让凸对偶变得困难

Eigil FjeldgrenRischel塔
林理工大学*

1 引言

凸函数的研究，尤其是对它们的优化（确切地说，是最小化），是应用数学中最重要的领域之一。凸性似乎是一种精心选择的假设：它既足够具体，能导出丰富的定理；又足够一般，能形成一套非平凡的理论（并产生大量重要的例子）。

凸优化可能因为具有“分析”而非“代数”的特征，并未得到应用范畴论学者的深入研究。一个值得注意的例外是 [4], ，它通过范畴论方法研究优化问题的分解。本文采取了不同的路径，试图定义一个以优化问题为对象的范畴，并通过范畴论方法证明有关优化的定理。作为示例，我们展示如何运用这些方法重新推导一些已有结果：一个极小极大型定理（定理 5.5），以及关于凸函数的事实（f^*）$^* = f$（其中 f 是勒让德对偶），命题 6.6。

2 凸优化

我们先回顾凸优化的一些基本对象，并借此统一术语。关于这一主题的标准参考资料，参见 [1] 。

定义 2.1（标准形式的凸优化问题）。 标准形式的凸优化问题 由以下部分组成：

1. 一个凸函数 $f_0 \colon \mathbb{R}^k \to \mathbb{R}$2. 一组凸函数 f_1, ... $f_n \colon \mathbb{R}^k \to \mathbb{R}$3. 一组仿射 函数 g_1，..... $g_m \colon \mathbb{R}^k \to \mathbb{R}$

问题是找到 $x \in \mathbb{R}^k$，使 $f_0(x)$ 最小，并满足约束条件 $f_i(x) \leq 0$，$g_i(x) = 0$。

凹优化问题是指待优化的函数 f 为凹函数，且目标是最大化的优化问题（注意，我们仍然是在凸集上进行优化）。

在下文中，我们将以略显非常规的方式，用 $\mathbb{R}_+$ 表示非负实数。 $\mathbb{R}_+^n$仅表示每个坐标均为非负的向量集合。

定义 2.2（优化问题的拉格朗日函数）。 设 $f_0 \colon \mathbb{R}^k \to \mathbb{R}$，$f_1$，..... f_n，g_1，..... g_m 为定义 2.1 中所述的标准形式凸优化问题。则该问题的拉格朗日函数是函数 $L \colon \mathbb{R}^k \times \mathbb{R}_+^n \times \mathbb{R}^m \to \mathbb{R}$，其定义为

$$L(x;\lambda,\nu) = f_0(x) + \sum_i \lambda_i f_i(x) + \sum_i \nu_i g_i(x).$$

*本研究得到 ARIA 安全保障人工智能计划的支持

A. Hadzihasanovic& J.-S. P. Lemay：应用范畴论 2025（ACT 2025）EPTCS 442，2026，第 1–14 页，
doi:10.4204/EPTCS.442.1

©Eigil Fjeldgren Rischel 本作品采用知识共享署名许可协议授权。

请注意，若 x 满足问题的约束，则 $\sup_{\lambda,\ \nu} L(x,\ \lambda,\ \nu)$ 为 $f_0(x)$；否则为 ∞。因此，我们可以将这个最小化问题视为一场零和博弈：我们选择 x，对手选择 $\lambda,\ \nu$，而我们的损失函数为 L。

自然会让人想问这个博弈中（纳什）均衡是否存在——注意，存在一个均衡（$x^*,\ \lambda^*,\ \nu^*$）意味着 $\inf_x \sup_{\lambda,\ \nu} L(x,\ \lambda,\ \nu) = \sup_{\lambda,\ \nu} \inf_x L(x,\ \lambda,\ \nu) = L(x^*,\ \lambda^*,\ \nu^*)$。

这对求解原问题非常有用。

对偶问题 是一个对函数 $\inf_x L(x,\ \lambda,\ \nu)$进行最大化的问题。这始终是一个凹问题（从下确界是 $\lambda,\ \nu$的凹函数这一意义上说）。

在凸优化领域，如果两个问题的约束条件所界定的 $\mathbb{R}^k$子集相同（且待优化的函数也相同），就称它们是等价的。但显然不能认为它们同构，因为约束函数的选择对优化理论有重要影响（例如，可能会导出不同的对偶问题）。在这里，我们采取这样的观点：拉格朗日函数才是凸优化中真正的基本对象——通过转向一个恰当的拉格朗日函数范畴，我们可以将对偶问题转化为该范畴上的真正自对偶性。

3 凸空间

定义 3.1（凸空间）。 凸空间的范畴是离散有限支撑分布单子$\Delta : \mathsf{Set} \to \mathsf{Set}$ 的代数范畴。态射称为 Δ-同态 或凸空间同态。

为避免不必要地增加记号，我们直接用 Set^Δ 表示凸空间的范畴，沿用 Eilenberg-Moore 范畴的通常记法。

凸空间已被多次研究——例如参见 [3]，其中对 Set^Δ 作了系统描述。

定义 3.2. 向量空间之间的函数称为仿射，如果它保持那些线性组合$\sum_i \lambda_i x_i$，其中 $\sum_i \lambda_i = 1$

定义 3.3. 设 X 为一个凸空间。一个凸函数，定义在 X 上，是一个函数 $f\colon X \to \mathbb{R}$，使得

$$f(\theta x + (1-\theta)x') \le \theta f(x) + (1-\theta) f(x')$$

凹函数 是指满足 $-f$ 为凸函数的函数（换言之，f 满足相反方向的不等式）。

从这个意义上说，“凸函数”一词与通常的命名习惯相冲突：通常会以函数所保持的结构来命名保持该结构的函数（因为凸函数并不保持凸结构）。遗憾的是，这种用法已经根深蒂固，难以更改。（凸函数之所以称为凸，是因为且仅因为其图像上方的区域是 $X \times \mathbb{R}$ 的一个凸子集。虽然除了惯例之外，似乎没有什么特别的理由不能互换“凸”和“凹”这两个术语。）

这个不等式

$$f(\theta x + (1-\theta)x') \le \theta f(x) + (1-\theta) f(x'),$$

在 f 为凸函数时成立，称为詹森不等式。有时，这个名称指的是该不等式的一个更强版本，例如：若 X 是取值于 f 定义域中的随机变量，则 $f(\mathbb{E}X) \le \mathbb{E} f(X)$。这些结论通常仅由 f 的凸性即可推出。

