当代经济学系列丛书
Contemporary Economics Series
主编 陈昕

国家“十二五”重点图书

（第二版）

经济理论中的最优化方法

当代经济学
教学参考书系

[美] 阿维纳什·K.迪克西特 著
冯曲 吴桂英 译

格致出版
格致出版社
上海三联书店
上海人民出版社

当代经济学系列丛书
Contemporary Economics Series
主编 陈昕

（第二版）

经济理论中的最优化方法

[美] 阿维纳什·K.迪克西特 著
冯曲 吴桂英 译

当代经济学教学参考书系

格致出版社
上海三联书店
上海人民出版社

图书在版编目（C I P）数据

经济理论中的最优化方法:第2版/(美)迪克西特(Dixit, A. K.)著;冯曲,吴桂英译.—上海：格致出版社,2013
(当代经济学系列/陈昕主编.当代经济学教学参考书系)
ISBN 978-7-5432-2253-3

Ⅰ. ①经… Ⅱ. ①迪…②冯…③吴… Ⅲ. ①经济理论-最优化算法 Ⅳ. ①F0

中国版本图书馆CIP数据核字(2013)第107473号

责任编辑　钱　敏
装帧设计　敬人设计工作室
　　　　　吕敬人

[美]阿维纳什·K.迪克西特 著
经济理论中的最优化方法(第二版)
冯曲　吴桂英 译

格致出版社·上海三联书店·上海人民出版社
(200001　上海福建中路193号24层　www.ewen.cc)

格致出版

编辑部热线　021-63914988
市场部热线　021-63914081
www.hibooks.cn

世纪出版集团发行中心发行
浙江临安曙光印务有限公司印刷
2013年6月第1版
2013年6月第1次印刷
开本：787×1092　1/16
印张：11.25　插页：5　字数：165,000

ISBN 978-7-5432-2253-3/F·639
定价:28.00元

出版前言

为了全面地、系统地反映当代经济学的全貌及其进程,总结与挖掘当代经济学已有的和潜在的成果,展示当代经济学新的发展方向,我们决定出版"当代经济学系列丛书"。

"当代经济学系列丛书"是大型的、高层次的、综合性的经济学术理论丛书。它包括三个子系列:(1)当代经济学文库;(2)当代经济学译库;(3)当代经济学教学参考书系。该丛书在学科领域方面,不仅着眼于各传统经济学科的新成果,更注重经济学前沿学科、边缘学科和综合学科的新成就;在选题的采择上,广泛联系海内外学者,努力开掘学术功力深厚、思想新颖独到、作品水平拔尖的"高、新、尖"著作。"文库"力求达到中国经济学界当前的最高水平;"译库"翻译当代经济学的名人名著;"教学参考书系"则主要出版国内外著名高等院校的通用教材。

本丛书致力于推动中国经济学的现代化和国际标准化,力图在一个不太长的时期内,从研究范围、研究内容、研究方法、分析技术等方面逐步完成中国经济学从传统向现代的转轨。我们渴望经济学家们支持我们的追求,向这套丛书提供高质量的标准经济学著作,进而为提高中国经济学的水平,使之立足于世界经济学之林而共同努力。

我们和经济学家一起瞻望着中国经济学的未来。

中文版前言

我非常高兴为我这本关于最优化的书的中译本写一个简短的前言。在这本书首次问世27年、再版13年之后，其内容和阐述依旧被认为足够有趣而值得翻译成中文，这确实是一件让人感到欣慰的事。中国正在迅速崛起而成为世界上一个重要的经济力量。如果本书能够为中国经济决策的质量作出即使是非常微小的贡献，那也将是一个很大的成就。

自我写作本书以来，经济学中出现了很多新的思想。行为经济学中新的经验的和理论的发展，向关于消费者和厂商目标的某些传统假定提出了挑战。因而值得强调的是，除了内部逻辑的一致性之外，最优化的方法并不依赖于任何关于这些目标的特定的假定。这些方法只不过向我们显示了如何最好地追求我们的目标，而不论这些目标具体是什么。新的应用领域已经出现了，而其中不对称信息和金融经济学领域或许最为重要。如果我是在今天首次写作本书，那么除了书中目前提到的关于这两个领域的少量例题之外，我还将为这两个领域提供更深入的内容。但是书中的最优化方法对于新的或成长中的领域依旧有用。

我要特别感谢冯曲先生和吴桂英小姐为翻译本书所作的工作。遗憾的是，我对中文一无所知，所以不能进行直接的判断。但是我确实知道他们是多么仔细地阅读了本书的英文版本。特别是冯曲先生，他发现了若干排印上和确实存在的错误，并使我注意到这些错误。因而该中译本应该会是对原书的一个实际的改进，而且也会有助于英文版下一次印刷的改进。

Avinash K. Dixit

Princeton 大学经济系

2003年12月9日

前　言

最优地利用稀缺资源，或者说，在约束条件下求最大化的问题，是经济学的中心主题。但约束条件下最大化的数学方法往往作为数学的一个分支被传授给经济学系的学生们，并且它的经济学应用也是分开进行的。也许，从一开始就将数学与经济学联系起来统一论述，能给学生提供更快更深刻的理解。此书就旨在给出这样的阐述。但我所强调的是经济学上的直觉而并非数学上的严谨。数学定理证明的构造是为了说明经济学上感兴趣的观点和便于经济学上的应用。我也挑选了一些说明性的例题和习题，而之所以挑选这些是因为它们在经济学上是有趣和有用的。

本书的第一版出版于 1976 年。时隔多年，本书的中心目标依然不变，但是主题有了很大的变化。因而我对原书作了大量的修改。例如，有关不确定性以及金融和不对称信息等主题，现在已是必不可少。因此，我新增一章专门来讨论这些问题，同时也扩充了关于动态规划的那章来讨论不确定性。

大部分章节已完全重写，并增加了很多新的例子和习题。有一个创新特别值得提及。在我写第一版时，初级和中级微观经济学中主要的阐述模式是几何模式，即基于预算线和无差异曲线之间，或者成本线和等产量线之间的相切。而今，相切这一老套做法似乎不再流行。因而我使用了一个更加简单，在经济学上更加直观的出发点，即通过“套利”操作对无成本改进的搜寻。这一做法允许统一地处理相切和角点最优解，并且其直觉更容易扩展到有时间和不确定性等的情形。

自本书第一版问世至今，经济学专业学生的数学

训练已大大改进。利用这一点，我引进了一些更深一点的论题，加深的内容体现在最后三章中，并在数学附录中概述了约束最大化的中心结论——库恩—塔克定理——的证明。但是本书的主要对象依旧是大部分经济学专业的高年级本科生和硕士一年级的学生，而并非少数打算成为数理经济学家的学生们。

本书的主要用途是作为中级和高级微观经济学的一个补充教材。它也可以用作主要教材，但是应当配以其他的书本或论文，包括那些列在每章末作为"进一步阅读"的论文和书本。关于消费者和生产者行为的短期课程可以将第1—8章作为基础。第9和第10章则相互独立。因而，包括时间维度上的最优化问题但不包括不确定性的课程可以略过第9章以及第11章的部分内容。相反地，包括不确定性但是与时间无关的课程可以略过第10章和第11章。

就推理的主线而言，书中的例题都得到了充分的解答，但是某些细节被省略了。习题包含了正文中的理论的进一步发展和应用。例题和习题也是学习本书不可缺少的部分，我强烈建议读者们仔细地完成它们。

我要感谢那些对改进第一版给了我有用建议的诸多读者，尤其是 Pete Kyle。同时，我也非常感谢 Richard Quandt，他极其仔细地阅读了第二版的全部手稿并且指出了其中的几个错误。我还要感谢 Barry Nalebuff 和 Carl Shapiro，他们也阅读了本书的部分章节。Peter Kenen 的法则是说，总会有你想象不到的排印错误，我为本书中仍然可能存在的错误负责。

A. K. Dixit

Princeton

1989 年 12 月

上海市版权局著作权合同登记号:图字09-2004-131号。

目　录

1 导论

经济学已经被定义为最优地利用稀缺资源的研究，即，在约束条件下最大化的研究。被最大化的对象以及施加在选择上的约束，从一个背景变化到另一背景：家庭的消费和劳动供给，企业的生产和政府的政策。但是所有约束条件下的最大化问题有一个共同的数学结构，这一数学结构又反过来为分析这些问题提供了一个共同的经济学直觉。本书旨在概述这一数学内容并且培养这种直觉。

消费者选择的标准模型提供了一个好的起点。其基本概念为一条预算线和一组无差异曲线。预算线上的点代表了所有支付得起的两种商品的组合。而无差异曲线组则代表了目标，即要达到尽可能高的曲线。最优选择落在一条无差异曲线和预算线的相切之处。图 1.1 给出了我们熟悉的图形。

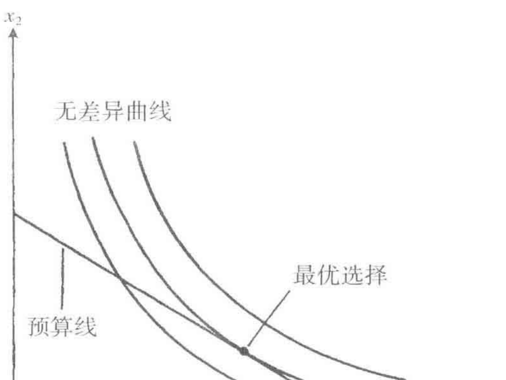

图 1.1　消费者的最优选择

在这一章中，我将运用文字和几何的方法进一步发展这个模型，但着眼于对数学方法的锐化和一般化，虽然我们在后面的章节才强调这些方法。

赋予图 1.1 中各个不同的量以一些代数上的含义是有益的。记 I 为消费者的货币收入，p_1 和 p_2 为两种商品的价格，x_1 和 x_2 为它们的数量。因而预算线，即支出恰好等于收入的线，可以用方程

$$p_1 x_1 + p_2 x_2 = I \tag{1.1}$$

来表达。

消费者对这两种商品的数量 x_1 和 x_2 的偏好用一个称为效用函数的数值尺度(numerical scale)来表示。它给商品的每一个消费束(x_1, x_2)赋予一个称为效用水平的数值 $U(x_1, x_2)$。在对任何可供选择的消费束进行比较时，最终被偏好的消费束就是得到最高效用数值的那个消费束。沿着一条无差异曲线，所有的点必须具有相同的效用数值。因而这样的曲线有一个方程

$$U(x_1, x_2) = \text{常数} \tag{1.2}$$

现在通过计算预算线和无差异曲线的斜率可以继续这一理论的阐述。由于两者相切，斜率必然相等。稍后我将继续阐述。但是让我从一个更简单更直观的方法开始。

1.1 套利方法

这一想法是让消费者从他预算的任意试验性的配置开始，考虑一个变动。如果这种变动产生了一个给他带来更高效用的消费束，那么它就将作为一个新的试验性配置而被消费者采取。一旦发现某个消费束已经无法用这种方式来改进，那么它就将成为最优配置，而且也将在实际中被消费。所以，不可能再找到一个改进将作为最优化的检验条件。

