略，并定义这些策略与连续博弈中的混合策略之间的混合策略等价。
- 我们证明，连续博弈中的纯策略纳什均衡是一个等价的多目标标准式博弈中的混合策略均衡。我们将这一结果推广到连续博弈中的混合策略均衡以及多目标标准式博弈中的层级均衡。
- 我们保证效用函数连续的多目标标准式博弈中存在层级纳什均衡。
- 我们通过使用适用于多目标标准式博弈的虚拟博弈算法，在两个连续博弈中计算纳什均衡，展示了我们的等价概念在算法上的价值。这两个博弈分别是多项式博弈和伯特兰价格博弈。

第*22*届自主智能体与多智能体系统国际会议论文集，*A. Ricci*、*W. Yeoh*、*N. Agmon*、*B. An*（编），*2023*年*5*月*29*日至*6*月*2*日，伦敦，英国。 © 2023 国际自主智能体与多智能体系统基金会（www.ifaamas.org）。保留所有权利。

[1]本文的补充材料可在 [25] 中找到。

2 背景

2.1 连续博弈

连续博弈有时也称为无限博弈，它将标准式博弈模型扩展到这样的情形：玩家拥有非空紧致度量空间的纯策略，而不是有限集合。直观地说，这意味着玩家可以从无限多个行动中进行选择。我们将连续博弈定义如下 [26],

定义 2.1 （连续博弈）。连续博弈是一个元组 $(N, \mathrm{S}\ v)$，其中：

- N是一个有限的玩家集合，索引为 i；$\bullet\ \mathrm{S} = S_1 \times \cdots \times S_n$，其中 S_i是玩家 i可用纯策略的非空紧致度量空间；$\bullet\ v = (v_1, \ldots, v_n)$，其中 $v_i \colon \mathrm{S} \to \mathbb{R}$是玩家 i的实值连续效用函数。

连续博弈中玩家 i 的混合策略集合定义为 S_i上的博雷尔概率测度集合，记为 $\mathrm{B}(S_i)$。对于给定的联合混合策略 $\mu \in \mathrm{B(S)}$，我们将玩家 i的期望效用定义如下 [5, 26],

$$v_i(\mu) = \int_{S_1 \times \cdots \times S_n} v_i(s_1, \cdots, s_n) d\mu_1(s_1) \cdots d\mu_n(s_n). \tag{1}$$

对于不熟悉这些测度论概念的读者，我们建议参阅附录 A 中对这些概念的简要介绍。

2.2 多目标标准式博弈

多目标标准式博弈（MONFGs）是将（单目标）标准式博弈推广到向量化收益的一种形式。其形式化定义如下 [20]。

定义 2.2 （多目标标准式博弈）。一个（有限的、含 n 个玩家的）多目标标准式博弈是一个元组（$N, \mathrm{A}\ \boldsymbol{p}$），具有 d 个目标，其中：

- N是一个有限的玩家集合，玩家以 i 为索引；$\bullet\ \mathrm{A} = A_1 \times \cdots \times A_n$，其中 A_i 是玩家 i 可选行动的非空有限集合；
- $\boldsymbol{p} = (\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n)$，其中 $\boldsymbol{p}_i \colon \mathrm{A} \to \mathbb{R}^d$是玩家 i 的向量收益函数。

任意给定玩家的混合策略定义为其行动集合上的概率分布。有限个点上的概率分布集合称为概率单纯形，其形式化定义如下。

定义 2.3. 概率 k单纯形 Δ^k 是一个点集，其中，

$$\Delta^k = \left\{ (x_0, \ldots, x_k) \in \mathbb{R}^{k+1} \;\middle|\; \sum_{i=0}^{k} x_i = 1 \text{ and } x_i \geq 0 \text{ for } i = 0, \ldots, k \right\}.$$

因此，玩家 i 的混合策略是一个概率分布$\delta_i \in \Delta^{k_i}$，其中 $k_i = |A_i| - 1$。令 $\Delta = \Delta^{k_1} \times \cdots \times \Delta^{k_n}$ 表示联合混合策略集合。玩家 i采用混合策略 $\delta \in \Delta$ 时的期望收益自然定义为：

$$\boldsymbol{p}_i(\delta) = \sum_{a \in \mathcal{A}} \boldsymbol{p}_i(a) \prod_{j=1}^{n} \delta_j(a_j). \tag{2}$$

多目标决策通常采用基于效用的方法 [22], ，这也是一种标准的博弈论方法。在多目标标准式博弈模型中，玩家的效用取决于对各项收益的特定权衡；也就是说，我们额外假设，已知任意玩家 i的效用函数 $u_i \colon \mathbb{R}^d \to \mathbb{R}$。我们将混合策略 δ的效用计算为：

$$u_i(\boldsymbol{p}_i(\delta)) = u_i\left(\sum_{a \in \mathcal{A}} \boldsymbol{p}_i(a) \prod_{j=1}^{n} \delta_j(a_j) \right). \tag{3}$$

我们注意到，对于多目标博弈中混合策略的效用，还有另一种定义：玩家首先对收益应用效用函数，随后计算其期望效用。在多目标决策文献中，这种方法称为期望标量化回报准则，而式（2）中的方法称为标量化期望回报准则 [10, 19]。

2.3 纳什均衡

我们考虑两种模型中的一个核心解概念，即纳什均衡（NE）。通俗地说，纳什均衡是一种联合策略，任何玩家都无法通过单方面偏离来提高自己的效用。定义 2.4 给出了连续博弈中的定义，而定义 2.5 则给出了多目标标准式博弈中的定义。

定义 2.4 （连续博弈中的纳什均衡）。如果混合策略组合 μ^* 满足，

$$v_i(\mu_i^*, \mu_{-i}^*) \geq v_i(\mu_i, \mu_{-i}^*),$$

对于所有玩家 i和混合策略 $\mu_i \in \mathrm{B}(S_i)$。

定义 2.5 （多目标标准式博弈中的纳什均衡）。若混合策略组合 δ^* 满足以下条件，则称其为纳什均衡：

$$u_i(\delta_i^*, \delta_{-i}^*) \geq u_i(\delta_i, \delta_{-i}^*),$$

对于所有玩家 i和混合策略 $\delta_i \in \Delta^{k_i}$。

一个重要的早期结果是，每个连续博弈都必定存在混合策略纳什均衡 [8]。然而，在多目标标准式博弈中，采用非线性效用函数时，并不保证纳什均衡存在 [20]。在第 3.3 节中，我们将进一步探讨这一问题，并为多目标标准式博弈提出一个新的纳什均衡存在性结果。

3 等价关系

我们的主要贡献是建立连续博弈与多目标博弈之间的等价关系。为此，我们首先引入一种特殊的博弈，称为恒等博弈，它在连接这两种模型方面起着关键作用。接下来，我们证明，连续博弈中的纯策略到多目标标准式博弈中的混合策略总是存在双射。为实现完全等价，我们为多目标标准式博弈引入一个新概念，称为层级策略。最后，我们证明纳什均衡在这一过程中得以保留，从而保证多目标标准式博弈的层级策略中存在纳什均衡。

3.1 恒等博弈

顾名思义，恒等博弈返回的收益向量与其接收到的输入策略相同。我们将在此类博弈

第3.2节中使用它们来证明，对于每个连续博弈，都可以构造出一个等价的多目标标准式博弈，反之亦然。

引理 3.1（恒等博弈）。对于任意有限玩家集合及有限的纯策略集，存在一组收益函数 $\boldsymbol{p}$，使得对于每位玩家 i，$\boldsymbol{p}_i(\delta) = \delta_i$。

证明概要。证明的核心思路是将所有纯策略对应的收益向量定义为纯策略本身。严格来说，这依赖于联合策略的记号，我们将联合策略定义为各个单独策略拼接而成的向量。随后可以利用全概率公式证明，这样便能得到所需性质。□

形式化证明见附录B。为了更详细地说明收益机制，我们在下面给出一个示例。

	A	B
A	$(1,0,1,0);(1,0,1,0)$	$(1,0,0,1);(1,0,0,1)$
B	$(0,1,1,0);(0,1,1,0)$	$(0,1,0,1);(0,1,0,1)$

图1：双人双行动情形下的恒等博弈。

例 1. 考虑图 1 中的恒等博弈。假设玩家一采用混合策略 $\delta_1 = (\frac{1}{2}, \frac{1}{2})$，玩家二采用混合策略 $\delta_2 = (\frac{1}{3}, \frac{2}{3})$。这将产生联合策略 $\delta = (\delta_1, \delta_2) = (\frac{1}{2}, \frac{1}{2}, \frac{1}{3}, \frac{2}{3})$。根据引理 3.1，期望收益向量也应为 $(\frac{1}{2}, \frac{1}{2}, \frac{1}{3}, \frac{2}{3})$。我们来验证这一点：

$$
\begin{aligned}
\boldsymbol{p}_i(\delta) &= \sum_{a \in \mathcal{A}} \boldsymbol{p}_i(a) \prod_{j=1}^{n} \delta_j(a_j) \\
&= (1,0,1,0) \cdot \frac{1}{6} + (1,0,0,1) \cdot \frac{1}{3} + (0,1,1,0) \cdot \frac{1}{6} + (0,1,0,1) \cdot \frac{1}{3} \\
&= \left(\frac{1}{6}, 0, \frac{1}{6}, 0\right) + \left(\frac{1}{3}, 0, 0, \frac{1}{3}\right) + \left(0, \frac{1}{6}, \frac{1}{6}, 0\right) + \left(0, \frac{1}{3}, 0, \frac{1}{3}\right) \\
&= \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{3}, \frac{2}{3}\right) \\
&= \delta.
\end{aligned}
$$

3.2 纯策略等价

我们提出连续博弈与多目标标准式博弈之间的一种新等价概念，称为纯策略等价（PSE）。通俗地说，当连续博弈中的纯策略能够双射映射到多目标标准式博弈中的混合策略，同时保持对应效用相等时，这两个博弈就是纯策略等价的。我们将在下文对其进行正式定义。

定义 3.1（纯策略等价）。连续博弈 $G_c = (N_c, \mathrm{S}\ v)$ 与一个效用函数为 u 的有限多目标标准式博弈 $G_m = (N_m, \mathrm{A}\ \boldsymbol{p})$ 纯策略等价，当且仅当存在一组函数 (π, φ)，使得：

- $\pi\colon N_c \to N_m$ 是一个双射，称为玩家双射；
- $\varphi = \varphi_1 \times \cdots \times \varphi_n$ 其中 $\forall i \in N_c$，$\varphi_i\colon S_i \to \Delta^{k_{\pi(i)}}$ 是一个连续双射函数，且其逆函数连续，称为策略双射；
- $\forall i \in N_c, \forall s \in \mathcal{S}, v_i(s) = u_{\pi(i)}\left(\boldsymbol{p}_{\pi(i)}\left(\varphi(s)\right)\right)$.