有一种自然的方法，可以将 $\mathbb{R}$ 的凸结构扩展到 $[-\infty, \infty)$ 和 $(-\infty, \infty]$，约定任何涉及无穷大的非平凡凸组合都等于该无穷大。这

也使形容词“凸”和“凹”在用于函数 $X \to (-\infty, \infty]$ 时具有了意义。例如，函数 $f\colon X \to (-\infty, \infty]$ 是凸的，当且仅当其取有限值的子集是 X 的凸子集，并且它在此集合上按通常意义是一个凸函数。

这对于扩张实直线 $\mathbb{R} = [-\infty, \infty]$不起作用，因为$\theta \cdot -\infty + (1-\theta)\infty$没有合理的解释。我们不可避免地会遇到一些取值于完整扩张实数集的函数，但我们仍希望讨论它们的凸性（或凹性）。

因此，我们约定，若函数 $f\colon X \to \mathbb{R}$ 在有意义时满足詹森不等式，则称其为凸函数；也就是说，当我们没有 $f(x) = -\infty$，$f(x') = \infty$ 或反之时。

命题 3.4. 向量空间之间的函数是仿射函数，当且仅当它是一个 Δ-同态 .

证明。 显然，仿射函数是一个 Δ-同态。设 $f\colon X \to Y$ 是一个 Δ-同态。注意，只需证明 f保持二元仿射组合 $\theta x + (1-\theta)x'$（其中 θ不一定属于$[0,1]$）。若 $\theta \in [0,1]$，则根据假设即可得证。否则设 $\theta > 1$（若不是，则利用对称性将其替换为 $1-\theta$）。于是

$$x = (1/\theta)(\theta x + (1-\theta)x') + (1 - 1/\theta)x'$$

这是一个凸组合，因此

$$f(x) = (1/\theta) f(\theta x + (1-\theta)x') + (1 - 1/\theta) f(x')$$

重新整理后，我们得到

$$\theta f(x) + (1-\theta) f(x') = f(\theta x + (1-\theta')x)$$

正如所需。 □

由命题3.4可知，我们将适当地使用术语仿射 来指称 Δ-同态，即使是在非向量空间的凸空间之间也是如此。通常不会因此产生混淆，但值得强调的是，使用这一术语并不意味着定义域对任意仿射组合封闭，例如。

凸空间既有笛卡尔积（由底层集合的积赋予逐点运算构成），也有张量积；张量积（余）表示“双同态”。这与向量空间的情形类似。然而，与向量空间不同，由于所有常值映射都是同态，投影 $X \times Y \to X$，Y都是双同态，这诱导出一个映射 $X \otimes Y \to X \times Y$。因此，同态 $X \times Y \to Z$ 是双同态的子集。

定义 3.5（单纯形）。 由有限集 $\{0, \dots n\}$ 上的自由凸空间称为n-单纯形，记作 Δ^n（编号看似不匹配，是因为 n-单纯形的维数为n）。注意， Δ^n 中的元素是一个元组（s_i）$_{i=0,\dots,n}$，因此 $\sum_i s_i = 1$ 且 $s_i \geq 0$。特别地，$\Delta^1 \cong [0,1]$。

定义 3.6（拓扑凸空间）。 拓扑凸空间 是配备了拓扑的凸空间 X，使得任意仿射映射 $\Delta^n \to X$ 都是连续的（其中 $\Delta^n \subseteq \mathbb{R}^{n+1}$ 取子空间拓扑）。

4 极小极大问题范畴

定义 4.1（极小极大问题）。 一个 极小极大问题 是一个三元组（X，Y，L），其中 X，Y是凸空间，且 $L\colon X \times Y \to \mathbb{R}$ 是一个函数，它

1. 在 X 中逐点 凸 ——对每个 y，给定 $x_1,\ x_2 \in X,\ \theta \in [0,1]$,

$$L(\theta x_1 + (1-\theta)x_2, y) \le \theta L(x_1, y) + (1-\theta)L(x_2, y)$$

2. 逐点凹 Y——对于每个 x，给定 $y_1,\ y_2 \in Y,\ \theta \in [0,1]$,

$$L(x, \theta y_1 + (1-\theta)y_2) \ge \theta L(x, y_1) + (1-\theta)L(x, y_2)$$

极小极大问题的态射 $(X,\ Y,\ L) \to (X',\ Y',\ L')$ 是一对函数 $\phi^+ \colon X \to X'$ 和 $\phi^- \colon Y' \to Y$，使得 $L(x,\ \phi^-(y')) \ge L'(\phi(x),\ y')$

我们将看到，在这个范畴上进行的各种构造，从范畴论的角度来看既自然又良好，这些构造体现了凸优化理论中的相关构造。

1. Minmax 在 $\mathrm{Set}^{\Delta} \times \mathrm{Set}^{\Delta,\ \mathrm{op}}$ 上是双纤维化的，而笛卡尔提升和余笛卡尔提升则体现了对原始变量进行最小化或对对偶变量进行最大化的操作。2. 强对偶 性质意味着某个特定图表具有局部 Beck-Chevalley 性质。3. 与此相关，与 L 对应的博弈存在纳什均衡，意味着存在某个态射。这一结论蕴含强对偶的事实，可以纯粹通过范畴论方法推导出来。

命题 *4.2.* 设 (X, A, L) 是一个极小极大问题。假设 A 是向量空间 V 的凸子空间，并且对于每个 $x,\ L(x, -) : A \to \mathbb{R}$ 是仿射的。假设 A 包含 V 的一个开子集。那么存在唯一的凸函数 $f \colon X \to \mathbb{R}, g \colon X \to V^*,$ 使得 $L(x, a) = f(x) + \langle g(x),\ a \rangle$。

注意，关于 A 仿射的极小极大问题与标准形式凸优化问题非常相似，主要区别在于，A 中允许的点集可能受到其他方式的约束，而不一定是要求某些坐标非负。

为简洁起见，我们省略证明，但请注意，关键在于，A 上的仿射函数总是可以唯一延拓到 V，且该延拓总可表示为一个线性函数加上一个常数。

定义 4.3（原始与对偶优化问题）。设 $L \colon X \times Y \to \mathbb{R}$ 是一个极小极大问题。与 L 相关联的原始优化问题 是函数

$$L^+(-) = \sup_y L(-, y) : X \to \mathbb{R}$$

（该问题是最小化此函数）。

对偶优化问题是函数 $L^-(-) = \inf_x L(x, -) \colon Y \to \mathbb{R}$

定义 4.4（对偶极小极大问题）。设 $L = (X, Y, L)$ 为一个极小极大问题。然后令 L^* 表示由 $(Y, X, L^*(y, x) = -L(x, y))$ 给出的对偶问题。