这样的变动不会造成任何额外的支出；它仅仅是某一数量的货币从购买一种商品转到另一种商品的重新配置。如果最初的配置并非最优，这样的变动可以提高消费者的效用。当消费者已作出了最优选择，达到了他的个人均衡后，这种获得效用改进而同时支出没有净增加的机会就消失了。金融市场中有与此非常相似之处。在市场均衡之外，利用相同资产在不同

市场间的价格差，参与人能够以零的净费用获得“套利”利润，而在市场均衡时就不存在这样的机会。事实上，正是人们寻求套利利润的过程导致了均衡。通过将这整个推理过程记为“套利方法”，并将由此导致的最优条件记为“无套利条件”，我将充分利用这种直觉上的相似之处。

如果商品是不可分的，这些变动必须按离散的步骤发生。然而，假定商品是完全可分的，通常是一个好的近似。因而，这些变动能够以无穷小的量进行，或者以经济学家们所说的“边际调整”进行。甚至对那些看起来不可分的商品，例如汽车或其他大宗耐用消费品，也存在质量等其他维度，允许连续的调整。无论如何，这样的边际变动是本书中我们分析的主题。

变量 x 的“一个微小(边际、无穷小量)变动”的标准符号是 $\mathrm{d}x$。它不可以被看作两个变量 d 和 x 的乘积，而本身就是一个整体 $\mathrm{d}x$。使用这种无穷小量的合理性可以被严格证明，但是大部分时候，我将以宽松而富有启发性的方式使用它们。当你对涉及使用无穷小量的陈述有所质疑，或不能确信我是否正确地使用它们时，你应该使用正确的微积分方法重新论证，从有限的变动 Δx 开始，然后趋向极限如 $\Delta x \to 0$。

首先，假定预算的最初配置对两种商品都有正的消费量 x_1 和 x_2。现在考虑一个小的套利操作，或将少量的但是数量为正的收入 $\mathrm{d}I$ 从商品 2 转到商品 1 进行边际上的再配置。从实物角度来看，这意味着多购买 $\mathrm{d}I/p_1$ 单位的商品 1，少购买 $\mathrm{d}I/p_2$ 单位的商品 2。用 MU_1 和 MU_2 表示这两种商品的边际效用。这意味着商品 1 数量上的微小变动 $\mathrm{d}x_1$ 改变了 $MU_1\,\mathrm{d}x_1$ 个单位的效用，类似地，商品 2 数量上的微小变动 $\mathrm{d}x_2$ 导致了 $MU_2\,\mathrm{d}x_2$ 个单位的效用变化。当两种商品的数量同时变动，这两种效应可以被加在一起，所以效用的变化为

$$MU_1\,\mathrm{d}x_1 + MU_2\,\mathrm{d}x_2 \tag{1.3}$$

在后面的章节中，我将使用偏导数和泰勒级数更严格地表达此式，但在此处这样简单的陈述就足够了。关键之处在于，任何一个边际调整都是如此之小，以至于在这个过程中边际效用自身的变化可以被忽略不计。当然，如果很多这样的边际调整同时进行，边际效用就会随着这一系列的变动而逐渐变化。特别地，如果 x_2 上升，和/或 x_1 下降，MU_2/MU_1 将会下降；这就是消费中的边际替代率递减原理。但是在目前，我所讨论的仅是一个边

际调整。

这样，套利操作对效用的影响就很容易计算。商品 1 的数量增加 $\mathrm{d}I/p_1$ 使得效用提高 $MU_1\mathrm{d}I/p_1$，而商品 2 数量的减少 $\mathrm{d}I/p_2$ 使得效用降低 $MU_2\mathrm{d}I/p_2$。因而效用的净增加为

$$(MU_1/p_1 - MU_2/p_2)\mathrm{d}I$$

如果这一表达式的值为正，消费者就会实行这个再配置并试图作出相同方向上的进一步的再配置。如果最初的消费束是最优的，那么这一表达式的值不可能为正。这是最优化的"无套利"标准的一部分。由于所选择的 $\mathrm{d}I$ 是正的，因而我们可以将该表达式除以 $\mathrm{d}I$ 并把这个标准写作

$$(MU_1/p_1 - MU_2/p_2) \leqslant 0 \tag{1.4}$$

现在假定这个标准没有被满足，也就是假定(1.4)式的左边大于零，那么把一些支出转到商品 1 上就是值得的。这一转移应该进行到何种程度呢？回忆起当 x_1 增加而 x_2 减少时，相对于 MU_2，MU_1 会逐渐下降，最终会到达使得(1.4)式的左边为零的那一点，并且在这个方向上的继续变动将无法再提高消费者的效用水平。

接下来考虑一个相反方向的再配置。这会改变上面所有变量的符号。如果最初的配置是最优的，我们必然会有

$$(MU_1/p_1 - MU_2/p_2) \geqslant 0 \tag{1.5}$$

如果这是错的，即如果上式的左边小于零，那么增加 x_2 而减少 x_1 的过程就一直会进行下去，直到上式的左边等于零为止。

我们可以将这两个关于最初消费束为最优的标准结合成一个：如果数量均为正的配置(x_1, x_2)是最优的，那么在这点处边际效用必须满足

$$MU_1/p_1 = MU_2/p_2 \tag{1.6}$$

这就是完整的"无套利条件"。其经济学解释为，在最优选择处消费者对于将额外一单位的货币配置给这种商品还是那种商品是无差异的。

1.2 使用微积分的相切条件

更复杂但更常用的导出相同的最优化条件的方法，是基于预算方程

(1.1)所表示的预算线和效用函数(1.2)所表示的无差异曲线的相切。将预算线方程写为

$$x_2 = (I/p_2) - x_1(p_1/p_2)$$

我们马上看到这条线的斜率的数值为(p_1/p_2)。无差异曲线的斜率是消费中的边际替代率(MRS),它等于边际效用的比率(MU_1/MU_2)。对此的一个启发性的推导如下。如果 dx_1 单位的商品 1 的边际损失恰好被 dx_2 单位的商品 2 的边际收益所补偿,那么边际替代率(MRS)就是比率 dx_2/dx_1。但是对这一收益的精确补偿可以写成以效用为单位的一个等式

$$MU_1 dx_1 = MU_2 dx_2$$

所以

$$MRS = dx_2/dx_1 = MU_1/MU_2 \tag{1.7}$$

顺便说一句,注意在这两个比例的分子和分母之间下标 1 和 2 的颠倒。这并非是排印的错误;而是这两个比例之间的关系。你可以看到通过交叉相乘就回到上面的等式了。

在最优选择处,无差异曲线的斜率(消费中的边际替代率)等于预算线的斜率(价格比)。因而

$$MU_1/MU_2 = p_1/p_2 \tag{1.8}$$

这等价于前面我们导出的最优化标准(1.6)式。但是(1.6)式所蕴涵的文字论证方法要比(1.8)式所蕴涵的几何论证方法更有优势。

1.3 角点解

套利方法在其中一种商品根本不被购买的情形下更容易被采用。由于这种情况发生在预算线的顶端,这样的最优选择通常被称为角点解。假定在最初配置中所有的收入都被花在了商品 2 上,于是 $x_1=0$,$x_2=I/p_2$。现在,在两种改进方向之中,即试图增加 x_1 或者减少 x_1,只有前者是可能的。因而,只有对应于这种检验条件的标准(1.4)式成立。由于 x_1 不能再减少,得出(1.5)式的论证也就不能进行了。对消费束$(0, I/p_2)$成立的(1.4)式的经济学解释同样简洁明了:即使是花在商品 1 上的第一个单位的

收入所带来的效用,也不足以抵过花在商品 2 上的最后一单位收入所带来的效用。

1.4 收入的边际效用

套利方法的第二个优点更为重要。回到两种商品最初都被购买的情形和最优化标准(1.6)式。现在假定消费者获得额外的数量为 dI 的收入可用于消费。他可以把它全部用于商品 1,即购买额外的(dI/p_1)单位的商品 1 并取得额外的($MU_1\,dI/p_1$)单位的效用。或者他可以把它全部用于商品 2,增加效用($MU_2\,dI/p_2$)单位。但是根据(1.6)式,这两个效用的增量必须相等。因而在边际上将无穷小量的额外收入 dI 配置到商品 1,或商品 2,抑或两种商品的任意组合,对消费者而言都是无差异的。于是我们可以把由每单位额外收入所带来的效用的增量简单地称为收入的边际效用,而无须具体说明这额外的收入是被如何花费的。将 λ 记作这一单位额外收入的边际效用。那么 dI 单位的额外收入增加了 λdI 单位的效用。将此等同于另外两种对花费在两种商品中的每一种商品上所获得的相同的效用增量的记法,我们有

$$\lambda = MU_1 / p_1 = MU_2 / p_2 \tag{1.9}$$

我们发现最优化标准(1.6)式两边相同的值具有一个非常有用的经济学解释——它是收入的边际效用。所有约束条件下最优化问题的无套利条件,都可能有一个非常类似的解释。我将会对此作出更为详细的解释,并在拉格朗日乘子或者影子价格的概念下将它统一起来。

1.5 多种商品和多个约束条件

将分析扩展到有多种商品,并且某些商品可能根本不被消费的情形是一样容易的。假定有 n 种商品,价格和数量分别为($p_1, p_2, \cdots, p_n$)和($x_1, x_2, \cdots, x_n$)。对于所有在最优时都被购买了正的数量的商品,边际效用对价格的比例必定有一个相同的值,这个值可以被解释为收入的边际效用 λ。对于所有在最优时没有被购买的商品,边际效用对价格的比例一定小于或至

多等于这个值。用符号来表示，即对于任意商品 i，

$$MU_i/p_i\begin{cases}=\lambda, x_i>0\text{ 时}\\ \leqslant\lambda, x_i=0\text{ 时}\end{cases}$$

或者

$$MU_i-\lambda p_i\begin{cases}=0, x_i>0\text{ 时}\\ \leqslant 0, x_i=0\text{ 时}\end{cases}\tag{1.10}$$

这个方法可以被扩展到允许有多个约束条件的情形。对于每个约束条件，我们需要一个单独的 λ，它可以被解释为放松那个约束条件的边际效用。我将在第 3 章中讨论这一点。

1.6 非紧的约束条件

最后，考虑一个与消费者理论并不非常相关的扩展，但这一扩展在其他应用中将是非常重要的。设想一个消费者有如此之多的收入以至于他心满意足并且无法花光所有收入，那么预算方程(1.1)应该用一个不等式来替代

$$p_1 x_1+p_2 x_2\leqslant I$$

通过定义一种新商品 x_3，我们可以很容易地把这种情形纳入前述理论的范围中。x_3 是“没有被花掉的收入”，其价格为 1 并且不产生效用(注意，我所说的并非储蓄，而是完全无用的没有被花掉的收入，储蓄会使得一个并不满足的消费者在将来消费更多的合意商品)。现在预算方程变为

$$p_1 x_1+p_2 x_2+x_3=I$$

且 $MU_3\equiv 0$。由于我们假定消费者选择了一个数量为正的 x_3，对于 $i=3$，(1.10)式给出了 $\lambda=0$。这给出了直观的含义：如果消费者甚至没有花完他所拥有的所有收入，那么收入增量所带来的边际效用应该为零。反过来，在对应于其他商品的(1.10)式中，我们可以利用这一点，获得 $MU_i=0$($i=1$ 和 2)。因此，这些商品被消费到产生零边际效用的水平，即达到了满足点。

我们从消费者的选择问题的一个简单而直观的讨论开始，仅仅计算了边际调整的“套利”收益和损失，用它们得到了最优化标准。这就给我们带

来了刻画约束条件下最优解的一个非常重要的一般方法。事实上,(1.10)式只是约束条件下最优化理论的基本结论的一种形式,即库恩一塔克定理。所扩展得到的一个已满足的消费者情形,是互补松弛性(Complementary Slackness)这一熟知的一般原理的一个例子。在以后的章节中,我将更加系统地逐步发展一般理论,利用微积分,发展和强化这里引入的那些直观的思想。

▶ 基础阅读

为了方便而有效地阅读此书,你应该熟悉初级或者中级微观经济学中的基本概念:转换曲线和无差异曲线,以文字或几何形式等进行的边际上的计算。如果你需要获得这方面的背景知识,任何好的初级经济学或者中级微观经济学教材都可充当此任。在此,我仅各提及一种。

William J. Baumol and Alan S. Blinder, *Economics: Principles and Policy*, Chicago: Harcourt Brace Jovanovich, 4th edn., 1987.