我们首先证明，对于每个多目标标准式博弈，都存在一个纯策略等价的连续博弈。我们注意到，这一性质已在 [23, 引理 1], 中使用，但尚未从纯策略等价的角度予以明确描述。

定理 3.2。对于每个具有连续效用函数的多目标标准式博弈，都存在一个与之纯策略等价的连续博弈。

证明。设 $G_m = (N_m, \mathrm{A}\ \boldsymbol{p})$ 是一个多目标标准式博弈，并设 u 是玩家在 G_m 中使用的连续效用函数集合。我们构造一个与 G_m 纯策略等价的连续博弈 $G_c = (N_c, \mathrm{S}\ v)$。

首先，令 $N_c = N_m$，从而得到玩家双射 $\pi = \mathbb{1}$。为简化记号，我们可以直接代入 $\pi(i) = i$。

回顾一下，玩家 i in G_m 的混合策略集合被定义为一个单纯形 Δ^{k_i}。对于所有 $S_i \in \mathrm{S}$，我们定义 $S_i = \Delta^{k_i}$，从而满足其为非空紧致度量空间的条件。因此，每个策略双射 $\varphi_i = \mathbb{1}$。

最后，定义每个 $v_i = u_i \circ \boldsymbol{p}_i$。注意，每个 v_i 都是连续的，因为它是两个连续函数的复合。因此，G_c 与 G_m 纯策略等价。□

此处给出的证明为任意多目标标准式博弈显式构造了一个纯策略等价的连续博弈，具体步骤概述于附录 C.1（见算法 2）。此外，由于两个博弈中的策略集合相同，策略双射就是恒等函数。从算法角度来看，这一点至关重要，因为它确保可以高效地计算该映射。这一结果有两个重要的意义。首先，我们注意到，也可以在单人情形下进行这一构造，例如多目标规划和强化学习中研究的情形。这表明，只要事先知道智能体的效用函数，就可以使用为连续行动空间设计的算法。其次，只要所得效用函数（二阶）可微，所得博弈就属于可微博弈这一类别，可以使用高效的基于梯度的方法求解 [15]。

现在我们证明定理 3.2 的逆命题，即每个策略集为凸集的连续博弈都可以映射为一个纯策略等价的多目标标准式博弈。

定理 3.3. 对于策略空间为欧几里得空间中凸子集的每个连续博弈，都存在一个纯策略等价的多目标标准式博弈。

证明。设 $G_c = (N_c, \mathrm{S}\ v)$ 是一个连续博弈，其中每个 $S_i \in \mathrm{S}$ 都是欧几里得空间中的紧致、凸且非空子集。我们构造一个有限多目标标准式博弈 $G_m = (N_m, \mathrm{A}\ \boldsymbol{p})$，其效用函数为 u，且该博弈与 G_c 纯策略等价。

令 $N_m = N_c$，即可得到显然的玩家双射，因此 $\pi = \mathbb{1}$。为简化记号，我们可以直接代入 $\pi(i) = i$，并将 N_m 或 N_c 简称为 N。

已知，所有具有非空内部的 $\mathbb{R}^k$ 紧凸子集都与概率 k-单纯形 [3, 同胚。定理 16.4]。我们首先假设每个策略集

S_i 确实具有非空内部。这意味着，对于每个玩家i，都存在一个连续双射函数 f_i，且其具有连续逆函数f_i^{-1}，使得，

$$f_i : S_i \to \Delta^{k_i}. \tag{4}$$

回顾一下，玩家 i 在 G_m 中的混合策略集合 Δ^{k_i} 定义为其行动上的概率单纯形。因此，我们可以利用这个同胚构造关于 $k_i + 1$ 行动的混合策略集合，

$$\forall i \in N : f_i(S_i) = \Delta^{k_i}. \tag{5}$$

注意，单纯形的顶点自然表示 G_m中的行动。这确保了每个联合策略都可以从 G_c 到 G_m 进行双射映射，且

$$\forall s \in \mathcal{S} : f(s) = f_1(s_1) \times \cdots \times f_n(s_n). \tag{6}$$

因此，我们可以定义策略双射 $\varphi = f$。由于 f是一个同胚，按定义，它是一个连续双射函数，且其逆映射连续。

令 G_m 的收益函数 $\boldsymbol{p}$ 为引理 3.1 中恒等博弈的收益。我们在 G_m 中定义效用函数 u 如下，

$$\forall i \in N, u_i = v_i \circ f_i^{-1}. \tag{7}$$

我们证明这确保了，

$$\forall i \in N, \forall s \in \mathcal{S} : v_i(s) = u_i\left(\boldsymbol{p}_i\left(\varphi_i(s)\right)\right). \tag{8}$$

代入所需的值后，我们得到 $\forall i \in N$， $\forall s \in \mathrm{S}$,

$$\begin{aligned} u_i\left(\boldsymbol{p}_i\left(\varphi_i(s)\right)\right) &= u_i\left(\boldsymbol{p}_i\left(f_i(s)\right)\right) && (9)\\ &= u_i\left(\boldsymbol{p}_i(\delta)\right) && (10)\\ &= u_i(\delta) && (11)\\ &= v_i \circ f_i^{-1}(\delta) && (12)\\ &= v_i(s). && (13) \end{aligned}$$

因此， G_c 与 G_m纯策略等价。

如果一个或多个 S_i 的内部为空，那么这些 S_i 相对于其仿射包具有非空内部。于是，S_i 与 k-单纯形同胚，其中 $k = \dim \mathrm{aff}(S_i)$。其余构造与前述情况类似。□

该证明的核心思路是利用恒等博弈收益构造一个多目标标准式博弈，并沿用连续博弈中的效用函数。由于我们假设策略集具有凸性，因此可以为每个策略集构造一个与概率单纯形之间的同胚。这些单形就是该多目标标准式博弈中的混合策略。随后，可以将连续博弈的效用函数与每个同胚的逆映射复合，从而定义该多目标标准式博弈中的效用函数。

本证明中定义的构造也建立了一种将连续博弈转化为多目标标准式博弈的计算方法，具体形式见附录 C.1（参见算法 3）。然而，与相反方向的情形不同，这里确实存在策略双射。因此，该算法要求明确地定义策略双射，这带来了两个不同的挑战。首先，这些策略双射可能不易获得；对此，我们在附录 C.2 中提出一种标准方法：先经过单位球，再映射到概率单纯形，可作为一种合适的初步尝试。其次，所得函数不保证能够高效地计算，因此在某些应用中，该过程可能难以处理。我们将定理 3.3 在此类情形中的应用留待未来工作。

定理 3.3 的一个有趣推论是，只要连续博弈的策略集仍与单纯形同胚，该定理就可以直接推广到策略集非凸的连续博弈。然而，虽然在连续博弈中通常会假设策略集具有凸性，但也存在不具备这一性质的相关博弈。在这种情况下，例如可以取每个策略集的凸包，用另一个满足凸性要求的连续博弈来近似原连续博弈。随后，可以为这个近似博弈构造一个纯策略等价多目标正规形式博弈，从而得到一个与原连续博弈近似纯策略等价的多目标标准式博弈。我们将在第 4.3 节展示这一方法，并说明它仍有可能捕捉到原博弈。

鉴于定理 3.2 和定理 3.3 中概述的构造，我们不禁要问，这些构造能否唯一确定纯策略等价博弈。下面的备注表明并非如此。我们在附录 C.1 中通过构造多个纯策略等价博弈对此进行了形式化证明。

备注1。 一个连续博弈可能对应多个纯策略等价多目标标准式博弈，反之亦然。

备注1的一个重要推论是，有些纯策略等价博弈比其他博弈更易于分析。例如，我们已经证明，对于一个连续博弈，可以构造一个收益为恒等博弈收益的纯策略等价多目标标准式博弈。如果还能证明该连续博弈与一个在其收益中包含被支配行动且具有单调效用函数的多目标标准式博弈纯策略等价，那么后者无疑会更容易分析。利用博弈之间的等价关系，通过较简单的等价博弈求解复杂博弈，过去已有成功案例 [11]。

3.3 混合策略等价

鉴于纯策略等价的相关结果，一个亟待解决的问题是，如何在等价的多目标标准式博弈中处理连续博弈的混合策略。为此，我们在多目标标准式博弈中引入一种新颖的策略概念，称为层级策略。层级策略允许对混合策略进行混合，并针对每个玩家 i 定义为其混合策略集合 Δ^{k_i} 上的博雷尔概率测度集合，记作 $\mathrm{B}(\Delta^{k_i})$。层级策略 μ 的期望效用定义方式与式 (1) 中连续博弈混合策略的期望效用定义方式类似，即，

$$u_i(\mu) = \int_{\Delta^{k_1} \times \cdots \times \Delta^{k_n}} u_i\left(\boldsymbol{p}_i(\delta_1, \cdots, \delta_n)\right) d\mu_1(\delta_1) \cdots d\mu_n(\delta_n). \tag{14}$$

我们强调，层级策略中的概率不能重新分配，从而形成等价的混合策略。这是因为所得混合策略的效用未必等于层级策略的期望效用，尤其是在使用非线性效用函数时。我们在例 2 中对此进行说明。