若 $\phi = (\phi^+, \phi^-) \colon L \to L'$ 是极小极大问题的态射，则 $\phi^* = (\phi^-, \phi^+) \colon L'^* \to L^*$ 仍是另一方向上的态射。该指派使 $(-)^*$ 成为极小极大问题范畴上的自逆函子

我们将经常利用这种对偶来简化证明，例如先对正向证明某个结论，再以“由对偶”论证其对反向同样成立。

定义 4.5（反向与正向态射）。称 $\mathrm{Set}^{\Delta} \times \mathrm{Set}^{\Delta,\mathrm{op}}$ 中的态射 $\phi = (\phi^+, \phi^-)$ 为正向的，若 ϕ^- 是同构；称其为反向的，若 ϕ^+ 是同构。

令 F 表示正向态射的集合， B 表示反向态射的集合。那么显然，(F,B) 构成一个正交分解系统——事实上，(F,B) 和 (B,F) 也都构成正交分解系统。

如果一个态射在 Minmax 中被视为 $\mathrm{Set}^{\Delta}\times\mathrm{Set}^{\Delta,\,\mathrm{op}}$中的态射时是正向的或反向的，我们就称它在 Minmax 中分别是正向的或反向的，并沿用记号 F， B 表示这些态射子类。

引理 4.6. 设 X,Y 是凸空间，且 $A\subset X\times Y$ 是一个凸子空间。设 $f\colon A\to\mathbb{R}$ 是一个凸函数。那么 $x\mapsto\inf_{y:(x,y)\in A} f(x,y)$仍然是凸的。

证明。 给定 $\theta\in[0,1]$， x， $x'\in X$，并考虑:

$$\inf_{y:(\theta x+(1-\theta)x',y)\in A} f(\theta x+(1-\theta x'),y).$$

因为如果 $(x,\ y),\ (x',\ y')\in A$那么 $(\theta x+(1-\theta)x',\ \theta y+(1-\theta)y')\in A$，我们有它小于 :

$$\leq \inf_{y,y':(x,y),(x',y')\in A} f(\theta x+(1-\theta)x',\theta y+(1-\theta)y'),$$

因为在后者中，我们是在一个更小的 f 集合上取下确界

应用凸性，得到

$$\leq \inf_{y,y':(x,y),(x,y')\in A} \theta f(x,y)+(1-\theta)f(x',y')$$

$$\leq \theta \inf_{y:(x,y)\in A} f(x,y)+(1-\theta)\inf_{y':(x',y')\in A} f(x',y')$$

这正是所需的不等式。 □

命题 4.7。 遗忘函子 $\mathrm{Minmax}\to\mathrm{Set}^{\Delta}\times\mathrm{Set}^{\Delta,\,\mathrm{op}}$ 是一个双纤维化。此外，我们对反向映射和正向映射上的（余）笛卡尔态射有如下描述。

1. 正向态射 $(\phi,1_A):(X,A,L)\to(Y,A,L')$ 是笛卡尔态射，当且仅当对所有 x,a， $L(x,a)=L'(\phi(x),\ a)$ *2.* 正向态射 $(\phi,1_A):(X,A,L)\to(Y,A,L')$ 是余笛卡尔态射，当且仅当 $L'(y,a)=\inf_{\phi(x)=y} L(x,a)$ *3.* 反向态射 $(1_X,\phi):(X,A,L)\to(X,B,L')$ 是笛卡尔态射，当且仅当 $L(x,a)=\inf_{\phi(b)=a} L'(x,b)$ *4.* 反向态射 $(1_X,\phi):(X,A,L)\to(X,B,L')$ 是余笛卡尔态射，当且仅当 $L'(x,b)=L(x,\phi(b))$

证明。 注意，只需为反向与正向态射提供笛卡尔提升和余笛卡尔提升（定义 4.5），因为这类提升可以复合。因此，只需验证所给描述是否正确，因为显然它们足以计算任意此类态射上的（余）笛卡尔提升。

还要注意，由于遗忘函子是忠实的，要验证态射 ϕ 是（余）笛卡尔的，只需证明基底中的任意分解都能提升——唯一性是自动成立的。

于是，设 $\phi=(\phi,\ 1_A):(X,\ A,\ L)\to(Y,\ A,\ L')$，且 $L(x,\ a)=L'(\phi(x),\ a)$。注意， Δ-同态与凸函数的复合仍是凸函数，因此这确实是 Minmax 的一个对象

现在，设 $\psi=(\psi^-,\ \psi^+):(Z,\ B,\ K)\to(Y,\ A,\ L')$ 是某个态射，使得我们有如下分解$\psi=\phi\psi'$ 在 Set 中$^{\Delta}\times\mathrm{Set}^{\Delta,\,\mathrm{op}}$。现在的目标是证明 $\psi':(Z,\ B,\ K)\to(X,\ A,\ L)$ 是一个同态。这就是不等式

$$K(z,(\psi')^-(a))\geq L((\psi')^+(z),a)=L'((\phi\psi')^+(z),a),$$

根据假设，该不等式成立

设 ϕ如上，但假设 $L'(y,a) = \inf \phi(x) = yL(x,a)$。首先注意，根据引理 4.6，这个函数实际上在 y 上是凸的，正如所需。

给定 $\psi\colon (X,A,L) \to (Z,B,K)$，现在假设我们在基底中有一个分解 $\psi = \psi'\phi$。我们必须证明 $L'(y,(\psi')^-(b)) \geq K((\psi')^+(y), b)$，但由于 $L'(y,(\psi')^-(b)) = \inf_{\phi(x)=y} L(x,(\psi')^-(b))$，这等价于等式 $L(x,(\psi')^-(b)) \geq K((\psi')^+\phi(x), b)$，而根据假设，该等式仍然成立。