Hal R. Varian, *Intermediate Microeconomics*, New York: Norton, 1987.

所需的数学知识为大学的微积分水平。在第 2 章中你很快就会用到偏导数的知识。第 3 章会出现向量和矩阵的表示法及其简单的相加和相乘。在本书的必要之处会提出一些有关凸性的概念。第 8 章会用到正定矩阵和二次型。而第 9—11 章会用到简单的积分。

那些需要获得这些数学技能中的一部分技能的读者可以按自己的情况选择适当的数学书,或者是为经济学家们而设计的数学书,或者是解释数学的经济学书籍。这里我再次各提及一种。

Serge Lang, *Calculus of Several Variables*, Berlin: Springer-Verlag, 3rd edn., 1987, chs. 1—6, 13—16.

Alpha C. Chiang, *Fundamental Methods of Mathematical Economics*, New York: McGraw-Hill, 3rd edn., 1984, chs. 4,5,7,9.

Alasdair Smith, *A Mathematical Introduction to Economics*, Oxford, UK: Blackwell, 1982.

▶2 拉格朗日方法

2.1 问题的陈述

利用一个和第 1 章中的消费者选择模型非常接近的背景，我们开始深入讨论约束条件下最优化问题的一般理论。假定选择变量为 x_1 和 x_2。我把它们写作一个更简洁的列向量的形式，

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

最初，使用向量只是为了缩略掉对分量的罗列，而向量和矩阵的实际运算在后面章节中会逐步出现。

我们需要表示法以区别一般向量 x 和 x 的某个特定值，如最优解，同时标记两者的共性：它们都是选择变量的向量。通常我用符号 $\bar{x}$ 代表一般变量 x 的最优值，而 $\bar{x}$ 的分量被写作 $\bar{x}_1$ 等。

那个要被最大化的函数，被称为目标函数 $F(x)$。约束函数是一般的非线性函数，

$$G(x) = c \tag{2.1}$$

其中 G 为函数，而 c 是给定常数。第 1 章的模型是一个特例：F 为效用函数，G 是表示支出的线性函数，

$$G(x) = p_1 x_1 + p_2 x_2$$

其中 c 为收入。如果有助于理解，你可以把 G 想象成一个更一般的非线性

支出或成本函数，比如出现在消费者面对数量折扣或保费价格表时的支出函数。

有了这些表示法，本章的问题就是要找到在满足 $G(x)=c$ 的情况下使 $F(x)$ 最大化的值 $\bar{x}$。

2.2 套利方法

正如我们在第 1 章中所做的那样，我们从一个试验性的点和一个无穷小的变动开始。让这个特定的试验性的点为 $\bar{x}$，无穷小的变动为

$$\mathrm{d}x=\begin{pmatrix}\mathrm{d}x_1\\ \mathrm{d}x_2\end{pmatrix}$$

我们要求($\bar{x}+\mathrm{d}x$)满足约束条件，并看它是否能产生一个更高的目标函数值。

由于 x 的变动是无穷小的，我们可以用泰勒级数的一阶线性项将目标函数值的变动近似为 x 的函数。用函数的下标表示该函数对相应变量的偏导数，例如 F_1 就是$\partial F/\partial x_1$。我们必须牢记的是，一般而言，每一个偏导数本身都是整个向量 x 的一个函数。只有在函数 F 为线性的特殊情形下，偏导数才会是常数。也只有在函数 F 是加性可分(x_1 的函数加上 x_2 的函数)的特殊情形下，F_1 才会独立于 x_2，反之亦然。所以，我们应当把偏导数写作函数 $F_1(x)$和 $F_2(x)$。当我们在最初的试验性的点 $\bar{x}$ 处计算这两个值时，它们的值就被写作 $F_1(\bar{x})$和 $F_2(\bar{x})$。

从任意一点 x 到 $x+\mathrm{d}x$ 的微小变动所引起的 $F(x)$的变动的一阶泰勒近似为

$$\begin{aligned}\mathrm{d}\,F(x)&=F(x+\mathrm{d}x)-F(x)\\&=F_1(x)\mathrm{d}x_1+F_2(x)\mathrm{d}x_2\end{aligned}\tag{2.2}$$

注意它和第 1 章中效用变动的表达式(1.3)是相似的。那个式子中从经济学的角度得到的边际效用在本式中为效用函数对两种商品的消费数量的偏导数。

对 $G(x)$ 的变动也有一个类似的表达式：

$$dG(x) = G_1(x)dx_1 + G_2(x)dx_2 \tag{2.3}$$

在第 1 章中 G 的偏导数就是价格。如果我们把 G 看成是一个更一般的非线性费用或支出函数,那么它的偏导数就是相应商品的边际价格。

现在我们可以对第 1 章中的套利方法进行修改以应用到新的更一般的背景中去。从满足约束条件(2.1)式的点 $\bar{x}$ 开始,考虑变动 dx 以使得 $(\bar{x}+dx)$ 仍然满足这个约束条件,那么 $dG(\bar{x})=0$。对那个特定的初始点应用(2.3)式,我们有

$$G_1(\bar{x})dx_1 = -G_2(\bar{x})dx_2$$

把等式两边相同的那个值称为 dc,那么我们的套利就是在函数 $G(x)$ 的值中把数量 dc 从 x_2 重新配置到 x_1。

首先假定 $G_1(\bar{x})$ 和 $G_2(\bar{x})$ 都是非零的,那么

$$dx_1 = dc/G_1(\bar{x}), dx_2 = -dc/G_2(\bar{x})$$

代入(2.2)式,得到目标函数值的变动为

$$dF(\bar{x}) = [F_1(\bar{x})/G_1(\bar{x}) - F_2(\bar{x})/G_2(\bar{x})]\, dc \tag{2.4}$$

如果方括号中的表达式非零,那么通过选择一个同方括号中的表达式有相同符号的 dc, $F(x)$ 的值就可以增加。和前面一样,我们可以反一个方向找到一个当 $\bar{x}$ 是最优选择时成立的无套利条件。只要 $\bar{x}_1$ 和 $\bar{x}_2$ 都不碰到某个自然边界,如零,那么变动量 dc 就可以是任意符号的。如果 $\bar{x}$ 是最优的,那么就不会有这样的变动使得 $F(x)$ 增加。所以(2.4)式的方括号中的表达式应该为零,即

$$F_1(\bar{x})/G_1(\bar{x}) = F_2(\bar{x})/G_2(\bar{x}) \tag{2.5}$$

这和前一章中的条件(1.6)式类似。

注意以下精确的表述:如果最优选择为 $\bar{x}$,那么它满足(2.5)式。我并没有给出相反方向的任何含义,所以并不能确保一个满足(2.5)式的解就是最优解。这正是必要条件和充分条件的区别所在,我将在本章后面部分进行更具体的讨论。

如果 $\bar{x}$ 位于某个自然边界上,例如当两者都必须非负时 $\bar{x}_1$ 和 $\bar{x}_2$ 中的一个为零,那么只有一个方向的变动 dc 才是有意义的。我们得到对应于(1.4)式和(1.5)式的一个不等式。在本章中我将不考虑此种情形,但将会

在后面回到这上面来。

和在第 1 章中一样，我们定义 λ 为(2.5)式两边相同的值。于是我们可以把该方程写成两个方程

$$F_j(\bar{x}) = \lambda G_j(\bar{x}),\ j = 1,2 \tag{2.6}$$

回忆一下，在第 1 章中 λ 可以解释为收入的边际效用。同理，此处引入的 λ 正是$F(x)$的最优值对应于 c 的一个变动而作出相应变动的比例。我将在第 4 章中深入分析这一解释及其含义。本章余下的部分主要讨论采用一种容易扩展到更一般背景的方式来推导(2.6)式，并提供一个找到最优解 $\bar{x}$ 的方法。不过，接下来的先是一些必要的略显离题的内容。

2.3 约束规格

如果 $G_1(\bar{x})$为零，那么情况会如何呢？现在在不会影响到约束条件的情况下 $\bar{x}_1$ 可以略微变动。如果 $F_1(\bar{x})$不等于零，那么这种变动就是合意的。例如，如果 $F_1(\bar{x})$为正，那么增加 x_1 可以使 $F(x)$增加。这种改进可以一直继续下去，直到要么 $F_1(x)$下降为零，要么 $G_1(x)$变得不再为零。从消费理论的角度来解释，一个消费者会持续地过多使用免费商品直到他满足为止，或者直到使用额外一个单位的这种商品不再是免费的为止。因而，如果 $G_1(\bar{x})$为零并且 $\bar{x}$ 是最优解，那么 $F_1(\bar{x})$也必定为零。我们可以把这两个零的比值定义为任何我们想要的数，当然将其定义为满足(2.5)式的值也无妨。

当 $G_1(\bar{x})$和 $G_2(\bar{x})$均为零时，情况又会如何呢？这也许意味着这两种商品都是免费的，从而应该被消费到满足点。但是这种情况下，存在因代数和微积分的扭曲所造成的一个更坏的可能性。取第 1 章中的预算线(1.1)式，并把预算方程写作

$$(p_1 x_1 + p_2 x_2 - I)^3 = 0$$

这是(1.1)式的重写方式，一种没有必要这么复杂的方式，但是两者在数学上是完全等价的。我们应该看看这样的变化是否会产生不同的结果。令 $G(x)$为这个式子左边的函数，那么

$$G_1(x) = 3\ p_1(p_1x_1 + p_2x_2 - I)^2$$

当 x 满足预算约束时 $G_1(x)$ 永远为零。对于 $G_2(x)$ 也是如此。商品在边际上并不是免费的，然而它们的数量对约束函数却没有影响。在这种情况下，我们的方法就陷入了困境。