例 2. 考虑图 2 中的博弈及其效用函数

$$u_1(x, y) = u_2(x, y) = x^2 + y^2. \tag{15}$$

	A	B
A	(3，1)；(3，1)	(1，3)；(1，3)
B	(1，3)；(1，3)	(3，1)；(3，1)

图2：例 2 中使用的博弈。

对于行玩家，我们定义两种策略：$\delta_{1,1}=(1,0)$ 和 $\delta_{1,2}=(0,1)$。为简单起见，我们假设列玩家采用确定性策略 $\delta_2=(1,0)$，即始终选择行动 A。对于双方玩家而言，联合策略 $(\delta_{1,1},\delta_2)$ 会产生 (3，1) 的期望收益向量，而 $(\delta_{1,2},\delta_2)$ 会产生 (1，3) 的期望收益向量。这两种联合策略各自带来的效用均为 10。

现在考虑行玩家的以下层级策略，

$$\mu_1=\left(P(\delta_{1,1})=\frac{1}{2},P(\delta_{1,2})=\frac{1}{2}\right).$$

这种层级策略表示，他们会以 50% 的概率采取 $\delta_{1,1}$，并以 50% 的概率采取$\delta_{1,2}$。由于这两种策略的效用均为 10，因此 μ_1 的期望效用也为 10。然而，将 μ_1 中的概率分配到各项中以形成混合策略 $\delta_{1,3}=(\frac{1}{2},\frac{1}{2})$，得到的期望收益向量为 $(2,2)$，效用为 8。这表明，层级策略无法拆分为与之等价的混合策略。

虽然层级策略主要作为我们分析中的一种工具，但在多目标博弈中也可能具有实际用途。具体而言，当智能体必须选定一种策略，并在给定的一段时间内执行该策略，且期间无法进行后续调整时，考虑采用层级策略是合适的。例如，参与拍卖的智能体可能必须事先向自动拍卖主持人承诺采用单一混合策略，之后拍卖将进行若干轮次，期间智能体无法再提供任何输入。

现在，我们定义连续博弈与多目标标准式博弈之间的混合策略等价。非正式地说，这一等价概念将纯策略等价推广到连续博弈中的混合策略与多目标标准式博弈中的层级策略之间的对应关系。

定义 3.2（混合策略等价）。设 $G_c=(N_c,S,v)$ 是一个连续博弈，$G_m=(N_m,A,\boldsymbol{p})$ 是一个有限多目标标准式博弈，其效用函数为 u。若 G_c 与 G_m 在 (π,φ) 下纯策略等价，且存在一个函数 ψ 使得，

- $\psi=\psi_1\times\cdots\times\psi_n$其中 $\forall i\in N_c$，$\psi_i:\mathrm{B}(S_i)\to\mathrm{B}(\Delta^{k_{\pi(i)}})$ 是一个双射函数；
- $\forall i\in N_c$，$\forall\mu\in\mathrm{B}(S)$ $v_i(\mu)=u_\pi(i)(\psi(\mu))$。

我们指出，根据定义，混合策略等价意味着纯策略等价。此外，请回想一下，连续博弈中混合策略的定义与多目标标准式博弈中层级策略集合的定义相似。这并非巧合，也使我们能够证明：只要一个连续博弈与某个多目标标准式博弈纯策略等价，它们也必然混合策略等价。正式证明见附录 D。

定理 3.4。如果一个连续博弈与一个多目标标准式博弈纯策略等价，那么它们也混合策略等价。

证明概要。根据纯策略等价的定义，存在一个带有连续逆函数的连续双射函数，将一个博弈中的策略映射到另一个博弈中。此外，混合策略和层级策略分别被定义为其纯策略和混合策略上的所有博尔测度的集合。我们可以利用这一事实证明，在一个策略集合上定义的测度，在另一个由映射后策略构成的集合上也必然是一个测度。这也意味着效用相等，从而完成证明。□

3.4 纳什均衡的映射

我们最后的理论贡献探讨了纯策略等价博弈中的纳什均衡。定理 3.5 首先指出，连续博弈中的纯策略纳什均衡对应于纯策略等价多目标标准式博弈中的混合策略纳什均衡。直观而言，这是显而易见的，因为连续博弈中纯策略的效用已被保证与多目标标准式博弈中混合策略的效用相等。因此，如果某个联合策略在任一博弈中都无法通过单方面行动加以改进，那么映射到相关博弈中的联合策略也同样无法改进。

定理 3.5。在连续博弈中，纯策略是纳什均衡，当且仅当它是纯策略等价多目标标准式博弈中的混合策略纳什均衡。

我们在下文定义层级纳什均衡，以便构造类似的论证，证明连续博弈中的混合策略纳什均衡必然对应于纯策略等价多目标标准式博弈中的层级策略纳什均衡。

定义 3.3（多目标标准式博弈中的层级纳什均衡）。若层级策略组合 μ^* 满足以下条件，则称其为层级纳什均衡：

$$u_i(\mu_i^*,\mu_{-i}^*)\geq u_i(\mu_i,\mu_{-i}^*),$$

对于所有玩家 i 以及备选层级策略 $\mu_i\in\mathrm{B}(\Delta^{k_i})$。

定理 3.6。在连续博弈中，混合策略是纳什均衡，当且仅当它是纯策略等价多目标标准式博弈中的层级纳什均衡。

定理 3.5 和定理 3.6 的完整证明见附录 E。这些性质非常重要，因为它们引入了算法方法，使得可以在一种博弈模型中计算纳什均衡，并将其用于另一种模型。此外，特别是作为定理 3.6 的推论，我们可以给出关于多目标标准式博弈中纳什均衡存在性的首个一般性结果，前提是允许玩家采用层级策略，而非仅限于混合策略。推论 3.7 的证明如下。

推论 3.7。每个具有连续效用函数的多目标标准式博弈都有一个层级纳什均衡。

证明。定理 3.2 表明，每个效用函数连续的多目标标准式博弈（MONFG）都可以映射为一个纯策略等价的连续博弈。此外，已知每个连续博弈都存在混合策略纳什均衡 [8]。定理 3.6 保证，连续博弈中的混合策略纳什均衡是 MONFG 中的层级纳什均衡，因此保证了每个 MONFG 都存在层级纳什均衡。□

算法1多目标虚拟博弈

输入： 一个多目标标准式博弈 $G = (N, A, \boldsymbol{p})$，效用函数 u和最大时间步 T

输出： 一个联合策略 δ

$\delta_1 \leftarrow \frac{\mathbf{1}}{|A_1|}$
$h_1 \leftarrow \mathbf{0}$ ▷ 玩家 1 的历史
$\delta_2 \leftarrow \frac{\mathbf{1}}{|A_2|}$
$h_2 \leftarrow \mathbf{0}$
对于 $t \in 1, \cdots T$**执行**
  $\tilde{\delta}_2 \leftarrow \frac{h_1}{t}$ ▷ 计算经验混合策略
  $\tilde{\delta}_1 \leftarrow \frac{h_2}{t}$
  $\delta_1 \leftarrow$ 最佳响应$(\boldsymbol{p}_1, u_1, \tilde{\delta}_2)$ ▷ 计算最佳响应
  $\delta_2 \leftarrow$ 最佳响应$(\boldsymbol{p}_2, u_2, \tilde{\delta}_1)$
  $a_1 \leftarrow a \sim \delta_1$ ▷ 从最佳响应中采样一个行动
  $a_2 \leftarrow a \sim \delta_2$
  $h_{1,a_2} \leftarrow h_{1,a_2} + 1$ ▷ 更新历史
  $h_{2,a_1} \leftarrow h_{2,a_1} + 1$
结束循环
返回 δ

4 实证结果

我们为所提出的定理提供了实证依据，并表明当策略集不满足纯策略等价所需的条件时，也可以用它来计算近似均衡。我们将单目标博弈中广为人知的虚拟博弈算法改编用于多目标博弈，并用它来计算连续博弈中的纯策略纳什均衡。结果从实证角度展示了我们贡献的适用性，也可为未来的应用提供有益范例。我们的实现可在 https://github.com/wilrop/pure-strategy-equivalence 获取。

4.1 多目标虚拟博弈

虚拟博弈旨在通过反复进行博弈，学习能够达到纳什均衡的策略。虽然虚拟博弈无法保证在一般和单目标博弈中收敛，但它及其扩展在实践中得到了广泛应用。算法1展示了虚拟博弈向多目标博弈的扩展。为简单起见，我们考虑双人变体，但也可以轻松扩展到 n人博弈。在算法的每次迭代中，玩家根据博弈历史计算对手的经验策略，并针对该策略计算最佳响应。随后，玩家从新策略中采样一个行动，并更新各自的历史。

近期研究探讨了将虚拟博弈改编用于连续博弈的方法 [7]。在他们的算法中，会保留一个包含过去策略的动态增长数组，以便之后计算最佳响应，这会带来很大的内存需求。我们的方法有一个关键优势：只需使用固定长度的数组，即每个行动对应一个条目，并在每次采取某个行动时递增相应计数器。随后，通过计算对手采取各个行动的相对频率，得出其经验混合策略。这种方法的局限在于，它只能从连续博弈中学习纯策略均衡。

上文所示的虚拟博弈算法看起来与原始虚拟博弈算法完全相同。不过，两者的区别在于最佳响应计算步骤。在单目标博弈中，可以通过选择期望收益最高的行动来高效完成，即

$$BR(A_i, s_{-i}, p_i) = \arg\max_{a_i \in A_i} p_i(a_i, s_{-i}). \tag{16}$$

在多目标博弈中，只有采用拟凸效用函数 [23]，才能保证这种方法返回正确的最佳响应。在一般的多目标标准式博弈中，最佳响应可能是混合策略，因此需要执行优化子程序，以找到能够产生最大效用的策略。由于最佳响应必须是全局最大值，因此需要使用全局优化算法。在特定效用函数下，或者近似最佳响应已足够时，也可以使用局部优化器。