反向态射的情形由对偶性直接得出。 □

这里“真正”发生的事情是，Minmax 是一个双侧纤维化：它由如下函子的格罗滕迪克构造得到：该函子将 $\mathsf{Set}^{\Delta,\mathrm{op}} \times \mathsf{Set}^{\Delta,\mathrm{op}} \to \mathsf{Cat}$ 中的一对 X、Y 映射到极小极大问题的偏序集 $L\colon X \times Y \to \mathbb{R}$（取相反序），其中态射通过前复合作用；然后“在第一个变量上逆变、在第二个变量上协变”地应用格罗滕迪克构造。（接着注意到，前复合作用的左/右伴随分别由下确界/上确界给出，从而使其成为一个双纤维化。）不过，双侧纤维化理论通常相当复杂，我们在此不作讨论——定义和基本性质见 [5]。

还要注意，这个函子与将 Minmax 展示为拓扑的 非常接近（见 [2]）。如果去掉极小极大问题必须是凸/凹的限制，我们可以用类似的上确界公式构造所需的泛提升。问题在于，一般凹函数集合的上确界并不自动是凹函数（不过，在适当意义下，取凸集上的上确界则是凹函数）。

定义 4.8 (Conv 和 Conc)。 1. 令 Conv 为如下范畴：其对象是由一个凸空间和一个凸函数组成的二元组 $(X,\ f\colon X \to \mathbb{R})$；其态射 $\phi\colon (X,\ f) \to (Y,\ g)$ 是仿射映射，满足 $g(\phi(x)) \leq f(x)$。

2. 令 Conc 为如下范畴：其对象是由一个凸空间和一个凹函数组成的二元组 $(X,\ f\colon X \to \mathbb{R})$；其态射 $\phi\colon (X,\ f) \to (Y,\ g)$ 是仿射映射，满足 $g(\phi(x)) \geq f(x)$。

命题 4.9. *1.* 以下赋值

$$(X,Y,L) \mapsto (X,L^+),$$

$$(\phi^+,\phi^-)\colon L \to L' \mapsto \phi^+$$

定义了一个函子 $(-)^+\colon \mathsf{Minmax} \to \mathsf{Conv}$

2. 类似地，$(-)^-$ 定义了一个函子 $\mathsf{Minmax} \to \mathsf{Conc}^{\mathrm{op}}$。（以这种特殊方式书写反变函子的原因稍后便会明了。）

3. 赋值 $(X,f) \mapsto (X,-f)$ 定义了一个函子（范畴同构）$\mathsf{Conc} \to \mathsf{Conv}$，反之亦然。于是 $L^- = -(L^*)^+$

4. 赋值 $(X,f) \mapsto (X,*,f)$ 定义了一个全忠实函子 $\mathsf{Conv} \to \mathsf{Minmax}$，其本质像由满足 Y 为单元素集的那些元组 (X,Y,L) 构成。

5. 类似地，$(Y,f) \mapsto (*,Y,f)$ 定义了一个全忠实函子 $\mathsf{Conc}^{\mathrm{op}} \to \mathsf{Minmax}$

6. 我们将不严格遵循记号约定，把Conv 和Conc 与它们在这些包含下的像等同起来——例如，L^+ 将被视为Minmax的一个对象。

7. $(-)^+$ 是Conv的包含的右伴随，而 $(-)^-$（视为一个函子 $\mathsf{Minmax} \to \mathsf{Conc}^{\mathrm{op}}$）是 $\mathsf{Conc}^{\mathrm{op}}$ 的包含的左伴随

8. 利用这些对应关系，我们有 $(-)^- = (((-)^*)^+)^*$

请注意，如果 $\phi = (\phi^+, \phi^-)\colon\ L \to L'$ 是 Minmax 的一个态射，那么记号 ϕ^+ 的两种含义是一致的，ϕ^- 也同样如此。

另请注意，自反子范畴 $\mathsf{Conc}^{\mathrm{op}} \subseteq$Minmax 是相对于正向态射的局部子范畴——当且仅当 ϕ^- 是同构时，态射才是正向态射（按定义），而单位 $L \to L^-$ 是以 L 为定义域的终对象正向态射。$\mathsf{Conv} \subseteq$ Minmax 也有对偶的陈述（它是相对于反向态射类的余局部化）。

定义 4.10 （极小极大问题上的幺半结构）。 极小极大问题上存在一种幺半结构，由 $(L \otimes L') = (X \otimes X',\ Y \otimes Y',\ (x,\ x',\ y,\ y') \mapsto L(x,\ y) + L(x',\ y'))$。 其单位为 $(*,\ *,\ 0)$。

一个态（即态射 $I \to (L,\ X,\ A)$）是一个满足条件的点 x_0，使得对所有 y，$L(x_0,\ y) \leq 0$。更有意思的是，求 $L \otimes L^*$ 的一个态。这是一个有序对 $x \in X,\ y \in Y$，使得对所有 y'，x'，不等式 $L(x,\ y') \leq L(x',\ y)$ 成立。

注意，对于所有的 x, y，都有 $\sup_{y'} L(x, y') \geq \inf_{x} L(x', y)$，这就是 minmax 不等式（或“弱对偶”）。

因此，选择能给出一个态的 x, y，便能使该不等式取等号——这就是 minmax 博弈的解 。根据对偶性，并且由于 $(L \otimes L^*)^* \cong L \otimes L^*$，对于这样的对象，态与余态之间存在一一对应。

命题 4.11。 遗忘函子 $\mathsf{Minmax} \to \mathsf{Set}^{\Delta} \times \mathsf{Set}^{\Delta,\mathrm{op}}$ 是一个幺半纤维化，具体意义见 *[8]*，（另见 *[6]*）。它也是一个幺半余纤维化。$\mathsf{Conv}, \mathsf{Conc}^{\mathrm{op}}$ 是幺半子范畴，而 $(-)^+, (-)^-$ 是关于限制的幺半积的强幺半函子。

5 强对偶

命题 5.1 （弱对偶）。 设 L 是一个极小极大问题。那么

$$\inf_x \sup_y L(x, y) = (L^+)^- \geq (L^-)^+ = \sup_x \inf_y L(x, y),$$

这里我们滥用记号，将极小极大问题 $*, *, r$ 与数值 $r(*, *)$ 等同起来

这在直观上显而易见，也可以通过考察 $(-)^+, (-)^-$ 的伴随性质得出。

定义 5.2 （强对偶）.设 $L = (X, A, L)$ 是一个极小极大问题。根据命题 5.1，存在一个态射 $(L^+)^- \to (L^-)^+$。如果 L 满足强对偶，则称其满足强对偶，即它是一个同构。（注意，这实际上只是一个实数不等式，而该不等式必定是等式。）