通过假定一个称为约束规格(Constraint Qualification)的条件，正规的数学理论放弃了处理这种情况。在当前的这个例子中，这就等于假定 $G_1(x)$ 和 $G_2(x)$ 中至少有一个是非零的。在一个特定的应用中如果这一点不成立，那么像(2.6)式那样的条件在那时就可能失效。所幸的是，现实中几乎很少有不满足约束规格的情况出现。而那些非常少见的情况，也可以通过把预算约束写成不同的代数形式而得以规避，如上面预算约束的例子。但是学习这门课程的学生并不应该把这个问题全部忘记，这是考试中教师最喜欢出难题的地方。当你用标准方法做时如果出现一些奇怪的错误，那么你就应该检验那个问题是否违反了约束规格。但是，除了在正式的论述和数学证明中以外，我将不会再进一步提到这个复杂的内容。

2.4 相切方法

第二个离题的内容是将套利方法同我们更加熟悉的初级经济学教材中的相切条件联系起来。在第 1 章中，我们看到得到条件(1.6)式的另外一个方法，它以预算线和无差异曲线的相切为基础。对于(2.5)式也同样可以如此。图 2.1 描述了这个过程。沿着曲线 $G(x) = c$，我们有 $\mathrm{d}G(x) = 0$，从(2.3)式我们可以计算出曲线在 x 处的切线的斜率为

$$\mathrm{d}x_2/\mathrm{d}x_1 = -G_1(x)/G_2(x) \tag{2.7}$$

注意下标的颠倒，正如在第 1 章中的一样。还要注意，如果 $G_2(x) = 0$，这条曲线就是垂直的。这并非是一个严重的问题。如果 $G_1(x)$ 和 $G_2(x)$ 都为零，曲线的斜率就不能被很好地定义，这种方法也会陷入我们在上面所解释过的问题中。在绝大多数的经济学应用中，G 是一个关于这两个变量都递增的函数。因而 $G_1(x)$ 和 $G_2(x)$ 都是正的，沿着曲线的 $\mathrm{d}x_2/\mathrm{d}x_1$ 为负，对于约束条件我们有常见的向下倾斜的转换边界。

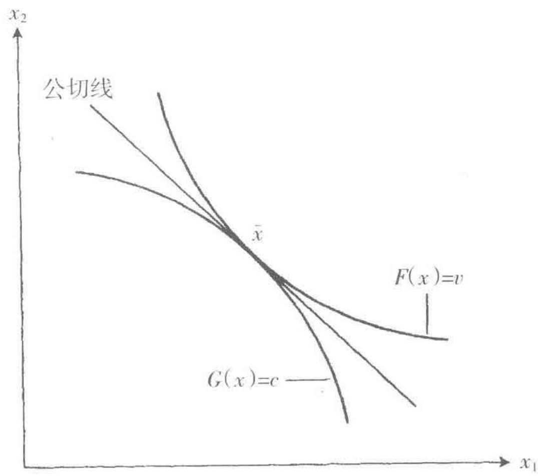

图 2.1　相切解

目标函数 $F(x)$ 的等值线，即 $F(x)$ 的值相等的曲线，穿过点 x。这条等值线在该点的切线的斜率也可以类似地计算：

$$dx_2/dx_1 = -F_1(x)/F_2(x) \tag{2.8}$$

同样地，在绝大多数的经济学应用问题中，F 是一个增函数，等值线（无差异曲线）是向下倾斜的。

如果 $\bar{x}$ 是最优解，那么这两条曲线必定是彼此相切的，也就是在这一点处有相同的斜率。令两个表达式相等，我们有

$$F_1(\bar{x})/F_2(\bar{x}) = G_1(\bar{x})/G_2(\bar{x}) \tag{2.9}$$

这与(2.5)式是等价的。

2.5　必要条件和充分条件

回想上述方法所推导得出的结论：如果在 $G(x)=c$ 的约束下在 $\bar{x}$ 点可以使得 $F(x)$ 取得最大值，那么(2.5)式成立。换言之，条件(2.5)式是 $\bar{x}$ 为最优解的一个逻辑上的结果。因而它被称为最优化的必要条件。更准确地说，由于它涉及函数 F 和 G 的一阶导数，它被称为一阶必要条件。

在寻找最优解时，一阶必要条件有助于缩小搜寻的范围。我们是从一个假定的已知最优解 $\bar{x}$ 开始，来建立起必要条件的。但是我们也可把

过程反过来，将分量 $\bar{x}_1$ 和 $\bar{x}_2$ 看作是未知的，将约束条件(2.1)式和必要条件(2.5)式看作决定它们的两个方程。通常，存在满足约束条件(2.1)式的 $\bar{x}$ 的整个都连续的值域。但是只有 $\bar{x}$ 的少数几个值，而且如果幸运的话，只有其中的一个值会同时满足条件(2.5)式。

如果我们从分离超平面定理导出，我们的问题确实有一个解，并且我们发现仅有惟一的 $\bar{x}$ 同时满足约束条件和一阶必要条件，那么它一定是我们所要找的最优解。如果同时满足(2.1)式和(2.5)式的解有多个，那么就我们目前的分析而言，它们都可能是最优解，所以还必须采用某个其他的方法来得到正确的最优解。但是即便如此，一阶必要条件(2.5)式也已经大大地削减了我们需要检验的可能为最优的解的数量。

一阶必要条件之所以不能总是带给我们正确的解，主要是因为相同的一阶条件也是在 $G(x)=c$ 的约束下最小化 $F(x)$ 这一问题的必要条件。最小化 $F(x)$ 和最大化 $-F(x)$ 是等价的。通过与得到(2.8)式相同的推理，我们会发现该函数的等值线的斜率为：

$$\mathrm{d}x_2/\mathrm{d}x_1=-[-F_1(x)]/[-F_2(x)]=-F_1(x)/F_2(x)$$

让此斜率在 $\bar{x}$ 处的取值和(2.7)式相等，我们再一次得到(2.9)式。

也可以这样理解，为得到这个条件我们看 x 的微小变动，是否能够提高 $F(x)$ 的值。如果不能，那么 x 比一个较小的邻域内的其他点好，或者说它对应了 $F(x)$ 的一个局部最高点。一般而言，一个函数可以有几个这样的局部最高点，也会有几个局部最低点。对于所有这些点同一个一阶必要条件都是成立的，而仅有一个点给出真正的或全局的最大值。

最后，最大值和最小值都能推出(2.5)式，但反过来推测不成立。因而甚至在一个非常小的邻域内，也可能有一个既不是最大值也不是最小值的点满足这个条件。来看一个简单的例子，考虑 $F(x)=x^3$，其中 x 为一个标量。我们有 $F'(0)=0$，但是在 $x=0$ 处，$F(x)$ 既不能取到最大值也不能取到最小值。

为了区别这些情形，任何满足一阶必要条件的点被称为稳定点(stationary point)，真正的最优解是这些点中的一个。为了把这个点从其他候选点中找出来，我们需要某些其他的检验方法。这样的检验通常要依赖函数的曲率或二阶导数。在图 2.1 中，给出的函数 F 和 G 的等值线的曲率恰

好能得到一个最大值。这样，当 x_1 沿着曲线 $G(x)=c$ 不断减少而 x_2 不断增加时，这条曲线就会变得越来越平坦，其经济解释为：当越来越多的 x_1 被转换成 x_2 时，它们之间的边际转换率就会减小。类似地，F 的等值线也显示了边际替代率递减。

涉及曲率和二阶导数的检验是第 6—8 章的主题。这些检验在另一个方面也有别于本章的一阶条件：如果这样的条件成立，那么至少在一个邻域内满足该条件的点就是最大值点；此条件保证了此点是最优解。因而，这样的条件就被称为二阶充分条件。

2.6 拉格朗日方法

现在让我们用一种易于记忆和使用的方式来推导一阶必要条件(2.6)式，因其发明者这种方法被称为拉格朗日方法。注意的是，我们希望用这个条件解出最优解 $\bar{x}$。我们引入 λ 作为(2.5)式两边相同的值，所以它和最优解 $\bar{x}$ 一样是未知的。也就是说，我们必须把它作为解的一个组成部分来决定它的值。同时，称其为待定的拉格朗日乘子。定义一个被称为拉格朗日函数的新的函数如下：

$$L(x,\lambda) = F(x) + \lambda[c - G(x)] \tag{2.10}$$

注意 L 也是 c 与任何出现在 F 和 G 的函数形式中的其他参数的函数。在具体问题中只有当 L 中的这些变量是重要的时候，它们才会被明确地表示出来。

记 L 的偏导数为

$$L_j \equiv \partial L/\partial x_j, L_\lambda \equiv \partial L/\partial \lambda$$

那么

$$L_j(x,\lambda) = F_j(x) - \lambda G_j(x)$$

和

$$L_\lambda(x,\lambda) = c - G(x)$$

一阶必要条件(2.6)式就变为 $L_j=0$，$j=1$ 和 2，而约束条件(2.1)式就是 $L_\lambda=0$。因而我们可以将目前为止整个推导的结果表述为一个简单的结论：

拉格朗日定理：假定 x 为一个二维的向量，c 为标量，函数 F 和 G 取标量值。同(2.10)式一样定义函数 L。如果没有其他约束(例如非负约束)条件，$\bar{x}$ 使 $F(x)$ 取得最大值且满足 $G(x)=c$，并且如果对于至少一个 j，$G_j(\bar{x})\neq 0$成立，那么存在一个 λ 值，使得

$$L_j(\bar{x},\lambda)=0, j=1,2;\ L_\lambda(\bar{x},\lambda)=0 \tag{2.11}$$

记住，这个定理只提供了最优化的必要条件。换言之，它从一个已知的最优解 $\bar{x}$ 开始，推导得出它必须满足条件(2.11)式。但在实际中，这个定理更多的是帮助我们缩小寻找最初的未知最优解的范围。我们把(2.11)式视为关于三个未知量 $\bar{x}_1$、$\bar{x}_2$ 和 λ 的三个方程。这些方程一般是非线性的，方程解的存在性和惟一性也不能得到保证。如果在这些条件下没有解，原因可能是这个最大化问题本身就没有解，或者是约束规格不被满足从而使一阶条件不能适用。如果在这些条件下有多个解，我们就需要用二阶条件在候选解中作出判断。但是在绝大部分的应用中，最优化问题能被很好地构造从而一阶必要条件就足以带给我们惟一的最优解。现在我给出几个使用拉格朗日方法的例题，并提供一些练习，由你自己尝试求出类似的解。在你获得了一些处理有两个变量和一个约束条件的问题的经验之后，你就对在下一章中将要考虑的一些扩展内容做好准备了。

虽然用符号 $\bar{x}$ 将一般点和最优点区别开来以保持理论的清晰展开，但在我们搜寻未知的最优解的应用中，这会显得很繁琐。因而在一些具体情形(如下面的例子)中，如使用像(2.11)式一样的条件时，我们经常省略掉 x 上面的横线。