4.2 多项式博弈

多项式博弈是连续博弈的一个子集，其效用函数保证是玩家策略 [26, 27] 的多项式函数。我们将展示，这类博弈也可以表示为多目标标准式博弈（MONFG），且无需使用任何针对连续博弈或多项式博弈的专用工具即可求解。我们将介绍 Parrilo [18] 所描述的一个简单示例。

考虑一个零和博弈，双方玩家都从区间 $[-1, 1]$ 中选择策略。玩家一的效用函数定义如下：

$$v_1(x, y) = 2xy^2 - x^2 - y, \tag{17}$$

其中，x 是玩家一选择的策略，y 是玩家二的策略。由于该博弈是零和博弈，玩家二的效用为 $v_2(x, y) = -v_1(x, y)$。该博弈所用的效用函数保证存在一个纯策略纳什均衡，且该均衡唯一，其中 $x^* = 0.397$ 且 $y^* = 0.630$。

由于策略集是线段，因此也是 1-单形，故必然存在一个纯策略等价的多目标博弈。要完成从多项式博弈到多目标博弈的转换，每个玩家 i都需要一个策略双射 $\varphi_i \colon S_i \to \Delta^{k_i}$。双方玩家的策略双射如下：

$$\varphi_i(s_i) = \left(\frac{s_i - s_{i,\min}}{s_{i,\max} - s_{i,\min}}, 1 - \frac{s_i - s_{i,\min}}{s_{i,\max} - s_{i,\min}} \right), \tag{18}$$

其中，对于两位玩家，$s_{i,\text{最小化}} = -1$ 和$s_{i,\text{最大化}} = 1$。逆策略双射定义为：

$$\varphi_i^{-1}(\delta_i) = s_{i,\min} + \delta_{i,0} \cdot \left(s_{i,\max} - s_{i,\min}\right). \tag{19}$$

因此，最终的多目标博弈有两位玩家，每位玩家各有两个行动，并具有相应的恒等博弈收益。此外，两位玩家的效用函数均为 $u_i = v_i \circ \varphi_i^{-1}$。由于原始效用函数 v_1 和 v_2 保证连续博弈中存在纯策略纳什均衡，因此虚拟博弈非常适合用于寻找该多目标标准式博弈中的混合策略均衡。

我们在构建的多目标博弈上执行虚拟博弈算法 200 次迭代，并将此过程重复 1000 次试验。图3展示了随时间变化而学习到的策略，其中阴影区域表示该时刻的标准差。我们说明

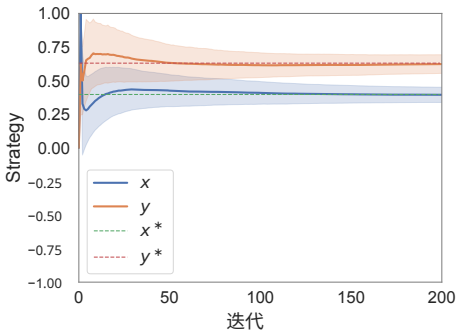


图3：多项式博弈的学习曲线。

纳什均衡（0.397,0.630）用虚线表示。显然，我们的算法在约 100 次迭代后学会了该均衡，并且能够随着时间推移不断改进其策略，使其更加接近精确均衡。

4.3 伯特兰价格博弈

定理 3.3 指出，只有对于策略空间为欧几里得空间凸子集的连续博弈，才能保证存在纯策略等价多目标正规形式博弈。我们通过对连续博弈进行近似，展示即使不满足这一条件，仍然可以应用纯策略等价。我们以 Judd 等人 [14] 描述的伯特兰价格博弈为例进行说明。

伯特兰价格博弈作为一种用于确定竞争环境中价格的经济模型，已得到广泛研究。在此例中，我们考虑两家企业 x 和 y，它们分别以价格 p_x 和 p_y 生产不同的商品。有三类顾客，他们对两种商品的需求各不相同。第一类顾客对商品有线性需求曲线 $d_{x,1}$ 和 $d_{y,1}$，且只想购买企业 x 的商品，

$$d_{x,1}(p_x, p_y) = a - p_x \qquad d_{y,1}(p_x, p_y) = 0, \tag{20}$$

其中 a 表示除价格之外影响需求的所有因素。第三类顾客对企业 y 商品的需求函数以类似方式定义，

$$d_{x,3}(p_x, p_y) = 0 \qquad d_{y,3}(p_x, p_y) = a - p_y. \tag{21}$$

最后，第二类顾客对两种商品都有需求，

$$d_{x,2}(p_x, p_y) = n \cdot p_x^{-\sigma} \left(p_x^{1-\sigma} + p_y^{1-\sigma}\right)^{(\gamma-\sigma)/(-1+\sigma)} \tag{22}$$

$$d_{y,2}(p_x, p_y) = n \cdot p_y^{-\sigma} \left(p_y^{1-\sigma} + p_x^{1-\sigma}\right)^{(\gamma-\sigma)/(-1+\sigma)}. \tag{23}$$

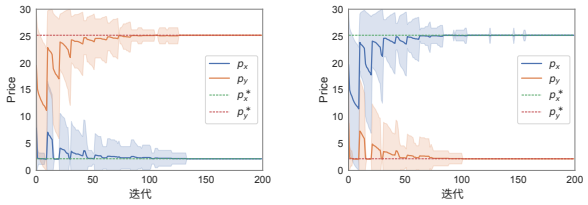
其中，n 为二型顾客的数量，σ 为 x 与 y 之间的替代弹性，γ 为复合商品的需求弹性。各商品的总需求，分别为 d_x 和 d_y，通过对每种类型顾客的个体需求求和得到。最后，设 m 为每家企业的单位生产成本，则两家企业的利润定义为：

$$r_x(p_x, p_y) = (p_x - m) \cdot d_x(p_x, p_y) \tag{24}$$

$$r_y(p_x, p_y) = (p_y - m) \cdot d_y(p_x, p_y). \tag{25}$$

p_x	p_y	r_x	r_y
2.168	25.157	**724.337**	608.981
25.157	2.168	608.981	**724.337**

表1：纳什均衡策略及其利润。企业 x 和 y 的最高利润已突出显示。



(a) 均衡点(2.168,25.157)对应的学习曲线。(b) 均衡点(25.157,2.168)对应的学习曲线。

图4：区间 $[1, 30]$ 内伯特兰价格博弈的学习曲线。

博弈中所考虑的可能价格范围是开区间 $(0, +\infty)$。因此，连续博弈中的策略空间是非紧的，从而违反了纯策略等价所需的一个必要条件。不过，我们可以通过对策略空间作紧凸近似，并改用这些近似空间来解决这一问题。具体而言，我们将价格限制在闭区间 $[p_{\text{最小值}}, p_{\text{最大值}}]$ 内，从而确保策略集与前一个示例一样是1-单形。由于采用了这一近似，我们可以重新使用公式（18）和（19）中定义的相同策略双射。需要注意的是，通过改变策略集来近似连续博弈，可能会使现有均衡变得不可达，或引入新的均衡。在这个具体示例中，由于我们同时对策略集设置了下界和上界，因此某个均衡可能落在界限之外，并且多目标标准式博弈中可能会产生一个在原始博弈中并非均衡的新均衡。

对于以下实验，我们定义 $\sigma = 3, \gamma = 2,$ $n = 2700$ $m = 1$ $a = 50$，和 。在这些参数下，价格博弈有两个不同的纳什均衡，如表1所示。

4.3.1 合适的近似。为使两个均衡都有机会被选中，我们设定 $p_{\text{最小值}} = 1$ 和 $p_{\text{最大值}} = 30$。每次执行虚拟博弈算法时均迭代200次，并像上一节一样，对1000次试验的结果取平均。

在图4中，我们展示了通向两个均衡的轨迹。在较早的轮次中，轨迹并不平滑，标准差也很高，因为基于有限的博弈历史进行最佳响应计算，会导致智能体迅速改变策略。然而，约150轮后，信念趋于收敛，纳什均衡便会持续出现。我们还发现，两个均衡对应的学习轨迹相似，这表明学习过程中最终遵循的具体轨迹主要取决于早期的随机化。这些结果表明，即使是对连续博弈的近似，只要能够充分捕捉原博弈的特征，也可以应用多目标算法。

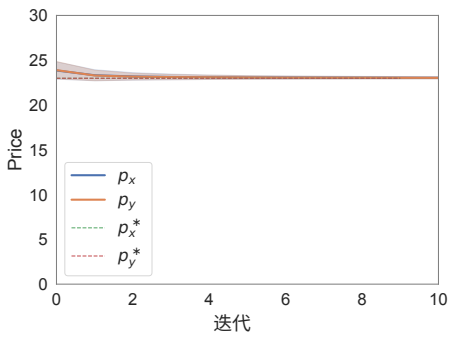


图5：区间 [4, 30] **内伯特兰价格博弈的学习曲线。**

4.3.2 不合适的近似。接下来，我们考察使用不合适的策略集近似会产生什么结果。具体而言，我们将最低价格提高到4，这使表1中的两个均衡都不再可能。直观来说，这些均衡中，一家企业选择价格较低的大众市场策略，另一家则选择价格较高的利基策略。提高最低价格后，大众市场策略便不再可行。图5展示了这一实验产生的学习轨迹。

我们发现，在所有试验中，企业都迅速收敛至联合策略(22.987,22.987)，使双方的利润均达到 673.38。事实上，该联合策略是多目标标准式博弈中的纳什均衡，但不是原始连续博弈的纳什均衡，从而证明了上述近似技术的局限性。有趣的是，在近似的多目标标准式博弈中，纳什均衡所对应的联合策略在考虑利润总和以及最低利润最大化时，都能带来更高的社会福利。具体而言，该策略带来的总利润为 1346.754，而表1中的两个均衡带来的总利润均为 1333.318，且最低利润更高。因此，即使对原始博弈采用的近似并不充分，我们的贡献仍可能产生有趣的解决方案，例如从机制设计的角度来看。