注意，对于由标准形式的凸优化问题的拉格朗日函数所产生的极小极大问题，这恰好就是强对偶的经典概念，参见 [1, 第 5.2 节。3]

命题 5.3. 设 L 是一个极小极大问题。假设存在 $\phi\colon I \to L \otimes L^*$。那么强对偶成立，即 $(L^+)^- \cong (L^-)^+$。$((\phi)^+)^-$ 给出一个态射

$$I = (I^+)^- \to ((L \otimes L^*)^+)^- \cong (L^+)^- \otimes ((L^*)^+)^- \cong (L^+)^- \otimes ((L^-)^+)^*$$

这里我们使用同构 $(L^+)^* = (L^*)^-$ 及其逆，以及 $(-)^-$和 $(-)^+$的强幺半性。该态射的存在意味着 $(L^+)^- \leq (L^-)^+$，这正是我们所需态射的另一个方向。

如果极小极大问题是一个零和博弈，那么点 $I \to L \otimes L^*$ 就是该博弈的一个纳什均衡选择。

命题 5.4 .设 (X, Y, L) 是一个极小极大问题。那么存在一个典范交换图

$$\begin{array}{ccc} (X,*) & \xrightarrow{\pi_X} & (*,*) \\ \downarrow{\scriptstyle \pi_A} & & \downarrow{\scriptstyle \pi_A} \\ (X,A) & \xrightarrow{\pi_X} & (*,A) \end{array}$$

在$\mathsf{Set}^{\Delta}\times\mathsf{Set}^{\Delta,\mathrm{op}}$中，该图是一个拉回。 L满足强对偶，当且仅当此方块对 L满足局部贝克-切瓦利条件，也就是说，典范映射$\pi_{X,!}\pi_A^* L \to \pi_A^*\pi_{X,!}L$ 是一个同构。（关于贝克-切瓦利条件的更多内容，请参见 *[7]*）

证明。 回顾$\pi_{X,!}(L) = L^-$， $\pi_Y^*(L) = L^+$。因此，该断言就是说 $\inf_x \sup_a L(x,\ a) = (L^+)^- = (L^-)^+ = \sup_a \inf_x L(x,\ a)$，这正是强对偶。 □

现在，我们为一类 Minmax 问题证明一个极小极大定理。请注意，该定理关键依赖于紧致性，因此不适用于标准形式凸优化问题的拉格朗日函数。

定理 5.5 （极小极大定理）。 设（L、 X、 A） $\in \mathsf{Minmax}$。若 X、 A 都是有限维向量空间的凸紧子空间，且 L 连续，则对于 L，强对偶成立；此外，还存在一个均衡 $I \to L \otimes L^*$ 。

最初的极小极大定理由冯·诺依曼（[10]）提出，其中 X， A 都是标准单纯形 Δ^n， Δ^m， 而 f 是仿射的（而不仅仅是凸的）。此后，该定理经历了许多推广，其中一些推广将上述定理作为特殊情形（例如参见 [9]）。这里的新颖之处不在于该定理本身，而在于其证明所采用的范畴论方法。

我们的定理可以用与通常证明纳什定理（针对一般非零和博弈）非常相似的方式，从角谷不动点定理推出——不过请注意，这并非其特例，因为 X 和 A 不一定是单纯形，而且这里的收益函数只是凸函数，不一定像博弈论博弈中的收益函数那样是仿射的。

不过，我们将给出另一种证明，它更直接地利用了 Minmax 的结构。基本思路是利用紧致性将问题归约到单纯形的情形，再用归纳论证将问题归约到 $X = A = \Delta^1 = [0,1]$， 的情形，而这一情形可以通过直接的拓扑论证来证明。归纳步骤是一个纤维序列论证，其中我们利用贝克-切瓦利性质来刻画强对偶性，见命题 5.4。

定义 5.6 （可解对）。 设 X， A 是拓扑凸空间。若对于任意连续极小极大问题 $L\colon X\times A \to \mathbb{R}$，强对偶都成立，则称对 $(X,\ A)$ 是一个可解对。

命题 5.7。 数对 $([0,1],[0,1])$ （换句话说，(Δ^1,Δ^1)）是可解的。

证明。 设 $L\colon [0,1]\times[0,1] \to \mathbb{R}$是一个连续极小极大问题。假设强对偶不成立。那么，通过给 L 加上一个常数，我们可以使得

$$\sup_\theta \inf_s L(s,\theta) < 0 < \inf_s \sup_\theta L(s,\theta).$$

考虑集合 $P = \{(s,\theta) \mid L(s,\theta) > 0\}$。由于对每个 s 都必须有 $\sup_\theta L(s,\theta) > 0$， 第一投影 $P \to [0,1]$ 必须是满射。由于每个纤维都是凸的，因而是连通的，并且投影 $[0,1]\times[0,1] \to [0,1]$ 是开映射，所以 P 是连通的。作为凸空间中的开连通子集，它是路径连通的。因此，存在某条路径 $\gamma(t) \in P$， 其中 $\gamma(0) = (0,\theta_0)$， 且 $\gamma(1) = (1,\theta_1)$。换言之（将正方形想象为第一坐标沿水平方向），存在一条从立方体左侧通到右侧的路径，使得沿路径处处都有 $L(\gamma(t)) > 0$。对偶地，也存在一条从顶部通到底部的路径，使得沿该路径 L 处处严格为负。但这两条路径必定会在某处相交，这就产生了矛盾。因此， L 必定有一个态或余状态，证明完毕。 □

命题 5.8。 设 $f\colon\ E \to B$ 是紧豪斯多夫拓扑凸空间之间的仿射满射，并假设：

- (B, A) 是可解的。
- 对每一个 $b \in B$，(E_b, A) 是可解的，其中 $E_b \subseteq E$ 是纤维。

那么(E, A) 也是可解的。

证明。 首先注意到，若 $L\colon\ E \times A \to \mathbb{R}$ 是连续的，则 $f_!L = (b,a) \mapsto \inf_{e \in E_B} L(e,a)$也是连续的。对 $r \in \mathbb{R}$，考虑 $f_!L^{-1}((-\infty, r))$。这不过是像$(f \times 1_A)(L^{-1}((-\infty, r)))$，由于紧致豪斯多夫空间之间的连续满射总是开映射，故该像是开的。