例题

例题 2.1：隐含不变预算份额的偏好

考虑一个消费者在价格分别为 p 和 q 的两种商品 x 和 y 之间进行选择(在理论部分使用符号 x_1、x_2 等是因为它们更容易被一般化到多种商品和多个约束条件的情景中，但是在仅有两种商品的例子中用符号 x 和 y 就更简单)。他的收入为 I，所以预算约束为

$$px + qy = I$$

假定效用函数为

$$U(x,y) = \alpha\ln(x) + \beta\ln(y) \tag{2.12}$$

其中 α、β 是正的常数，ln 代表自然对数。

将拉格朗日函数写作

$$L(x,y,\lambda) = \alpha\ln(x) + \beta\ln(y) + \lambda[I - px - qy]$$

回忆 $\mathrm{d}\ln(x)/\mathrm{d}x = 1/x$，因而一阶必要条件(2.11)式变为

$$\partial L/\partial x \equiv \alpha/x - \lambda p = 0$$

$$\partial L/\partial y \equiv \beta/y - \lambda q = 0$$

和

$$\partial L/\partial \lambda \equiv I - px - qy = 0$$

为求解这些方程，将第一和第二个方程中的 x、y 代入第三个方程，得到

$$\lambda = (\alpha + \beta)/I \tag{2.13}$$

和

$$x = \frac{\alpha I}{(\alpha+\beta)p}, y = \frac{\beta I}{(\alpha+\beta)q} \tag{2.14}$$

这些就是需求函数，即用价格、收入和给定参数 α、β 表示的最优消费数量的解。

我们也可以把它们写作

$$\frac{px}{I} = \frac{\alpha}{\alpha+\beta}, \frac{qy}{I} = \frac{\beta}{\alpha+\beta} \tag{2.15}$$

换言之，对于这个特定的效用函数，花在这两种商品上的收入所占的份额是常数。这个性质应用起来非常方便，而且在有些时候和现实十分接近。在对理论模型进行最初的探索时，这种设定往往是一个非常重要的简化，可以使模型得到具体的解，为进一步分析和检验模型指出方向。因而这个函数颇受经济学家们的宠爱。

注意在(2.13)式中，收入的边际效用与收入成反比。这看上去似乎是直觉上很吸引人的边际效用递减思想的一个自然的结果。但是这是一个靠

不住的观点，见下面的练习 2.1。

例题 2.2：大炮对黄油

考虑一个有 100 单位劳动的经济。它可以生产 x 单位大炮或者 y 单位黄油。假设要生产 x 单位大炮，需要使用 x^2 单位的劳动，同样，生产 y 单位黄油需要使用 y^2 单位的劳动。所以这个经济面临的资源约束为

$$x^2 + y^2 = 100$$

从几何上看，你可以很容易发现这条生产可能性边界是四分之一圆。

我们想要最大化的目标函数为

$$F(x, y) = ax + by$$

其中 a、b 为给定的正的常数。

为了求解这个问题，建立拉格朗日函数

$$L(x, y, \lambda) = a\,x + b\,y + \lambda[100 - x^2 - y^2]$$

一阶条件为

$$\partial L/\partial x \equiv a - 2\lambda x = 0$$
$$\partial L/\partial y \equiv b - 2\lambda y = 0$$

和

$$\partial L/\partial \lambda \equiv 100 - x^2 - y^2 = 0$$

将前两个方程代入第三个方程可以得到

$$100 = (a^2 + b^2)/(4\lambda^2)$$

或者

$$\lambda = (a^2 + b^2)^{1/2}/20$$

则

$$x = 10a/(a^2 + b^2)^{1/2}, \; y = 10b/(a^2 + b^2)^{1/2} \tag{2.16}$$

你可以把 a、b 看作赋予这两种商品的权重或者社会价值，则(2.16)式给出了这个经济的最优供给，它们是这两个权重的函数。如果这两个权重

同比例增加，如都翻了一倍，则最优数量 x 和 y 不变。供给对这两个权重是零次齐次的，所以其仅与权重的相对值有关。当一种商品权重的相对值上升时，这种商品的供给也随之上升。在后面的内容中，特别是在有关比较静态的那一章中，我们会看到这些性质在一般情况下是如何奏效的。

习题

习题 2.1：柯布—道格拉斯效用函数

考虑同例题 2.1 中一样的消费者选择问题，只是他现在有一个不同的效用函数 $\widetilde{U}$，定义为

$$\widetilde{U}=x^{\alpha}y^{\beta}$$

证明它会产生一个和上面(2.14)式相同的不变预算份额的需求函数。（提示：为了简化求解过程，将拉格朗日乘子从关于两种商品的一阶条件方程中消去，得到 x 和 y 之间的一个关系式。将其尽可能地化简，再利用其和预算约束解出商品的消费量。）

注意，这两个效用函数的关系为

$$U(x,y)=\ln[\widetilde{U}(x,y)]\text{ 或者 }\widetilde{U}(x,y)=\exp[U(x,y)]$$

这表明对效用函数进行任何递增变换都不会影响消费者的最优选择。如果我们所关心的仅仅是观察到的需求行为，那么在这样的转换下效用函数的形式就是不确定的(也是无关紧要的)。任何依赖于对某一特定形式的选择而产生的性质都是没有意义的。

这种性质之一就是收入的边际效用递减。如果我们把这个问题中的乘子写作 $\widetilde{\lambda}$ 以区别于例题 2.1 中的 λ，那么你可以验证

$$\widetilde{\lambda}=\frac{\alpha^{\alpha}\beta^{\beta}}{(\alpha+\beta)^{\alpha+\beta}}\frac{I^{\alpha+\beta-1}}{p^{\alpha}q^{\beta}}\tag{2.17}$$

如果$(\alpha+\beta)>1$，则 $\widetilde{\lambda}$ 随着收入上升而增大。

在一些情况下，效用函数的具体形式发挥了特殊的作用。如果我们作了一些有关人与人之间可比性的假定，或者为了表示涉及时间或不确定性的情况下的偏好，而使用了一些在时间或状态之间是加性可分的函数，这种

情况就会发生。但是在所有这些情形中，采用特定的函数形式首先是出于对这些其他因素的考虑，而不是出于消费者个人选择的潜在机制。

在没有考虑这些其他因素时，为了计算上的方便我们可以自由地进行效用函数的转换。注意到以相同比例改变 α 和 β，在例题 2.1 和习题 2.1 中都不会改变需求。看一下(2.14)式或者(2.17)式就会发现，选择比例使得 $\alpha+\beta=1$ 是非常方便的。

习题 2.2　线性支出系统

再次回到例题 2.1 中的消费者，只是此时他的效用函数被修改为 $\hat{U}$，定义为

$$\hat{U}(x,y)=\alpha\ln(x-x_0)+\beta\ln(y-y_0)$$

其中 x_0 和 y_0 是给定的常数，并且 $\alpha+\beta=1$。证明在这两种商品上的最优支出是收入和价格的线性函数：

$$px=\alpha I+\beta px_0-\alpha qy_0$$
$$qy=\beta I-\beta px_0+\alpha qy_0$$

对效用函数的这一微小的修改给它带来了更大的可能为最优选择的范围。现在，这两种商品的预算份额可以随着收入和价格系统性地变动。一种商品可能是必需品而另一种商品可能是奢侈品(但是两种商品都不能是劣等品，因为 α 和 β 必须为正，以确保边际效用为正)。但支出仍然有一个简单的函数形式。出于这些原因，在早期的关于消费者需求的经验研究中这种设定非常普遍。

习题 2.3：生产和成本最小化

考虑一个生产者，他以每年 r 的价格租赁机器 K，以每年 w 的工资雇用劳动 L，以生产产出 Q，其中

$$Q=\sqrt{K}+\sqrt{L}$$

假定他想以最小的成本来生产固定数量的 Q。找出他的要素需求函数。证明拉格朗日乘子为下式：

$$\lambda=2wrQ/(w+r)$$

并为 λ 给出一个经济学解释。

现在用 p 代表产出的价格。假定这个生产者可以改变产出的数量并寻求最大的利润。证明他的最优产出供给量为

$$Q = p(w + r)/(2wr)$$

将上式和你对 λ 的解释联系起来。

▶ 进一步阅读

对拉格朗日方法的补充性分析，参见 Varian（1987，见本书第 1 章中的基础阅读）的第 5 章和第 20 章的附录；和 Smith（1982，见本书第 1 章中的基础阅读）的第 2 章中的第 1—4 节，第 4 章中的第 1—3 节。

本书中对理论的展开相对注重直觉和启发性，也有一些教材在数学上更为严谨，我仅提及一本：Michael D. Intriligator, *Mathematical Optimization and Economic Theory*, Englewood Cliffs, NJ: Prentice-Hall, 1971；其中的第 3 章就是关于拉格朗日方法的。

▶3 扩展与一般化

3.1 多个变量和多个约束条件的情形

如果有 n 个选择变量($x_1, x_2, \cdots, x_n$)，我们就让向量 x 包括 n 个分量。于是(2.11)式就扩展到 $j=1,2,\cdots,n$，我们有包含($n+1$)个未知量的($n+1$)个方程，即 x 的 n 个分量和数 λ。

如果有 m 个约束，将它们写作

$$G^i(x) = c_i \qquad i = 1,2,\cdots,m$$

其中方程是以上标来标记的，以避免和偏导数混淆，后者是以下标来表示的。现在我们忽略任何其他的限制，如对变量的非负约束。我们要求 $m<n$，因为对 n 个变量施加 n 个约束，一般会使选择缩减为一组离散的点；而更多的约束一般将不能相互一致。

拉格朗日方法可以非常容易地扩展到这种情形。对每一个约束我们定义一个乘子 λi，并将拉格朗日函数定义为：

$$L(x_1,\cdots,x_n,\lambda_1,\cdots,\lambda_m) = F(x_1,\cdots,x_n) + \sum_{i=1}^{m}\lambda_i[c_i - G^i(x_1,\cdots, x_n)] \tag{3.1}$$

在最优解 $\overline{x}$ 处满足的一阶必要条件就是

$$\partial L/\partial x_j = 0, j = 1,2,\cdots,n \tag{3.2}$$

和

$$\partial L/\partial\lambda_i = 0, i = 1,2,\cdots,m \tag{3.3}$$

当利用这些条件寻找最优解时，我们把它们当作包含 $\bar{x}_1,\bar{x}_2,\cdots,\bar{x}_n,\lambda_1,\lambda_2,\cdots,\lambda_m$ 这$(m+n)$个未知量的$(m+n)$个方程。

采用向量矩阵的形式可以把它们表述得更简洁。令 c 为由分量 c_i 组成的列向量，G 为由分量函数 G^i 组成的列向量值函数。因而可以用一个向量方程 $G(x)=c$ 将所有的约束写在一起。接下来，偏导数 $F_j(x)$也应该组成一个向量，我将其写作 $F_x(x)$，下标 x 表示求导所对应的向量变量。我将形成以下惯例，当一个函数的变量为列向量时，其偏导数向量则为行向量，反之亦然。这在数学上有很充分的理由，但是此处主要的优点则是，它可以使我们不必始终对矩阵进行转置。每一个 G^i 都有一个偏导数行向量 $G^i_x(x)$，其分量为 $G^i_j(x)$。这些行向量的垂直排列就会形成一个 $m\times n$ 的矩阵，写作 $G_x(x)$。而那些乘子则会形成一个行向量 λ。