我们注意到，为任意连续博弈设计适当的近似并非易事。对于大多数有趣的应用而言，合适的凸紧近似并非现成可得，因此需要借助领域知识或进行后处理，以确认所获得的解在原始博弈中是否成立。

5 相关工作

多目标博弈最早由 Blackwell [2] 提出，此后得到了广泛研究。一种常见的解概念是帕累托-纳什均衡，它采用效用无关方法，因此直接基于向量化收益定义 [13]。采用基于效用的方法时 [22], , 研究表明纳什均衡未必存在 [20]。后续研究表明，只假设效用函数为拟凹函数 [23] , 就足以再次保证均衡存在。从计算角度来看，强化学习算法以及对手建模 [21, 28] 和通信 [24] 等其他技术近期均已得到成功探索。

本研究考虑的第二种博弈模型是连续博弈。关于这类博弈，已有一般性的存在性保证，例如针对纳什均衡 [8] 和相关均衡 [9]的研究。从计算角度来看，虚拟博弈 [7]和无悔学习均已得到探索 [12], , 后者还获得了强有力的收敛保证。需要指出的是，尽管我们遵循 Stein等人 [26], 对连续博弈的定义，但连续博弈还有其他定义，它们对策略集或效用函数 [1, 7, 12]作出了不同假设。最后，多项式博弈作为连续博弈的一个子类，其纳什均衡和相关均衡已得到详尽研究，并取得了理论与算法方面的成果 [26, 27]。

我们的工作与博弈论中的其他等价概念有关。策略等价是这类等价概念中首个值得关注的例子 [17]。策略等价的一个优点是能够保留纳什均衡，因此可作为计算博弈中纳什均衡的有用工具。例如，对于一个难以求解的博弈，可以构造一个与之策略等价的零和博弈，再利用已有的高效求解方法计算其纳什均衡 [11]。

第3.2节中定义的纯策略等价与博弈同构的概念最为接近。博弈同构将两个博弈定义为同构，当且仅当存在一个从其中一个博弈到另一个博弈的映射 [6]。博弈同构有两个变体，即强同构和弱同构：强同构保留所有纳什均衡，而弱同构仅保留纯策略纳什均衡。

6 结论

我们提出了一种新颖的等价类，称为纯策略等价，用于描述连续博弈与多目标标准式博弈之间的关系。我们证明，对于策略空间为欧几里得空间凸子集的每个连续博弈，都可以构造出一个与之纯策略等价的多目标标准式博弈，反之亦然。此外，这种等价关系意味着纳什均衡得以保留。我们通过使用多目标虚拟博弈算法，在两个连续博弈中学习纳什均衡，展示了纯策略等价的适用性。

这种等价概念为这两种模型带来了多项新的理论贡献和计算方法。在理论层面，已知其他允许更复杂交互的博弈模型，例如贝叶斯博弈和扩展式博弈，可以归约为标准式博弈 [17]。因此，在连续博弈与多目标博弈之间建立桥梁，有望进一步得到具有无限多个纯策略的这类博弈与其相关多目标变体之间的等价性结果。从算法角度来看，我们预计这将有助于更高效地求解连续博弈，因为多目标标准式博弈的表格结构可能更适合计算方法。未来，我们计划在结构更复杂的大型连续博弈上评估这一方法。最后，由于等价概念并非唯一，对于任意给定的博弈，如何找到最佳的纯策略等价博弈，仍是一个值得探讨的问题。

致谢

我们感谢 Raphael Avalos 在本研究早期阶段提供的宝贵讨论。WR 和 RR 获得佛兰德斯研究基金会（FWO）资助，资助编号为 1197622N 和 1286223N；CG 则获得玛丽·居里资助，编号为 101063180。本研究获得了佛兰德斯政府通过“佛兰德斯人工智能研究计划”项目提供的资助。

参考文献

[1] Lukáš Adam、Rostislav Horčík、Tomáš Kasl 和 Tomáš Kroupa。2021。用于计算连续博弈中均衡的双重预言机算法。*AAAI*人工智能大会论文集35，6（2021年5月），5070–5077。https://doi.org/ 10 1609 35 6 16641 。/aaai.v i。David Blackwell。。《向量收益极小极大定理的一个类比》。*Pacific J. Math.* [2] 1954 6, 1 1954 1 8 10 2140 1956 6 1 ( )，–。https://doi.org/。/pjm.。。Glen E. Bredon。。《一般拓扑学》。载于拓扑与几何，Glen E. Bredon（编）。施普林格纽约，纽约州纽约市，1–62。https://doi.org/10.1007/ 978-1-4757-6848-0 1 [3] 1993 – Marek Capiński 和 Peter Ekkehard Kopp。2004。测度、积分与概率。施普林格，伦敦。https://doi.org/ 10 1007 978 1 4471 0645 6 。/- - - - Haifeng Fu。。《关于具有无限行动的标准式博弈中帕累托非支配混合策略纳什均衡的存在性》。《经济学快报》201（2021年4月 2021 109771 10 1016 2021 109771），。https://doi.org/。/j.econlet.。Joaquim Gabarró、Alina García 和 Maria Serna。。《博弈同构的复杂性》。《理论计算机科学》412（2011），6675–6695。第48期。https://doi.org/ [5] 2021 [6] 2011 10 1016 2011 07 022 。/j.tcs.。。Sam Ganzfried。。《在连续博弈中计算近似纳什均衡的算法及其在连续布洛托博弈中的应用》。*Games* 12，2（ 2021 47 10 3390 12020047 年6月），。https://doi.org/。/g I. L. Glicksberg。。《角谷不动点定理的进一步推广及其在纳什均衡点中的应用》。*Proc. Amer. Math. Soc.* 3，1 1952 170 170 10 2307（），。https://doi.org/。/ [7] 2021 [8] 1952 ,2032478 [9] 1989 Sergiu Hart 和 David Schmeidler。。《相关均衡的存在性》。《运筹学数学》14，1（1989年12月），18–25。[10] Conor F. Hayes、Roxana Rădulescu、Eugenio Bargiacchi、Johan Källström、Matthew Macfarlane、Mathieu Reymond、Timothy Verstraeten、Luisa M. Zintgraf、Richard Dazeley、Fredrik Heintz、Enda Howley、Athirai A. Irissappane、Patrick Mannion、Ann Nowé、Gabriel Ramos、Marcello Restelli、Peter Vamplew 和 Diederik M. Roijers。2022。多目标强化学习与规划实用指南。《自主智能体与多智能体系统》36，1（2022年4月），26。https://doi.org/10.1007/s10458-022-09552-y [11] Joseph Lee Heyman。2019。论策略等价博弈的计算。博士论文。俄亥俄州立大学。[12] Yu-Guan Hsieh、Kimon Antonakopoulos 和 Panayotis Mertikopoulos。2021。连续博弈中的自适应学习：最优遗憾界与收敛至纳什均衡。载于第三十四届学习理论会议论文集（《机器学习研究论文集》，第*134*卷），Mikhail Belkin 和 Samory Kpotufe（编）。PMLR，2388–2422。[13] Anisse Ismaili。2018。多目标博弈中的存在性、混合、计算与效率。载于*PRIMA 2018*：多智能体原理与实践系统，Tim Miller、Nir Oren、Yuko Sakurai、Itsuki Noda、Bastin Tony Roy Savarimuthu 和 Tran Cao Son（编）。Springer International Publishing，沙姆，210–225。[14] Kenneth L. Judd、Philipp Renner 和 Karl Schmedders。2012。求解连续策略博弈中的所有纯策略均衡。《数量经济学》3，2（2012年7月），289–331。https://doi.org/10.3982/QE165[15] Alistair Letcher、David Balduzzi、Sébastien Racanière、James Martens、Jakob Foerster、Karl Tuyls 和 Thore Graepel。2019。可微博弈机制。《机器学习研究杂志》20，84（2019），1–40。[16] Junlin Lu、Patrick Mannion 和 Karl Mason。2022。用于住宅电器调度的多目标多智能体深度强化学习方法。《*IET*智能电网》n/a，n/a（2022年5月）。https://doi.org/ . /stg . Michael Maschler、Eilon Solan 和 Shmuel Zamir。. 博弈论。剑桥大学出版社，剑桥。https://doi.org/ 10 1017 9780511794216 . /CBO Pablo [18] 2006 A. Parrilo。. 多项式博弈与平方和优化。载于第*45*届*IEEE*决策与控制会议论文集。IEEE，美国加利福尼亚州圣迭戈， 2855 2860 10 1109 2006 377261 –。https://doi.org/ . /CDC. . Roxana Rădulescu、Patrick Mannion、Diederik M. Roijers 和 Ann Nowé。2020。多目标多智能体决策：基于效用的分析与综述。《自主智能体与多智能体系统》34，1（2020年4月），10–10。https://doi.org/10.1007/s10458-019-09433-x [20] Roxana Rădulescu、Patrick Mannion、Yijie Zhang、Diederik M. Roijers 和 Ann Nowé。2020。多目标标准式博弈中均衡的基于效用的分析。《知识工程评论》35（2020），e32–e32。https: //doi.org/10.1017/S0269888920000351[21] Roxana Rădulescu、Timothy Verstraeten、Yijie Zhang、Patrick Mannion、Diederik M. Roijers 和 Ann Nowé。2022。多目标标准式博弈中的对手学习意识与建模。《神经计算与应用》 [19] 34, 3 2022 1759 1781 10 1007 00521 021 06184 3 （2022年2月），–。https://doi.org/ . /s - - - Diederik M. Roijers 和 Shimon Whiteson。. 多目标决策。载于《人工智能与机器学习综合讲座》，第34卷。Morgan and Claypool，129–129。https://doi.org/10.2200/ S00765ED1V01Y201704AIM034[23] Willem Röpke、Diederik M. Roijers、Ann Nowé 和 Roxana Rădulescu。2022。向量收益标准式博弈中的纳什均衡。《自主智能体与多智能体系统》36，2（2022年10月），53。https://doi.org/10.1007/ s10458-022-09582-6[24] Willem Röpke、Diederik M. Roijers、Ann Nowé 和 Roxana Rădulescu。2022。多目标标准式博弈中的偏好沟通。《神经计算与应用》（2022年7月）。https://doi.org/10.1007/s00521-022- 07533-6 [25] Willem Röpke、Carla Groenland、Roxana Rădulescu、Ann Nowé 和 Diederik M. Roijers。2023。弥合单目标与多目标博弈之间的差距。https://doi.org/ 10 48550 2301 05755 . /ARXIV. . Noah D. Stein、Asuman [22] 2017 Ozdaglar 和 Pablo A. Parrilo。2008。可分离与低秩连续博弈。《国际博弈论杂志》37，4（2008年12月），–。https://doi.org/ . /s - - - Noah D. Stein、Pablo A. Parrilo 和 Asuman Ozdaglar。. 连续博弈中的相关均衡：刻画与计算。《博弈与经济行为》 [26] 71, 2 2011 436 455 10 1016 2010 04 004 （3月），–。https://doi.org/ . /j.geb. . . Yijie Zhang、Roxana Rădulescu、Patrick Mannion、Diederik M. Roijers 和 Ann Nowé。2020。多目标标准式博弈中用于强化学习的对手建模。载于第*19*届自主智能体与多智能体系统国际会议论文集。自主智能体与多智能体系统国际基金会，奥克兰，新西兰，2080–2082–2080–2082。 [28]