另一方面，考虑 $f_!L^{-1}((r,\infty))$。设(b,a)属于该集合——则存在某个 ϵ，使得对 $e \in E_b$有 $L(e, a) > r + \epsilon$。由 L的连续性，对每个 $e \in E_b$都存在邻域 $V_e \subseteq E$和 $U_e \subset A$，使得 $e \in V_e$、 $a \in U_e$，并且 $L(V_e \times U_e) \geq r + \epsilon/2$。由紧致性，存在一个覆盖 E_b的 V_e的有限集。令 V 为这个并， U 为相应 U_e的交。则$L(V \times U) \geq r + \epsilon/2$。于是 $f(V) \times U$是(b,a)的一个开邻域，且包含在 $f_!L^{-1}((r,\infty))$中，因此该集合是开的，从而 $f_!L$ 是连续的。

现在回想，(E, A) 可解意味着下面的方块对连续的 L 满足贝克-切瓦利条件：

$$\begin{array}{ccc} (E,*) & \longrightarrow & (*,*) \\ \downarrow & & \downarrow \\ (E,A) & \longrightarrow & (*,A) \end{array}$$

现在我们可以将其作如下分解：

$$\begin{array}{ccccc} (E,*) & \longrightarrow & (B,*) & \longrightarrow & (*,*) \\ \downarrow & & \downarrow & & \downarrow \\ (E,A) & \longrightarrow & (B,A) & \longrightarrow & (*,A) \end{array}$$

根据前面的论证，以及 (B, A) 可解这一假设，这里右边的方块满足贝克-切瓦利条件。因此只需证明左边的方块也满足。对给定的 L，这意味着要证明 B 上的这两个函数相同

$$b \mapsto \inf_{e \mapsto b} \sup_a L(e,a)$$

$$b \mapsto \sup_a \inf_{e \mapsto b} L(e,a)$$

但该等式对某个给定的 b 而言，恰好就是把 L 限制到 (E_b, A) 上的强对偶，而由假设这是一个可解对，故该等式必然成立。□

推论 5.9. 若 $(X, [0,1])$ 可解，则对每个 n，(X, Δ^n) 也可解

$n = 0$ 的情形是平凡的，而 $n = 1$ 的情形由于 $\Delta^1 \cong [0,1]$ 不过就是假设本身。映射 $\Delta^n \to [0,1]$ 取出第一个坐标，其纤维除了 1 上方的纤维只是一个点之外，全都同构于 Δ^{n-1}，因此利用该命题，我们通过归纳法即可完成证明。

引理 5.10. 设 X是一个紧致拓扑凸空间。假设对所有 n，(X, Δ^n)均可解。那么对所有拓扑凸空间 A，(X, A) 均可解。

证明。反设 $L\colon X\times A\to\mathbb{R}$ 不具有强对偶性，并且不失一般性地假设 $\sup_a\inf_x L(x,a) < 0 < \inf_x\sup_a L(x,a)$。对每个 $a\in A$，令 X_a 由满足 $L(x,a)\le 0$ 的那些 $x\in X$ 构成。给定有限组 $a_0,\dots a_n$，考虑诱导映射 $a\colon \Delta^n\to A$，并将可解性应用于问题 $(a_!L, X, \Delta^n)$—— 这尤其意味着 $\inf_x\sup_i L(x,a_i)\le 0$。这意味着族 X_a 具有有限交性质，因此由紧致性可知其交集非空。但交集中的某个元素 x^* 必须满足 $\sup_a L(x^*,a)\le 0$，这就产生了矛盾。□

推论 5.11。若 X 是紧的，则对任意 A，(X,A) 均可解。

证明。由命题 5.8 及其推论应用于命题 5.7，可知对所有 n，$([0,1],\Delta^n)$均可解。由于 $[0,1]$ 是紧的，根据引理 5.10，这意味着对所有 A，$([0,1],A)$均可解。现在利用对偶性可知，对所有 X，$(X,[0,1])$均可解，这意味着(X,Δ^n)可解；再利用一次引理 5.10，即得证。□

定理 *5.5* 的证明。由推论 5.11，(X,A)这一对可解，且强对偶成立。由于X,A均为紧集，必存在 x^*，a^*，使其分别达到下确界 $\inf_x\sup_a L(x,a)$ 和上确界 $\sup_a\inf_x L(x,a)$。它们构成一个均衡。□

值得注意的是，这里用到了紧致性。回想一下，拓扑紧致性与 $\mathrm{Hom}(X,\ -)$ 保持滤余极限这一性质密切相关，该性质也称为紧致性（不过，在 Top 中实例化的这一性质实际上并不等同于拓扑紧致性）。这里我们利用紧致性，从“有限”子问题（$L,\ X,\ \Delta^n$）中态的存在性推出整个问题中态的存在性；这种用法并非上述形式（A 甚至也不是其子单纯形的余极限），不过或许可以改写证明，使这一步更具范畴论的意味。

按 n 归纳的思路受到 [11], 的启发，不过我们的证明相当不同——他们只考虑仿射博弈，因此归纳步骤完全不同（而且他们不需要我们所需的复杂 $n=1$ 基底情形）；此外，由于我们关注的不只是单纯形上的博弈，因此还需要一个额外的紧致性论证。

我们可以利用极小极大定理，推导出关于凸优化的其他一些重要结论

定理 5.12（分离超平面定理（紧致情形））。设 X、$Y\subset\mathbb{R}^k$ 是不相交的紧致凸子空间。那么存在 $v\in\mathbb{R}^k$ 和 $\alpha\in\mathbb{R}$，使得 $\langle v,x\rangle+\alpha < 0 < \langle v,y\rangle+\alpha$ 对任意$x\in X$、$y\in Y$均成立。

证明。考虑极小极大问题

$$(L, X\times Y, A=\overline{B(0,1)}\subseteq\mathbb{R}^k), L(x,y,v)=\langle v, y-x\rangle.$$

由于闭单位球是紧致的，根据极小极大定理，存在一个均衡态 $x^*,\ y^*,\ v^*$，从而满足

$$\langle v, y^*-x^*\rangle \le \langle v^*, y^*-x^*\rangle \le \langle v^*, y-x\rangle$$