有了这些表示法后，(3.1)式就可以写得更简洁：

$$L(x,\lambda) = F(x) + \lambda[c - G(x)] \tag{3.4}$$

让我们对此式进行验证。注意，λ 是一个 m 维的行向量，即一个 $1\times m$ 维的矩阵，而在方括号中的项是一个 m 维的列向量，即一个 $m\times 1$ 维的矩阵。这两个矩阵的乘积值，为行向量中的第 i 个分量乘以列向量中相应的分量，然后将所有这些乘积相加而得。其结果是一个 1×1 维的矩阵，即一个标量，它就是(3.1)式中的表达式。

用相同的方式，条件(3.2)和(3.3)也可以变得更加紧凑：

$$L_x(\bar{x},\lambda) = 0 \tag{3.5}$$

$$L_\lambda(\bar{x},\lambda) = 0 \tag{3.6}$$

此时约束规格又会怎样呢？记得在第 2 章的例子中，我们有两个变量$(n=2)$和一个约束$(m=1)$。矩阵 $G_x(x)$为 1×2 的矩阵，即一个简单的行向量$(G_1(x),G_2(x))$。约束规格就是该向量在 $\bar{x}$ 处非零的假设。而在一般情况下，该条件就是使这个矩阵为非奇异的。由于 $m<n$，这等价于要求它应该有 m 个线性无关的行向量，也就是它的秩应该为可能的最大值，即 m。

让我们把这些结果总结为一个简洁的结论：

拉格朗日定理：假定 x 是 n 维的向量，c 是 m 维的向量，F 为取标量值的函数，G 为取 m 维向量值的函数，并且 $m<n$。定义

$$L(x,\lambda) = F(x) + \lambda[c - G(x)] \tag{3.4}$$

其中 λ 为 m 维的行向量。如果 $\bar{x}$ 在 $G(x) = c$ 的约束下最大化 $F(x)$，假定没有其他的约束，且 $G_x(\bar{x})$ 的秩$= m$，那么存在一个值 λ，满足

$$L_x(\bar{x},\lambda) = 0 \tag{3.5}$$

$$L_\lambda(\bar{x},\lambda) = 0 \tag{3.6}$$

我没有证明地给出了这些一般化的结论。绝大多数读者大概会接受两个变量和一个约束情况下的直觉推理，并仅在应用中检验以上的一般结果。不过即便是这样的读者，也会发现上面的精简陈述是有用的。对于那些更加有数学导向的读者而言，我在附录中对本章中最一般化的结果，即库恩一塔克定理，给出了一个正式的证明。

3.2 非负变量

接下来假定变量 x_j 必须非负，以使一些经济学问题有意义。如果最优的 $\bar{x}$ 恰好非负，因而使得这些限制不起作用，也就是说，如果所有的 $\bar{x}_j$ 都严格为正，那么以上的条件(3.2)式和(3.3)式继续成立。比如说 $\bar{x}_1$ 为零，那么导出一阶条件的套利方法更加受限制。我们只能考虑那些 $dx_1 > 0$ 的无穷小量的变动 dx。将第1章中导出不等式条件(1.10)式的推导一般化就会给出，

$$L_1(\bar{x}) \equiv F_1(\bar{x}) - \sum_{i=1}^{m} \lambda_i G_1^i(\bar{x}) \leqslant 0$$

此处，我再次省略了正式的证明。

更一般地，$\bar{x}$ 的一些分量为正，而另一些则为零。那么在初始点，对于拉格朗日函数对应于每一个正的分量的偏导数，有类似(3.2)式的等式成立，而对于为零的分量，就会有类似上面的不等式。换句话说，对于每一个 j，我们应该有，

$$L_j(\bar{x}) \leqslant 0, \ \bar{x}_j \geqslant 0 \tag{3.7}$$

这其中至少有一个式子会以等式的形式成立。在非常少见的情形中，两个式子会同时以等式的形式成立。但最优化的要求在逻辑上排除了两个式子

同时为严格的不等式的可能性。

(3.7)式中的两个不等式至少有一个以等式形式成立的要求有时可以更简洁地表达为,

$$\bar{x}_j L_j(\bar{x}) = 0$$

要点是,只有当至少有一个因子为零时,乘积才能为零。

像(3.7)式那样的、不能同时是严格的一对匹配的不等式,被定义为是互补松弛的。对于单个不等式而言,比如 $\bar{x}_j \geqslant 0$,如果它是以等式形式成立的,则称它为紧的,也就是说,$\bar{x}_j$ 处于其允许范围的边界上;如果 $\bar{x}_j$ 为正,则称这个不等式是松的,因而在碰到边界之前还有一些调整的空间。(3.7)不等式组中的每一个式子都补充了另一个式子的松弛性,即:如果一个是松的,那么另一个一定是紧的。

我们可以把(3.7)式中所有的分量不等式整理成向量。这里我将利用以下的表示法:$x \geqslant 0$ 表示对于每个 j,都有 $x_j \geqslant 0$;$x > 0$ 表示至少有一个分量不等式是严格的;$x \gg 0$ 表示所有分量不等式都是严格的。于是(3.7)式变为

$$L_x(\bar{x}) \leqslant 0, \bar{x} \geqslant 0, \text{且满足互补松弛条件} \tag{3.7}$$

它被理解为,对向量不等式中的每一对分量不等式,互补松弛条件都成立。

我将结论再次总结如下,以便日后参考:

有非负变量的拉格朗日定理:假定 x 为 n 维向量,c 为 m 维向量,F 为取标量值的函数,G 为取 m 维向量值的函数,且 $m<n$。定义

$$L(x,\lambda) = F(x) + \lambda[c - G(x)] \tag{3.4}$$

其中 λ 为 m 维行向量。如果 $\bar{x}$ 在约束 $G(x) = c$ 和 $x \geqslant 0$ 下最大化 $F(x)$,而且约束规格 $G_x(\bar{x})$ 的秩 $= m$ 成立,那么存在一个值 λ,满足

$$L_x(\bar{x}, \lambda) \leqslant 0, \bar{x} \geqslant 0, \text{满足互补松弛条件} \tag{3.7}$$

和

$$L_\lambda(\bar{x}, \lambda) = 0 \tag{3.6}$$

一阶必要条件被假定缩小了我们搜寻最优解的范围。在这种情况下这是怎样进行的呢?开始时我们并不知道最优解中的哪些分量为正,哪些为零。如果我们假定某种特定的情形,比如 $\bar{x}_1 > 0, \bar{x}_2 = 0, \bar{x}_3 > 0, \cdots$,那么由

(3.7)式我们得到$L_1(\bar{x})=0$，$\bar{x}_2=0$，$L_3(\bar{x})=0$，…。换言之，一个特定的情形会从(3.7)式产生一组 n 个等式的方程。这些方程可能根本无解，或者即使它们有解，也有可能不满足该情形所要求的其他不等式条件。但是如果有一个满足所有这些条件的解，那么它就成为最优选择的一个候选的解。

在 n 维向量 x 中共有 2^n 种分量为正和零的情形。对这些情形中的每一个分量，我们可以重复上面的过程，以找出最优解的其他候选的解。于是我们的搜寻减少了所有这些候选解。

在一些应用中这种搜寻相对容易进行；求解线性规划问题的单纯形方法就是这种搜索过程的一个系统性算法。但是一般地，这种搜索过于繁琐费力。如果我们每次都要这么算一遍，那么求解约束条件下的最优化问题的前景就非常黯淡了。所幸的是，许多实际中有意思的问题为这些情形中的搜索提供了捷径和系统的方法。在经济理论的绝大部分基本情形中，我们可以很好地猜出等式和不等式可能的情形，并在此基础上进行计算，最后利用二阶充分条件验证由此得到的解确实是最优的。

3.3 不等式约束

现在我们可以考虑更一般的不等式约束。这在经济学中非常重要，因为通常并没有强迫人们将手头的收入或某些资源全部用完。在求解的过程中，我们应该决定将它们全部用完是否是最优的。

假定约束中的第一个分量只需要以不等式形式成立

$$G^1(x)\leqslant c_1$$

利用同第1章消费者问题中引入"未花掉的收入"变量一样的技巧，我们可以将此放入前面所讨论的情形中。让我们定义一个新的变量 x_{n+1}，

$$x_{n+1}=c_1-G^1(x) \tag{3.8}$$

通过这个扩大了的变量集，约束变成了一个精确的等式。新变量 x_{n+1} 受到非负约束，但是我们知道怎样去处理这一点。

记 $\hat{L}$ 为新问题的拉格朗日函数，以区别于原来问题中的 L。那么

$$\hat{L}(x_1,\cdots,x_n,x_{n+1},\lambda_1,\cdots,\lambda_m)$$

$$= F(x_1, \cdots, x_n, \lambda_1, \cdots, \lambda_m) + \lambda_1[c_1 - G^1(x_1, \cdots, x_n) - x_{n+1}]$$

$$+ \sum_{i=2}^{n} \lambda_1 [c_i - G^i(x_1, \cdots, x_n)]$$

$$= L(x_1, \cdots, x_n, \lambda_1, \cdots, \lambda_m) - \lambda_1 x_{n+1}$$

除了关于 x_{n+1} 的一个新的一阶条件外，其他所有的一阶条件都和以前一样，$\partial \hat{L}/\partial x_{n+1} \equiv -\lambda_1 \leqslant 0, x_{n+1} \geqslant 0$

满足互补松弛条件。

回想起

$$x_{n+1} = c_1 - G^1(x) = \partial L/\partial \lambda_1$$

因而这个条件可以写成：

$$\partial L/\partial \lambda_1 \geqslant 0, \lambda_1 \geqslant 0 \tag{3.9}$$

满足互补松弛条件 ：这是一个和非负变量的一阶条件(3.6)式非常对称的形式。

类似地，我们可以允许所有的约束都为不等式。如果是这样，一般就没有理由保留 $m<n$ 的限制，因为对选择变量 x 而言，由不等式约束的任何数量仍然可以有不小的变动范围。其他分量的一阶条件正好是(3.9)式的对应式，我们可以把这些一阶条件表示成一对向量不等式，而在每一对分量不等式中满足互补松弛条件。

约束规格也发生了变化。具体来说，假定有 k 个约束是紧的，即以等式的形式成立。从导数矩阵 $G_x(x)$ 中取出对应于这 k 个约束的行向量。由此得到的 $k \times n$ 子矩阵的秩应该为 k。

我们再次给出正式的表述：

有不等式约束的拉格朗日定理： 假定 x 为 n 维向量，c 为 m 维向量，F 为取标量值的函数，G 为取 m 维向量值的函数。定义

$$L(x, \lambda) = F(x) + \lambda[c - G(x)] \tag{3.4}$$

其中 λ 为 m 维行向量。如果 $\bar{x}$ 在满足 $G(x) \leqslant c$ 的约束下最大化 $F(x)$，而且约束规格成立，即由那些满足 $G^i(\bar{x}) = c_i$ 的行向量 i 组成 $G_x(\bar{x})$ 的子矩阵具有最大可能的秩，那么存在一个值 λ，满足

$$L_x(\bar{x}, \lambda) = 0 \tag{3.5}$$

和

$$L_\lambda(\bar{x},\lambda) \geqslant 0,\lambda \geqslant 0\text{ ,满足互补松弛条件} \tag{3.10}$$

最后,我们可以把非负变量和不等式约束的情形结合起来,从而得到这类问题最一般的结论:

库恩—塔克定理:假定 x 为 n 维向量,c 为 m 维向量,F 为取标量值的函数,G 为取 m 维向量值的函数。定义

$$L(x,\lambda) = F(x) + \lambda[c - G(x)] \tag{3.4}$$

其中 λ 为 m 维行向量。如果 $\bar{x}$ 在 $G(x)\leqslant c$ 和 $x\geqslant 0$ 的约束下最大化 $F(x)$,而且约束规格成立,即由那些满足 $G^i(\bar{x})=c_i$ 的行向量 i 组成 $G_x(\bar{x})$ 的子矩阵具有最大可能的秩,那么存在一个值 λ,满足

$$L_x(\bar{x},\lambda) \leqslant 0,\ \bar{x} \geqslant 0\text{,满足互补松弛条件} \tag{3.7}$$

和

$$L_\lambda(\bar{x},\lambda) \geqslant 0,\ \lambda \geqslant 0\text{,满足互补松弛条件} \tag{3.10}$$

再一次,对于$(m+n)$个互补松弛条件中每一种可能的情形,利用此定理找到解的这个繁琐程序即要从所有 $2^{(m+n)}$ 种情形中搜寻。不过,通常是有捷径的。我们会很快看到一些说明和应用这一定理的例子。

例题

例题 3.1:拟线性偏好

假定有两种商品 x 和 y,其数量必须非负,价格分别为 p 和 q,都为正。考虑一个有收入为 I 的消费者,其效用函数为

$$U(x,y) = y + a\ln(x)$$

其中 a 为一个给定的、正的常数,这样的偏好被称为拟线性偏好,因为对其中一种商品的数量而言效用函数是线性的。

为了得到这个消费者的需求函数,我们可以用库恩—塔克定理。构造拉格朗日函数

$$L(x,y,\lambda) = y + a\ln(x) + \lambda[I - px - qy]$$

在这个问题中一阶条件(3.7)式和(3.10)式即为

$$a/x-\lambda p\leqslant 0, x\geqslant 0 \tag{3.11}$$

$$1-\lambda q\leqslant 0\ , y\geqslant 0 \tag{3.12}$$

$$I-px-qy\geqslant 0, \lambda\geqslant 0 \tag{3.13}$$

每一个情形都满足互补松弛条件。

让我们机械地求解以培养起对这种方法的直觉。在两个非负变量和一个不等式约束的情况下,一共有 $2^3=8$ 种可能的等式和不等式情形。我们看其中哪些会成为最优解的候选解。

首先,注意到预算约束不可能是松的,其经济含义非常简单:因为商品的边际效用为正,预算不可能不被花完。正式地说,如果预算约束是松的,那么(3.13)式给出 $\lambda=0$,(3.11)式就会要求 $a/x\leqslant 0$,而这是不可能的;类似地(3.12)式也会要求 $1\leqslant 0$。

这把我们需要考察的情形缩减到四种,也就是(3.11)式和(3.12)式中的 x 和 y 大于零或者等于零的情形。

正如我们前面看到的,x 和 y 都为零不满足预算等式,而预算约束是一定成立的。因而这种情形在逻辑上是不可能的。其经济学道理依旧是因为商品具有正的边际效用。

如果 $x=0$ 和 $y=I/q>0$,那么(3.12)式会给出 $\lambda=1/q$,(3.11)式就会变成$p/q\geqslant\infty$,显然这是不成立的。因而这种情形也不可能出现。其在经济学上的理由是,最初一单位的 x 具有无穷大的边际效用,所以消费零单位 x 不可能是最优的。

接下来考虑 $x>0$ 和 $y=0$ 的情形。于是从预算约束我们得到 $x=I/p$,从(3.11)式得到$\lambda=a/I$。把它代入(3.12)式,我们会有 $1\leqslant aq/I$,或者 $I\leqslant aq$。这是针对问题给定参数的一个条件,这些参数可能满足也可能不满足这个不等式。如果它们满足的话,那么这种情形所要求的前提条件就是彼此一致的,从而我们得到最优解的一个候选解。

最后,如果 x 和 y 都是正的,(3.11)式和(3.12)式就会给出 $a/(px)=\lambda=1/q$,所以 $x=aq/p$。于是预算约束得出 $y=I/q-a$。如果 $I>aq$,那么这在逻辑上是一致的。同样地,这些参数可能满足也可能不满足这个不等式。如果它们满足的话,我们也得到了最优解的一个候选解。

对各种情形的讨论已经完毕。让我们从这些推导的过程和结果中得出一些有益的想法:第一个值得注意的是,在 8 种可能的情形中,有 6 种是可以通过经济学直觉排除掉的。因此,在积累了一定的经验之后,用这种方式将正式的推导过程大大简化是可能的。

其次,我们观察到使最后两种情形中的每一种内部一致的参数条件将 p,q 和 I 所有可能的值的空间划分成两个互补而不交的集合。一种情形要求 $I \leqslant aq$,而另一种要求 $I > aq$。在许多实际应用中,都会出现类似的巧妙的分类。

第三,当 $I = aq$ 时,我们有 $\lambda = 1/q$。所以

$$1 - \lambda q = 0, y = 0$$

即(3.12)式中的两个不等式都以等式成立。我在(3.6)式之后讨论互补松弛条件时,曾提及这种可能性,并称其为例外的情形。现在我们看这是为什么:它源于参数的一种特殊结构,此时问题的解正好位于互补松弛条件中等式和不等式的一种情形转换到另一种情形的点上。

最后,让我们重新表述最优选择的规则:

$$\text{如果 } I \leqslant aq, \text{那么 } x = I/p, y = 0$$
$$\text{如果 } I > aq, \text{那么 } x = aq/p, y = I/q - a$$

为了更清晰地理解这个解,让我们来想象一下。假定收入从非常低的水平开始,逐渐上升。起先,所有的收入都花在了商品 x 上,而不消费任何 y。在某一点上后,在 x 上的支出保持不变,所有额外的收入都用于 y 上。我们可以把 x 想作一个必需品的典型,它对收入有绝对优先的要求,但是一旦它的要求得到满足,所有额外的收入就可以用于其他商品了。

当我们想分离出一个部门或者产业,并希望避免对其商品需求的收入效应造成影响,拟线性偏好模型就很有用。这常被称为“局部均衡”,或更贴切地称为“产业分析”。显然,这并不意味着人们只是从字面上接受“收入的变化不影响我们所要考察的商品的需求”的假定,目前,这一假定常被证明是一个可接受的近似或者简化。

例题 3.2:技术性闲置

假定一个有 300 单位劳动和 450 单位土地的经济体。这些要素可用来

生产小麦或者牛肉。生产每单位的小麦需要 2 单位劳动和 1 单位土地，而生产每单位的牛肉需要 1 单位劳动和 2 单位土地。

一个生产 x 单位小麦和 y 单位牛肉的计划是可行的，如果其对每一种生产要素的需求都不大于该要素现有的数量：

$$2x + y \leqslant 300 \tag{3.14}$$

$$x + 2y \leqslant 450 \tag{3.15}$$

每一个不等式表示的是所有落在直线上或者直线以下的点。给定这两个约束，生产可行集合就是图 3.1 中的四边形 OABC。可行集的东北部的边界，即生产可能性边界就是 ABC。

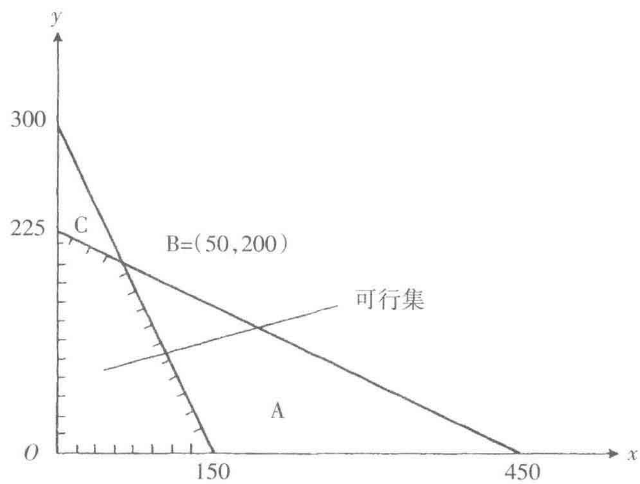

图 3.1 生产和闲置

沿着线段 AB，(3.14)式以等式形式成立，而沿着线段 BC，(3.15)式以等式形式成立。只有在 B 点，这两个式子都会以等式形式成立。在任何其他的点，都会有一种要素闲置。

你或许会禁不住地假定充分利用资源是最优的，即在 B 点进行生产，$x=50, y=200$。但是，情况并不一定如此。

假定这个社会有一个目标函数或者社会福利函数，其被定义为关于这两种产品数量的函数，具有我们前面用过的那种简单形式：

$$W(x, y) = \alpha \ln x + \beta \ln y \tag{3.16}$$

其中 α 和 β 是给定的正常数，且 $(\alpha+\beta)=1$。

根据前面的例子，我们直觉上知道，对于 x 和 y 的非负约束不会是紧的。所以从一开始我们就可以排除这种情形。将对应于约束(3.14)式和对应于约束(3.15)式的拉格朗日乘子分别记为 λ 和 μ。构造拉格朗日函数为：

$$L(x,y,\lambda,\mu)=\alpha\ln x+\beta\ln y+\lambda[300-2x-y]+\mu[450-x-2y]$$

一阶条件为：

$$\alpha/x-2\lambda-\mu=0 \tag{3.17}$$

$$\beta/y-\lambda-2\mu=0 \tag{3.18}$$

和

$$300-2x-y\geqslant 0,\lambda\geqslant 0,\text{满足互补松弛条件} \tag{3.19}$$

$$450-x-2y\geqslant 0,\mu\geqslant 0,\text{满足互补松弛条件} \tag{3.20}$$

在(3.19)式和(3.20)式之间，我们有四种可能的等式和不等式相互组合的情形。我们应该相信，两种要素都闲置的情形，即 $\lambda=0$，且 $\mu=0$ 是不合理的。让我们检验这一点：把 $\lambda=0=\mu$ 代入(3.17)式和(3.18)式，就会得到 $\alpha=0=\beta$，而这又是不可能的。所以，这种情形排除了之后，我们就剩下三种情形了。

首先考虑 $\lambda=0$ 和 $\mu>0$ 这种情形，我把它记作情形(i)。这里(3.17)式给出 $x=\alpha/\mu$，(3.18)式给出 $y=\beta/2\mu$。由于 $\mu>0$，那么(3.20)式就变成

$$450=x+2y=(\alpha+\beta)/\mu \text{ 或 } \mu=1/450$$

于是 $x=450\alpha$，$y=225\beta$。

还剩下检验(3.19)式中的可行性条件是否满足。我们需要检验：

$$300\geqslant 2x+y=900\alpha+225\beta=900-675\beta \text{ 或 } \beta\geqslant 8/9$$

其他的情形也可以用同样的方法来检验，我在此仅给出结果：

如果 $\lambda>0$ 和 $\mu=0$(情形(ii))，我们得到 $x=150\alpha$，$y=300\beta$。当 $\beta\leqslant 2/3$ 时，这种情形就是内在一致的。