A 测度论

在本研究中，尤其是在第3.3节和第3.4节中，我们运用了数学领域测度论的相关结果。为使本文内容相对自成一体，我们补充了一个简短章节，介绍定义和证明中用到的初等结果。关于这一主题的更深入论述，请参阅 Capiński and Kopp 的教材 [4]。

A.1 测度

测度推广了长度和体积的概念。非正式地说，测度是一个函数，它将某个母集的子集映射为一个非负实数，对应于其"质量"。

设 X 为一个集合。该集合上的 σ 代数是由 X 的子集构成的非空集合族 F，且在补集、可数并和可数交运算下封闭。我们称有序对 (X, F) 为可测空间。若函数 $\mu : \mathrm{F} \to [0, +\infty]$ 满足以下条件，则称其为测度，

(1) $\mu(\emptyset) = 0$;
(2) $\mu\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{n} \mu(B_i)$ for all $B_i \in \mathcal{F}$ with $B_i \cap B_j = \emptyset, i \neq j$.

给定可测空间 (X, F) 和定义在 (X, F) 上的测度 μ，称 (X, F, μ) 为测度空间。

为说明这一概念，我们来看一个具体的例子：狄拉克测度。狄拉克测度为每个包含预先指定元素 x 的子集赋值 1，为其他所有子集赋值 0。该测度可用于定义狄拉克δ函数，这是物理学中的一个有用工具。形式上，对于给定的 $x \in X$，可测空间 (X, F) 上的狄拉克测度 μ_x 定义为：

$$\mu_x(B) = \begin{cases} 0, & x \notin B \\ 1, & x \in B \end{cases} \qquad \text{for all } B \in \mathcal{F}$$

A.2 博雷尔概率测度

粗略地说，拓扑空间是一个集合，以及一族被指定为开集的子集。博雷尔集是拓扑空间中任何能够通过对开集取可数并、可数交和相对补集而构成的集合。X 上的所有博雷尔集构成一个 σ- 代数，称为博雷尔 σ- 代数；它是包含 X 中所有开集的最小 σ-代数。因此，给定集合 X 及其上的博雷尔 σ-代数 F，我们可以定义测度空间 (X, F)。该空间上的测度 μ 称为博雷尔测度。

设 (X, F) 是一个测度空间，μ 是该空间上的一个测度。若 $\mu(X) = 1$，则 μ 称为概率测度。自然地，如果 F 是博雷尔 σ-代数，那么 μ 也称为博雷尔概率测度。

在连续博弈中，行动集合是一个非空的紧致度量空间。因此，考虑混合策略时，不能简单地为每个行动分别指定一个概率。相反，我们需要在行动空间的开集上定义测度，以赋予概率。由于博雷尔 σ-代数是包含行动空间中所有开集的最小 σ-代数，因此通常将混合策略定义为该空间上的博雷尔概率测度。所有混合策略的集合也自然定义为所有博雷尔概率测度的集合。

A.3 推前测度

拓扑空间之间的同胚 $\varphi : X \to Y$ 是一个连续双射，且其逆映射连续；特别地，如果 $B \subseteq Y$，$A \subseteq X$ 是开集，那么 $\varphi^{-1}(B)$ 和 $\varphi(A)$ 也是开集。给定 X 上的博雷尔测度 μ，定义在 Y 上的推前测度 $\varphi_*(\mu)$ 为

$$\varphi_*(\mu)(B) = \mu\left(\varphi^{-1}(B)\right)$$

对于 Borel σ-代数 F_Y 中的所有集合 B，可以验证这确实在 Y 上定义了一个博雷尔测度。我们需要以下众所周知的结果，为完整起见，我们给出证明。

引理 A.1。设 X、Y 为拓扑空间，其博雷尔 σ-代数分别为 $\mathcal{F}_X$、$\mathcal{F}_Y$，博雷尔概率测度集合分别为 $\mathcal{B}(X)$、$\mathcal{B}(Y)$。若 $\varphi : X \to Y$ 是一个同胚，则映射

$$\psi : \mathcal{B}(X) \to \mathcal{B}(Y) : \mu \mapsto \varphi_*(\mu)$$

是一个双射。

证明。首先注意到

$$\{\varphi^{-1}(B) : B \subseteq Y \text{ open}\} = \{A : A \subseteq X \text{ open}\} \tag{26}$$

因为 φ 是一个同胚。因此，上述集合生成 F_X，而 X 上的博雷尔测度由其在该集合上的取值唯一确定。

为证明 ψ 是单射，令 μ，$\mu' \in \mathrm{B}(X)$。那么

$$\begin{aligned} &\psi(\mu) = \psi(\mu') \\ \implies &\varphi_*(\mu) = \varphi_*(\mu') \\ \implies &\mu\left(\varphi^{-1}(B)\right) = \mu'_i\left(\varphi^{-1}(B)\right) \forall B \subseteq Y \text{ open} \\ \implies &\mu = \mu'. \end{aligned}$$

最后一步再次利用了 (26) 是 F_X 的生成集这一事实，这意味着在该集合上取值相同的任意两个博雷尔测度必定相等。

为证明 ψ 也是满射，令 $\mu_Y \in \mathrm{B}(Y)$，我们只需再次在生成集 (26) 上定义 $\mu_X \in \mathrm{B}(X)$：

$$\mu_X(\varphi^{-1}(B)) = \mu_Y(B), \quad \forall B \subseteq Y \text{ open}.$$

根据构造，$\psi(\mu_X) = \mu_Y$，因为对于所有开集 $B \subseteq Y$，

$$\psi(\mu_X)(B) = \mu_X(\varphi^{-1}(B)) = \mu_Y(B),$$

因此，由于这些测度在 F_Y 的一个生成集上取值相同，它们在 F_Y 上也取值相同。□

B 恒等博弈

本节将对引理 3.1 给出形式化证明。下面重述该引理。

引理 3.1（恒等博弈）。对于任意有限玩家集合以及有限的纯策略集合，都存在一组收益函数 $\boldsymbol{p}$，使得对于每个玩家 i，$\boldsymbol{p}_i(\delta) = \delta$。

证明。我们证明，对于一个有限玩家集合 N，且每个玩家 $i \in N$ 都有一个有限纯策略集合 $A_i \in \mathrm{A}$，总能构造出一个恒等博弈。

我们首先说明恒等博弈中策略的描述方式 s_i，因为它决定了所需的收益。设 $m_i := |A_i|$。每位玩家的策略都是一个长度为 $|A_i|$ 的概率向量，

$$\forall i \in N : \delta_i = \left(P\left(a_{i,1}|\delta_i\right), \ldots, P\left(a_{i,m_i}|\delta_i\right)\right). \tag{27}$$

设 s 是一个联合策略，其中每位玩家 i 都有一个个体策略 δ_i。s 的描述是按升序排列的所有个体策略 δ_i 的串联。在每位玩家都拥有有限数量纯策略的博弈中，δ 必然具有有限的描述长度。具体而言，由于每个个体策略的长度为 m_i，而联合策略 δ 将所有此类策略合并到一个向量中，因此我们得到

$$|\delta| = \sum_{i=1}^{n} m_i. \tag{28}$$

我们用 $\delta_{i,l}$ 表示玩家 i 在联合策略 δ 中采取行动 $a_{i,l}$ 的概率，因此 $\delta_{i,l} = P(a_{i,l}|\delta)$。

恒等博弈确保对所有玩家 i, $\boldsymbol{p}_i(\delta) = \delta$。因此，我们只需定义一个由所有玩家共享的收益函数。此外，注意到每个收益向量的长度等于联合策略的长度，

$$\forall i \in N, \forall a \in \mathcal{A} : \boldsymbol{p}_i(a) \in \mathbb{R}^{|\delta|}. \tag{29}$$

纯联合策略 a 是联合策略 δ 的一种特殊情况，其中每个玩家 i 都确定性地采取行动 $a_i \in A_i$。当我们想要指定完整的联合策略记号时，将其记为 δ_a。最后补充一点记号说明：我们将 $\delta_i(a_i)$ 定义为在 δ_i 下采取 $a_i \in A_i$ 的概率。

现在我们构造恒等博弈的收益。令纯策略的有限收益集合等于联合策略向量，即

$$\boldsymbol{p}_i(a) = \delta_a \tag{30}$$

$$= \left(P\left(a_{1,1}|\delta_a\right), \cdots, P\left(a_{n,m_n}|\delta_a\right)\right). \tag{31}$$