由于不相交，y^*-x^* 必须非零，因此，通过适当地选择 v，显然可以使左侧的项严格为正。因此 $v^*,\ y^*-x^* =:\ \delta>0$。现在必定存在某个 $\alpha\in\mathbb{R}$，使得 $v^*,\ y^*+\alpha=-v^*,\ x^*-\alpha=\delta/2>0$。

由均衡性质可知，y^*必须在 Y上使 v^*,y 最小；类似地，x^*必须在 X 上使 v^*,x 最大。因此，对所有 x,y，我们有

$$\langle v^*,x\rangle+\alpha\le -\delta/2 < 0 < \delta/2\le\langle v^*,y\rangle+\alpha,$$

由此证明得证。 □

定理 5.13（分离超平面定理（一般情形））。设 $X, Y \subseteq \mathbb{R}^k$ 是不相交的凸子集。那么存在 v, α，使得对于所有 $x \in X, y \in Y$ 都有 $\langle v, x\rangle + \alpha \leq 0 \leq \langle v, y\rangle + \alpha$。

证明。设 $K_i, L_i, i = 1, \ldots$ 是满足以下性质的两列集合：

- 对于每个 i, K_i, L_i 都不相交。
- 对于每个 i, $K_i \subseteq K_{i+1}$
- K_i, L_i 中的每一个都是紧致凸集
- $\cup_i K_i = X, \cup_i L_i = Y$

例如，可以通过取 X和 Y与盒子 $[-i, i]^k$的交集来构造这些集合，从而得到紧致、凸且互不相交的子集，它们穷尽了 X和 Y。

现在应用定理 5.12，得到一个 $v_i \in B(0,1)$ 序列，使得 $\overline{v_i}, -$ 在 K_i 上为负，在 L_i上为正。由单位球的紧致性可知，该序列有一个密度点 v^*。现在，对于每一对 $x \in X, y \in Y$ 我们都可以找到某个 i，使得 v_i, x 与 v^*, x 的距离小于任意给定的 ϵ，对 y 也同样成立，并且满足 $x \in K_i$, $y \in L_i$。但这样一来，$v^*, y - x$ 与 $v_i, y - x$ 的距离小于 2ϵ, 而后者为正，因此 $v^*, y - x \geq 0$。

现在，对于每个 i，$v^*, -$ 在 K_i 上有一个最大值点 x_i^*，在 L_i 上有一个最小值点 y_i^*。因此，按照与定理 5.12 的证明类似的论证，存在一个非空闭区间 $[a_i, b_i]$，使得若 $\alpha \in [a_i, b_i]$，则 $v^*, -$ 在 K_i 上为 $+\alpha \leq 0$，在 L_i 上为 ≥ 0。但是，由于集合 K_i，L_i 是递增的，这列区间必定递减，因此它们的交集必定非空——于是，该交集中的任意 α 都会使 $v^*, -$ 在 X 上非正，在 Y上非负，正如所需。□

（这就是标准分离超平面定理，教科书中的介绍见 [1, 第 2.5 节。1]）。

6 勒让德变换

定义 6.1（凸共轭）。设 V 是一个（实）向量空间，$f: V \to \mathbb{R}$ 是一个函数（不一定是线性的）。那么，凸共轭 $f^*: V^* \to \mathbb{R}$ 定义为

$$f^*(\alpha) = \sup_x \alpha(x) - f(x)$$

凸共轭也称为勒让德变换或芬切尔-勒让德变换。它与凸对偶密切相关。我们将使用极小极大问题的范畴语言来证明凸共轭的以下基本性质，并在此过程中了解凸对偶所发挥的作用。请注意，我们在此处使用“强对偶”一词，比严格意义上的必要程度稍微复杂一些——通常只需直接引用分离超平面定理即可。

命题 6.2。设 $f: V \to \mathbb{R}$ 是凸的，因此 $(V, *, f)$ 是一个极小极大问题。于是我们可以构造修正的极小极大问题 $L = (V, V^*, (x, \alpha) \mapsto f(x) - \alpha(x))$——请注意，除定义域上的符号变化外，这相当于添加约束 $x = 0$。于是 $(L^*)^+ = -L^- = f^*$。请注意，此等式中星号有两种用法，彼此冲突：一种是通过交换变量得到的反向优化问题 L^*，另一种是凸共轭函数 f^*。不过，这不至于造成太大混淆。

命题 *6.3*。给定一个极小极大问题 $L=(X,Y,L)$，其中 X 是一个有限维实向量空间，令 $L|_0=(X, Y\oplus X^*, L\oplus -\langle -,-\rangle)$ 注意，当 $x=0$ 时，$(L|_0)^+(x)=\infty$；否则为 0。因此，这相当于添加约束 $x=0$。类似地，定义 $L|^0=(L^*|_0)^*=(X\oplus Y^*, Y, L\oplus\langle -,-\rangle)$。那么，勒让德变换 $f^*=((f|_0)^*)^+$（将 f 和 f^* 都视为极小极大问题，并使用包含关系 $Conv\to Minmax$）

命题 6.4。设 $f\colon X\to\mathbb{R}$ 是定义在向量空间上的连续凸函数。那么，极小极大问题 $f|_0$ 具有强对偶性

证明。注意，$\{x,t\mid f(x)\le t\}\subseteq X\oplus\mathbb{R}$ 是一个闭凸集。因此，存在一个经过 $(0,f(0))$ 的超平面，使整个集合位于同一个半空间内。这意味着存在一个非平凡的仿射方程 $A(x,t)\ge b$，只要 $t\ge f(x)$ 就满足该方程，并且 $A(0,f(0))=b$。

显然，对于某些 $\alpha_0\in X^*, a\in\mathbb{R}$，有 $A(x,t)=\alpha_0(x)-at$。如果 $a=0$，则对于所有 x，都有 $\alpha_0(x)\le b$，这是不可能的。因此，通过归一化，令 $a=1$。这意味着 $\alpha_0(x)+f(x)\ge b=f(0)$。

回顾一下，极小极大问题 $f|_0$ 由 $(X,\ X^*,\ L(x,\ \alpha)\mapsto f(x)-\alpha(x))$ 给出。强对偶意味着 $\inf_x\sup_\alpha L(x,\ \alpha)=\sup_\alpha\inf_x L(x,\ \alpha)$。我们总有不等式 $\ge$，因此只需找出一个 α^*，使得 $\inf_x\sup_\alpha L(x,\ \alpha)\le\inf_x L(x,\ \alpha^*)$