我们证明这确保了 $\boldsymbol{p}_i(\delta) = \delta$。混合联合策略 δ 的期望收益向量计算如下（见公式 2），

$$\boldsymbol{p}_i(\delta) = \sum_{a \in \mathcal{A}} \boldsymbol{p}_i(a) \prod_{j=1}^{n} \delta_j(a_j).$$

此外，由于策略彼此独立，

$$P(a|\delta) = \prod_{j=1}^{n} \delta_j(a_j).$$

因此，根据全概率公式，我们可以得出，

$$\boldsymbol{p}_i(\delta) = \sum_{a \in \mathcal{A}} \boldsymbol{p}_i(a) \prod_{j=1}^{n} \delta_j(a_j) \tag{32}$$

$$= \sum_{a \in \mathcal{A}} \left(P\left(a_{1,1}|a\right), \cdots, P\left(a_{n,m_n}|a\right)\right) P(a|\delta) \tag{33}$$

$$= \sum_{a \in \mathcal{A}} \left(P\left(a_{1,1}|a \cap \delta\right), \cdots, P\left(a_{n,m_n}|a \cap \delta\right)\right) P(a|\delta) \tag{34}$$

$$= \sum_{a \in \mathcal{A}} \left(P\left(a_{1,1}|a \cap \delta\right) P(a|\delta), \cdots, P\left(a_{n,m_n}|a \cap \delta\right) P(a|\delta)\right) \tag{35}$$

$$= \left(P\left(a_{1,1}|\delta\right), \cdots, P\left(a_{n,m_n}|\delta\right)\right) \tag{36}$$

$$= \delta. \tag{37}$$

请注意，式（34）成立，因为在联合行动 a 下，执行任意行动 $a_{i,\ l}$ 的概率都有唯一确定的定义，且独立于另一联合策略 δ。□

C 纯策略等价

我们提供了将一种类别的博弈转换为另一种类别的博弈的明确构造方法，并证明这类构造不一定是唯一的。此外，我们介绍了一种在连续博弈与多目标标准式博弈之间构造策略双射的标准方法。最后，我们简要讨论了这种方法可能存在的缺陷，以及未来的研究如何应对这些挑战。

C.1 构造等价博弈

第3.2节中给出的证明可以形式化，以计算纯策略等价博弈。我们首先展示如何从多目标标准式博弈出发，利用算法 2 构造一个纯策略等价的连续博弈。注意，由于玩家双射和策略双射都是恒等函数，因此可以高效地完成该构造。

算法 2 从多目标标准式博弈构造连续博弈。

输入： 一个多目标标准式博弈 $G_m = (N_m, \mathrm{A}, \boldsymbol{p})$ 和效用函数 u
输出： 一个连续博弈 $G_c = (N_c, \mathrm{S}, v)$

$N_c \leftarrow N_m$
$\mathcal{S} \leftarrow \Delta^{k_1} \times \cdots \times \Delta^{k_n}$
$v \leftarrow (u_1 \circ \boldsymbol{p}_1, \ldots, u_n \circ \boldsymbol{p}_n)$
$G_c \leftarrow (N_c, \mathcal{S}, v)$
return G_c

从连续博弈构造多目标标准式博弈的方法如算法 3 所示。这里，我们明确要求给定策略双射 φ_i，将策略集映射到单形，并用于构造效用函数。随后，对于给定策略，计算其在多目标标准式博弈中的效用的效率取决于计算策略双射的逆的效率。

算法 3 从连续博弈构造多目标标准式博弈。

输入： 一个连续博弈 $G_c = (N_c, \mathrm{S}, v)$ 和同胚 $\varphi_i : S_i \to \Delta^{k_i}$
输出： 一个多目标标准式博弈 $G_m = (N_m, \mathrm{A}, \boldsymbol{p})$ 和效用函数 u

$N_m \leftarrow N_c$
$\mathrm{A} \leftarrow [k_1 + 1] \times \cdots \times [k_n + 1]$
$\boldsymbol{p} \leftarrow \mathrm{IdentityPayoffs}(N_m, \mathrm{A})$
$G_m \leftarrow (N_m, \mathrm{A}, \boldsymbol{p})$
$u \leftarrow (v_1 \circ \varphi_1^{-1} \ldots, \ v_n \circ \varphi_n^{-1})$
返回 $G \quad m, u$

请注意，$[k_i + 1] = \{1, \ldots, k_i + 1\}$，其中 k_i 是与玩家 i 的策略集同胚的单纯形中的顶点数。函数 IdentityPayoffs 返回给定玩家基集和联合行动集的恒等博弈收益。

最后，为备注1提供证明。非正式地说，我们指出，可以构造具有不同收益和效用函数、且都与给定连续博弈纯策略等价的多目标标准式博弈，反之亦然。尽管这些博弈仍具有相同的结构，但不同的收益和效用函数可能使某些博弈比其他博弈在计算上更简单。识别与

可高效求解的博弈策略等价的博弈，在其他场景中也已得到成功研究 [11]。

备注1。 一个连续博弈可能对应多个纯策略等价的多目标标准式博弈，反之亦然。

	A	B
A	(0, 0); (0, 0)	(0, 3); (0, 3)
B	(3, 0); (3, 0)	(3,3); (3, 3)

图 6：备注1的证明中使用的多目标标准式博弈。

证明。令 G_c 为一个连续博弈，其具有 $S_1 = S_2 = [0,3]$ 和效用函数 $v_1(s_1, s_2) = v_2(s_1, s_2) = s_1 + s_2$。注意，策略空间本身已经是单形，但不是概率单纯形。为构造一个多目标标准式博弈，我们可以对两个玩家使用算法 3 和策略双射 $\varphi_i(s) = \frac{s}{3}$。由此得到的纯策略等价多目标正规形式博弈称为 G_m，其收益与图 1 所示相同，效用函数为 $u_i(p_1, p_2, p_3, p_4) = 3\cdot p_2 + 3\cdot p_4$。现在考虑第二个多目标标准式博弈 $G_{\mathcal{M}}$，如图 6 所示，其效用函数为 $u_i(p1, p2) = p1 + p2$，适用于两个玩家。显然，该博弈也与 G_c 纯策略等价，因为它直接表示了单形策略空间。因此，G_c 同时与 G_m 和 $G_{\mathcal{M}}$ 纯策略等价。

对于相反的方向，我们可以将算法 2 应用于图 6 所示的博弈，从而得到一个连续博弈 $G_{\mathcal{C}}$，其策略空间为 $S_i = [0,1]$，且两位玩家的效用函数为 $v_i = u_i \circ \boldsymbol{p}_i$。因此，$G_{\mathcal{M}}$ 与 G_c 和 $G_{\mathcal{C}}$ 均为纯策略等价。□

C.2 构造策略双射

要使用算法 3 从连续博弈构造多目标标准式博弈，必须为每位玩家 i 提供一个策略双射 $\varphi_i\colon S_i \to \Delta^{k_i}$。我们提出一种直接但可能效率不高的方法来构造此类函数。之后，我们将简要讨论其他可能更适合此任务的技术。

C.2.1 标准方法。 当没有显而易见的同胚可用时，可以先构造一个映射，将任意玩家 i 的策略空间 S_i 映射到闭单位球 $B \subset \mathbb{R}^{k_i}$。随后，我们可以构造一个从 $B \subset \mathbb{R}^{k_i}$ 到概率单纯形 Δ^{k_1} 的同胚。将这两个映射复合起来，即可得到一个从 S_i 到概率单纯形 Δ^{k_1} 的同胚。

设 C 是 $\mathbb{R}^d$ 中的一个紧凸子集，且具有非空内部，并将其边界定义为 ∂C。定义 $f\colon \partial C \to S^{d-1}$ 如下：

$$f(x) = \frac{x}{\|x\|}. \tag{38}$$

该映射将 C 的边界点映射到（$d-1$）-球面。直观来说，对于一个二维凸紧集 C，f 会将边界 ∂C 映射到一个圆。映射 f 是一个同胚，因此存在逆映射 f^{-1}，而映射 $k\colon B^d \to C$ 定义如下：

$$k(x) = \begin{cases} \|x\| f^{-1}\left(\frac{x}{\|x\|}\right) & x \neq 0, \\ 0 & x = 0, \end{cases} \tag{39}$$

也是一个同胚（例如参见 [3]）。请注意，该构造假设原点位于 C 的内部，这始终可以通过平移实现。

为了完成整个构造，我们还指定逆函数 $f^{-1}\colon S^{d-1} \to \partial C$ 和 $k^{-1}\colon C \to B^d$。首先，注意到 x 是穿过 y 的射线与 ∂C 的交点，即可得到 $x = f^{-1}(y)$。令 $p_A(y) := \inf\{\lambda > 0\colon y \in \lambda A\}$ 为闵可夫斯基泛函。通俗地说，对于输入点 y，闵可夫斯基泛函 p_A 返回一个最小正数，通过该数缩放 A 后，y 便位于所得空间中。我们定义

$$f^{-1}(y) = \frac{y}{p_{\partial C}(y)}. \tag{40}$$

最后，$k^{-1}\colon C \to B^d$ 可以通过先计算经过 y 的射线与 ∂C 的交点，再进行缩放来构造：

$$k^{-1}(y) = \begin{cases} f\left(\frac{y}{\|y\|}\right)\frac{1}{\|y\|} & y \neq 0, \\ 0 & y = 0, \end{cases}. \tag{41}$$

如前所述，我们可以利用这些函数，将欧几里得空间 C 中的任意紧凸子集映射到闭单位球，并采用相同的过程将概率单纯形映射到单位球。将这两个映射复合起来，即可得到 C 与概率单纯形之间的完整同胚。

C.2.2 讨论。尽管所提出的方法很容易解释，但在实践中可能难以实现。这是因为计算闵可夫斯基泛函 $p_{\partial C}$ 需要在连续范围内进行搜索。一种可能的解决方案是使用二分搜索算法，以所需精度定位一个 λ。然而，这种方法的计算开销可能较大，因此我们建议尽可能探索更高效的技术。另一种选择是缓存 λ 的值，并在可行时重复使用，以避免反复进行二分搜索。