显然，对于我们的 L，有 $\inf_x\sup_\alpha L(x,\ \alpha)=f(0)$，因为除非 $x=0$，否则上确界为 ∞。另一方面，取 $\alpha^*=-\alpha_0$，由构造可知，对于所有 x，都有 $f(0)\le f(x)-\alpha^*(x)$，证明完毕。□

引理 6.5。给定一个交换方块：

$$\begin{array}{ccc}(X_1,A_1) & \longrightarrow & (Y_1,B_1)\\ \downarrow & & \downarrow\\ (X_2,A_2) & \longrightarrow & (Y_2,B_2)\end{array}$$

在 $\mathsf{Set}^\Delta\times\mathsf{Set}^{\Delta,\ \mathrm{op}}$ 中，设从 Minmax 出发的纤维化具有贝克-切瓦利性质，并令 (Z,C) 为另一对凸空间。那么这个方块

$$\begin{array}{ccc}(X_1\times Z,A_1\times C) & \longrightarrow & (Y_1\times Z,B_1\times C)\\ \downarrow & & \downarrow\\ (X_2\times Z,A_2\times C) & \longrightarrow & (Y_2\times Z,B_2\times C)\end{array}$$

也满足贝克-切瓦利条件

命题 6.6. 设 f 是一个凸函数。那么，在将有限维向量空间与其双对偶等同的识别下，$f=(f^*)^*$（其中 f^* 表示勒让德变换）。

证明。回忆一下，$f^*=((f|_0)^*)^+$，是一个极小极大问题。于是要证明的是

$$(((((f|_0)^*)^+)|_0)^*)^+=f.$$

首先使用 将 $((-)^*)^+=((-)^-)^*$，改写，并采用记号 $((-)^*|_0)^*=-|^0$，，我们可以将其改写为 a s

$$((f|_0)^-|^0)^+$$

现在注意到，将其限制在由极小极大问题组成的子范畴 Minmax_l 上，其中 X、Y、L 满足：X、Y 是实向量空间，并且这些同态由线性映射（而不仅仅是仿射映射）给出；$(-)|_0$ 和 $-|^0$ 构成自函子，而 $(-)|_0\dashv(-)|^0$。

由于 $(-)^-$ 是包含的左伴随，我们有 $(-)|_0^- \dashv -|^0$。因此，对于任意 L，都有一个规范映射，即伴随的单位，$L \to (L|_0)^-|^0$。如果 $L = (X、\ *、\ f)$ 是 Conv 中的一个元素，那么根据泛性质，该映射可分解为 $((f|_0)^-|^0)^+$。这给出了不等式 $f \geq (f^*)^*$。

（注意，即使 f 不是凸的，这个不等式实际上仍然成立，而且我们确实尚未真正用到凸性）。

注意，利用自然同一 $(X^*)^* = X$，我们有

$$(f|_0)|^0 = (X \oplus X, X^*, (x, x'; \alpha) \mapsto f(x) - \alpha(x) + \alpha(x')).$$

显然 $\inf_x \sup_\alpha f(x) - \alpha(x) + \alpha(x') = f(x')$，因为除非 $x = x'$，否则上确界为无穷。但请注意 $(f^*)^*(x') = \sup_\alpha \inf_x f(x) - \alpha(x) + \alpha(x')$

我们现在的断言是，我们可以借助强对偶交换这些极值化子。这相当于断言，局部 Beck-Chevalley 性质在 $(f|_0)|^0$ 处对该方块成立：

$$\begin{array}{ccc} (X \oplus (X^*)^*, *) & \longrightarrow & ((X^*)^*, *) \\ \downarrow & & \downarrow \\ (X \oplus (X^*)^*, X^*) & \longrightarrow & ((X^*)^*, X^*) \end{array}$$

但根据命题 6.4，强对偶在每一个如下形式的方块中成立

$$\begin{array}{ccc} (X, *) & \longrightarrow & (*, *) \\ \downarrow & & \downarrow \\ (X, X^*) & \longrightarrow & (*, X^*) \end{array}$$

根据引理 6.5，这就证明了前一个交换方形同样满足贝克-切瓦利条件，从而完成证明。□

参考文献

[1] Stephen P. Boyd & Lieven Vandenberghe（2004）：凸优化。剑桥大学出版社，英国剑桥；纽约，doi:10.1017/CBO9780511804441。[2]Eduardo J. Dubuc & Luis Español（2006）：拓扑函子作为熟悉的纤维化，doi:10.48550/arXiv.math/0611701。arXiv:math/0611701。[3]Tobias Fritz（2009）：凸空间 I：定义与例子，doi:10.48550/arXiv.0903.5522。arXiv:0903.5522。[4] Tyler Hanks、Matthew Klawonn、Evan Patterson、Matthew Hale & James Fairbanks（2024）：一阶优化的组合框架，doi:10.48550/arXiv.2403.05711。arXiv:2403.05711。[5] Fosco Loregian & Emily Riehl（2020）：纤维化的范畴论概念。数学论述 38(4)，第496–514页，doi:10.1016/j.exmath.2019.02.004。[6] Joe Moeller & Christina Vasilakopoulou（2020）：幺半群 *Grothendieck* 构造。范畴理论及其应 用 (31)，第 1159–1207 页。可访问：http://www.tac.mta.ca/tac/volumes/35/31/ 35-31abs.html[7] Duško Pavlović（1991）：范畴插值：没有直接像的下降与贝克-切瓦利条件。载于Aurelio Carboni、Maria Cristina Pedicchio & Guiseppe Rosolini（编）：范畴论，Springer，柏林、海德堡，第306–325页，doi:10.1007/BFb0084229。

[8] Mike Shulman（2009）：框架双范畴与幺半纤维化。范畴理论及其应用(18)，第650–738页。可访问：http://www.tac.mta.ca/tac/volumes/20/18/20-18.pdf。[9] Maurice Sion：关于一般极小极大定理 8(1)，第171–176页。doi:10.2140/pjm.1958.8.171。可访问：http://msp.org/pjm/1958/8-1/p14.xhtml。[10] J. V. Neumann：关于博弈论 100(1)，第295–320页。doi:10.1007/BF01448847。可访问：http://link.springer.com/10.1007/BF01448847。[11] Jonathan Weinstein：极小极大定理的两个初等证明。可访问：https://bpb-us-w2.wpmucdn.com/sites.wustl.edu/dist/d/2034/files/2022/05/minimax.pdf。