在利用纯策略等价求解连续博弈时，采用只需进行有限次数函数求值的算法可能会有所裨益，从而避免代价高昂的计算。最后，与其为策略双射寻求精确解，不如学习这类双射，这或许更有用。未来的工作中，我们计划对此进行更深入的研究。

D 混合策略等价

本节首先给出定理 3.4 的证明。为完整起见，我们在此重述该定理。

定理 3.4。 如果一个连续博弈与一个多目标标准式博弈纯策略等价，那么它们也混合策略等价。

证明。设 $G_c = (N_c, S, v)$ 是一个连续博弈，$G_m = (N_m, A, \boldsymbol{p})$ 是一个具有效用函数 u、与之纯策略等价的多目标标准式博弈。为简化记号，我们假设在 G_c 与 G_m 之间的任何映射中，都隐式应用玩家双射 $\pi\colon N_c \to N_m$，并将玩家集合统称为 N。

令 $i \in N$。集合 $\mathrm{B}(S_i)$ 是玩家 i 的策略集上的博雷尔概率测度集合，因而表示其混合策略；类似地，$\mathrm{B}(\Delta^{k_i})$ 表示 Δ^{k_i} 上的博雷尔概率测度，其中 k_i 来自纯策略等价的定义。由于 G_c 和 G_m 纯策略等价，因此存在一个连续双射函数 φ_i，其连续逆映射将纯

策略从 G_c 映射到 G_m 中的混合策略，对于每个玩家 i 均如此。根据引理 A.1，该函数

$$\psi_i : \mathcal{B}(S_i) \to \mathcal{B}(\Delta^{k_i})$$

其中 $\psi_i(\mu_i) = \varphi_{i*}(\mu_i)$ 是一个双射。设 $\psi = \psi_1 \times \cdots \times \psi_n$，则它也同样是一个双射。

还需证明，混合策略 μ 在 G_c 中的效用等于其对应的层级策略 $\psi(\mu)$ 在 G_m 中的效用。

$$\begin{aligned}
\mathbb{E}v_i(\mu) &= \int_{S_1\times\cdots\times S_n} v_i(s_1,\cdots,s_n)d\mu_1(s_1)\cdots d\mu_n(s_n) \\
&= \int_{S_1\times\cdots\times S_n} u_i\left(\mathbb{E}\boldsymbol{p}_i\left(\varphi_i\left(s_1,\cdots,s_n\right)\right)\right)d\mu_1(s_1)\cdots d\mu_n(s_n) \\
&= \int_{\Delta^{k_1}\times\cdots\times\Delta^{k_n}} u_i\left(\mathbb{E}\boldsymbol{p}_i\left(\delta_1,\cdots,\delta_n\right)\right)d\psi_1(\mu_1)(\delta_1)\cdots d\psi_n(\mu_n)(\delta_n) \\
&= \mathbb{E}u_i\left(\psi\left(\mu\right)\right).
\end{aligned}$$

□

E 纳什均衡的映射

纯策略等价的一大优势是，博弈动态保持不变。我们在定理 3.5 中对这一点进行了形式化表述：在连续博弈中，纯策略是纳什均衡，当且仅当它在与之纯策略等价的多目标标准式博弈中是混合策略纳什均衡。下面我们将对此进行形式化证明。

定理 3.5. 在连续博弈中，纯策略是纳什均衡，当且仅当它在与之纯策略等价的多目标标准式博弈中是混合策略纳什均衡。

证明。设 $G_c = (N_c, S, v)$ 是一个连续博弈，$G_m = (N_m, A, \boldsymbol{p})$ 是一个与之纯策略等价的多目标标准式博弈，其效用函数为 u。为简化记号，我们假设在 G_c 与 G_m 之间进行任何映射时，都会隐式应用玩家双射 $\pi: N_c \to N_m$，并将玩家集合统称为 N。

首先，假设 s^* 是 G_c 中的纯策略纳什均衡。令 φ_i 将 G_c 中玩家 i 的纯策略映射为其在 G_m 中的混合策略。根据定义 3.1，我们知道，G_c 与 G_m 之间的纯策略等价性确保了

$$\forall s \in \mathcal{S} : v_i(s) = u_i\left(\boldsymbol{p}_i\left(\varphi\left(s\right)\right)\right). \tag{42}$$

此外，由于 s^* 是纳什均衡，我们知道，

$$\forall i \in N, \forall s_i \in S_i : v_i\left(s_i^*, s_{-i}^*\right) \geq v_i\left(s_i, s_{-i}^*\right). \tag{43}$$

将这些部分结合起来，我们得到 $\forall i \in N,\ \forall s_i \in S_i$，

$$v_i(s_i^*, s_{-i}^*) \geq v_i(s_i, s_{-i}^*) \tag{44}$$

$$\implies u_i\left(\boldsymbol{p}_i\left(\varphi\left(s_i^*, s_{-i}^*\right)\right)\right) \geq u_i\left(\boldsymbol{p}_i\left(\varphi\left(s_i, s_{-i}^*\right)\right)\right) \tag{45}$$

因此，s^* 的像是 G_m 中的纳什均衡。

现在假设 δ^* 是 G_m 中的一个混合策略纳什均衡。与之前类似，我们知道存在一个双射函数 φ_i，它将 G_c 中的纯策略映射到 G_m 中的混合策略，并且

$$\forall \delta \in \Delta : u_i\left(\boldsymbol{p}_i\left(\delta\right)\right) = v_i\left(\varphi^{-1}\left(\delta\right)\right). \tag{46}$$

由于 δ^* 是一个纳什均衡，我们知道，

$$\forall i \in N, \forall \delta_i \in \Delta^{k_i} : u_i\left(\boldsymbol{p}_i\left(\delta_i^*, \delta_{-i}^*\right)\right) \geq u_i\left(\boldsymbol{p}_i\left(\delta_i, \delta_{-i}^*\right)\right). \tag{47}$$

再次将这些部分结合起来可得，$\forall i \in N,\ \forall \delta_i \in \Delta^{k_i}$，

$$u_i\left(\boldsymbol{p}_i\left(\delta_i^*, \delta_{-i}^*\right)\right) \geq u_i\left(\boldsymbol{p}_i\left(\delta_i, \delta_{-i}^*\right)\right) \tag{48}$$

$$\implies v_i\left(\varphi^{-1}\left(\delta_i^*, \delta_{-i}^*\right)\right) \geq v_i\left(\varphi^{-1}\left(\delta_i, \delta_{-i}^*\right)\right). \tag{49}$$

请注意，最后一个不等式表明，G_c 中的任何玩家都不存在能够提高其效用的纯策略偏离。然而，由于期望的线性性，这也意味着不存在能够提高其效用的混合策略偏离。因此，δ^* 的像是 G_c 中的纯策略纳什均衡。

□

最后，我们给出定理 3.6 的形式化证明。该定理指出，先前的结果可推广至连续博弈中的混合策略纳什均衡，以及纯策略等价的多目标标准式博弈中的层级纳什均衡。

定理 3.6。在连续博弈中，一个混合策略是纳什均衡，当且仅当它是纯策略等价的多目标标准式博弈中的层级纳什均衡。

证明。设 $G_c = (N_c, S, v)$ 是一个连续博弈，$G_m = (N_m, A, \boldsymbol{p})$ 是一个纯策略等价的多目标标准式博弈，其效用函数为 u。为简化记号，我们假设在 G_c 与 G_m 之间进行的任何映射中，均隐式应用玩家双射 $\pi: N_c \to N_m$，并将玩家集合简记为 N。

首先，假设 μ^* 是 G_c 中的纳什均衡。由定理 3.4 可知，对每个玩家 i，都存在一个函数 ψ_i，它将 G_c 中的混合策略双射地映射到 G_m 中的层级策略。这保证了：

$$\forall \mu \in \mathcal{B}\left(\mathcal{S}\right) : v_i\left(\mu\right) = u_i\left(\psi\left(\mu\right)\right). \tag{50}$$

此外，由于 μ^* 是一个混合策略纳什均衡，

$$\forall i \in N, \forall \mu_i \in \mathcal{B}\left(S_i\right) : v_i\left(\mu_i^*, \mu_{-i}^*\right) \geq v_i\left(\mu_i, \mu_{-i}^*\right). \tag{51}$$

将两者结合，我们得到 $\forall i \in N,\ \forall \mu_i \in \mathrm{B}(S_i)$

$$v_i\left(\mu_i^*, \mu_{-i}^*\right) \geq v_i\left(\mu_i, \mu_{-i}^*\right) \tag{52}$$

$$\implies u_i\left(\psi\left(\mu_i^*, \mu_{-i}^*\right)\right) \geq u_i\left(\psi\left(\mu_i, \mu_{-i}^*\right)\right) \tag{53}$$

因此是 μ^* 的像的一个层级纳什均衡。现在假设 μ^* 是 G_m 中的一个层级纳什均衡。我们知道，

$$\forall \mu \in \mathcal{B}(\Delta) : u_i(\mu) = v_i(\psi^{-1}(\mu)). \tag{54}$$

根据层级纳什均衡的定义，我们还可以指出，对于 μ^*，

$$\forall i \in N, \forall \mu_i \in \mathcal{B}(\Delta^{k_i}) : u_i\left(\mu_i^*, \mu_{-i}^*\right) \geq u_i\left(\mu_i, \mu_{-i}^*\right). \tag{55}$$

Finally, we get $\forall i \in N, \forall \mu_i \in \mathcal{B}(\Delta^{k_i})$,

$$u_i\left(\mu_i^*, \mu_{-i}^*\right) \geq u_i\left(\mu_i, \mu_{-i}^*\right) \tag{56}$$

$$\implies v_i\left(\psi^{-1}\left(\mu_i^*, \mu_{-i}^*\right)\right) \geq v_i\left(\psi^{-1}\left(\mu_i, \mu_{-i}^*\right)\right) \tag{57}$$

As such, the image of μ^* is a Nash equilibrium in G_c. □