如果 $\lambda>0$ 和 $\mu>0$(情形(iii))，我们得到两种要素的充分利用点 $x=50$ 和 $y=200$。当 $2/3<\beta<8/9$ 时，这种情形就是内在一致的。

这些结论也给出了几个有用的见解。同样地，参数的范围恰好被分为互补且不相交的区域，在每一个区域中，仅有一种情形产生一个最优化的候

选解。对于值较小的β,解落在线段 AB 上。对于值在中间范围的β,解落在点 B 上。最后,对于值很大的β,解落在线段 BC 上。

目标函数(3.16)式的社会无差异曲线类似于双曲线。赋予 y 的权重β(相对于 x 的权重)越大,社会为了得到 y 更愿意牺牲 x,即双曲线更平坦。因此,对于一个较小的β,社会福利的等值线与生产可能性集合的切点就会落在线段 AB 上;对于中间大小的β,我们会在 B 处得到一个角点解;对于一个较大的β,切点就会落在 BC 上。

接着,注意到在 AB 上的任何一点(除了 B 点以外),表示让一些土地闲置是最优的。为了理解这一点,我们注意到,物品的生产有固定的投入系数要求,小麦需要相对更多的劳动投入。如果我们希望使用闲置的土地,我们就必须生产更少的小麦和更多的牛肉。而其他的方法将增加劳动投入的要求,而此时劳动已经被充分利用了。但这是一种β相对较小的情形,说明人们对牛肉的评价没有对小麦的评价那么高,因而要求牺牲小麦就是不值得的。

如果生产中足够的替代是可能的,那么上述难题就不会出现,两种要素都会得到充分利用。而且在这个理论框架下出现的要素闲置就是技术刚性的一个结果,而并非因为任何有效需求的不足或者协调的失败。

最后,再来看看互补松弛条件(3.19)式和(3.20)式。如果在最优解处有一种要素未被充分利用,那么对应于它的约束的拉格朗日乘子就为零。在第 1 章的例子中,消费者预算约束的乘子就是货币的边际效用。同样,在这个例子中的每一个乘子都给出了每一额外单位的那种要素对社会福利的影响。于是互补松弛条件在经济学上的意义就非常直观:如果不充分使用现有的资源是最优的,那么这种资源的边际增量一定是没有价值的。下一章中我将更详尽地展开这一思想。

习题

习题 3.1:配额

假定一个消费者的效用函数如下:

$$U(x_1, x_2, x_3) = \alpha_1 \ln(x_1) + \alpha_2 \ln(x_2) + \alpha_3 \ln(x_3) \tag{3.21}$$

其中 α_j 是相加为 1 的正常数。他的预算约束为：

$$p_1x_1 + p_2x_2 + p_3x_3 \leqslant I$$

此外，该消费者还面临着一个配额约束：他购买的商品 1 不能超过 k 单位。

求解这个最大化问题，在什么样的参数条件下，配额约束会是紧的？

证明：当配额约束为紧时，消费者想花在商品 1 上，但是又不能花的这部分收入会以 $\alpha_2 : \alpha_3$ 的比例分配到商品 2 和商品 3 上。你觉得对购买面包的配额约束会同样地影响对黄油和大米的需求吗？效用函数(3.21)式的什么性质会产生这样的结果？你觉得面包—黄油—大米的例子同这个题目的情形有何不同？

习题 3.2：嫉妒性消费者之间的分配

假定社会中有一笔总量固定为 Y 的商品可供配置。有两个消费者，她们彼此嫉妒。如果消费者 1 得到 Y_1，消费者 2 得到 Y_2，那么她们的效用函数分别为：

$$U_1 = Y_1 - kY_2^2, U_2 = Y_2 - kY_1^2$$

其中 k 为正的常数。商品的配置必须满足 $Y_1 + Y_2 \leqslant Y$，并最大化 $U_1 + U_2$。

证明：如果 $Y > 1/k$，那么在最优解处资源约束一定是松的。解释这个结果。

习题 3.3：投资配置

现有一笔总量为 C 的资本可以在 n 个项目间进行分配。如果非负数量的资本 x_j 分配给了项目 j，$j=1,2,\cdots,n$，那么项目的投资组合的期望收益为：

$$\sum_{j=1}^{n}\left[\alpha_j x_j - \frac{1}{2}\beta_j x_j^2\right]$$

我们选择资本的分配额以使期望收益最大。

利用库恩—塔克条件找出一阶必要条件。定义：

$$H = \sum_{j=1}^{n}(\alpha_j/\beta_j), K = \sum_{j=1}^{n}(1/\beta_j)$$

证明：

(i) 如果 $C>H$，那么总资金中有一部分未被使用。

(ii) 对所有的 j，如果有 $\alpha_j>(H-C)/K$，那么每个项目都会得到一些资金。

(iii) 如果有某个项目未得到资金，那么它对应的 α 一定比任何获得资金的项目小。

▶ 进一步阅读

对更一般的拉格朗日和库恩—塔克定理，上一章提到的 Intriligator (1971，见本书第 2 章中的进一步阅读）一书的第 4 章给出了更为严格的分析。

拟线性偏好及其应用在 Hal Varian，*Microeconomic Analysis*，New York：Norton，2nd edn.，1984，pp. 278—83 上有更详细的讨论。

关于配额的一个较好的现代的分析在：Peter Neary and Kevin Roberts，"The Theory of Household Behavior under Rationing"，*European Economic Review*，13 (1980)，pp. 25—42；他们使用了本书到第 8 章才会展开的方法，到那时再回头看这篇文献将会非常值得。

4 影子价格

4.1 比较静态分析

拉格朗日方法及其在第3章中的扩展和一般化，为每一个约束都引入了一个值未定的乘子。这些乘子的值作为解的一部分被求解出来。第1章中对消费者选择问题的富有启发性的讨论，为其中的拉格朗日乘子提供了一个经济学上的解释：它正是收入的边际效用。在第2章和第3章中，我也曾提示过，对于约束条件下的最大化问题，类似的解释在更一般的意义上也成立。这正是本章的主题。

在约束条件下求最大化的问题有几个作为数据的参数。在满足 $G(x)=c$ 的约束下最大化 $F(x)$ 的问题中，参数 c 就是一个明显的例子。还有其他一些出现在函数 F 和 G 的定义中的参数，例如在第2章的例题和习题中出现的权重 α、β 和价格水平。经济学家们经常需要了解，如果这些参数取不同的数值，最大化问题的解会发生怎样的变化。在消费者理论中，我们通过比较不同预算约束下的最优选择来讨论价格变动的收入效应和替代效应。在厂商的生产和供给理论中，一个厂商的边际成本，是当厂商对每一种产出水平都选择成本最小的投入组合时，生产两种不同的产出水平的成本的差额。比较最优解如何随着参数的变动而变动的一般方法被称为比较静态分析。而事实上拉格朗日乘子的重要性在于，它们为一个非常重要的比较静态问题提供了答案。

4.2 等式约束

让我们从第 2 章中的简单分析框架开始，两个选择变量(x_1, x_2)，一个目标函数 $F(x)$和一个约束等式 $G(x) = c$。令 $\bar{x}$ 代表最优选择，$v = F(\bar{x})$表示最高能达到的值。现在假定 c 增加一个无穷小量 $\mathrm{d}c$。令$(\bar{x} + \mathrm{d}\bar{x})$为新的最优选择，$v + \mathrm{d}v$ 为新的最优值。

注意此处 $\mathrm{d}x$ 的用法和第 2 章中的用法略有不同。第 2 章的目的是检验 $\bar{x}$ 是否为最优解，因此我们考虑它的一个任意偏离 $\mathrm{d}x$，从而得到在最优解 $\bar{x}$ 处成立的一阶必要条件。但是此处的增量 $\mathrm{d}\bar{x}$ 不是任意的，它是选择变量随参数的微小变动而作出的最优的微小变动。

对于这些微小的变动，我们可以将 F 和 G 值的变化进行一阶泰勒近似。我们有：

$$
\begin{aligned}
\mathrm{d}v &= F(\bar{x} + \mathrm{d}\bar{x}) - F(\bar{x}) \\
&= F_1(\bar{x})\mathrm{d}\bar{x}_1 + F_2(\bar{x})\mathrm{d}\bar{x}_2 \\
&= \lambda[G_1(\bar{x})\mathrm{d}\bar{x}_1 + G_2(\bar{x})\mathrm{d}\bar{x}_2] \\
&= \lambda[G(\bar{x} + \mathrm{d}\bar{x}) - G(\bar{x})] \\
&= \lambda[(c + \mathrm{d}c) - c] = \lambda \mathrm{d}c
\end{aligned}
$$

在各步推导中，第二行和第四行是泰勒近似，第三行使用了一阶条件(2.6)式，第五行使用了约束(2.1)式，结果现在可以写为：

$$\mathrm{d}v/\mathrm{d}c = \lambda \tag{4.1}$$

所以，乘子 λ 就是目标函数能达到的最大值的变动与约束式子右边参数的变动的比例。现在我们可以将第 1 章中收入的边际效用看作是该更一般的结论的一个特殊情形。

多个选择变量和多个约束等式的情形并不会增加难度。若采用向量矩阵的符号，实际上的推导完全一样。回顾第 3 章中第一部分。令向量约束的右边变动 $\mathrm{d}c$，把 $\mathrm{d}\bar{x}$ 记作由此导致的最优解向量 $\bar{x}$ 的变动，那么

$$
\begin{aligned}
\mathrm{d}v &= F(\bar{x} + \mathrm{d}\bar{x}) - F(\bar{x}) = F_x(\bar{x})\mathrm{d}\bar{x} \\
&= \lambda G_x(\bar{x})\mathrm{d}\bar{x} = \lambda[G(\bar{x} + \mathrm{d}\bar{x}) - G(\bar{x})] = \lambda \mathrm{d}c
\end{aligned}
$$

让我们暂停一会儿来检验一下各个向量和被乘矩阵的维度。例如在第二行的第一个表达式中，λ 为 m 维行向量，$G_x(\bar{x})$ 为 $m\times n$ 的矩阵，$d\bar{x}$ 为 n 维列向量。最后的结果为行向量 λ 和具有相同维度 m 的列向量 dc 的乘积，因而它为标量。事实上它是两个向量的内积：

$$\lambda dc = \sum_i \lambda_i dc_i$$

这个结果非常重要，值得单独陈述以便日后参照：

拉格朗日乘子的解释：如果 v 为 $F(x)$ 在向量约束 $G(x) = c$ 下的最大值，λ 为对应约束的乘子组成的列向量，那么由无穷小量变动 dc 引起的变动 dv 由下式给出：

$$dv = \lambda dc \tag{4.2}$$

应该强调的是，如果 c 的变动较大，那么(4.2)式仅给出了 v 变化的一阶部分或线性近似。对于 c 的较大变动，我们可以泰勒展开到高阶，找到一个更好的近似。在第 8 章中我们将这样做，尽管目的并不相同。

4.3 影子价格

为了描述和解释(4.2)式，考虑一个计划经济，它要选择一个生产计划 $\bar{x}$ 以最大化社会福利函数 $F(x)$。这个计划要求资源向量 $G(x)$，这些资源的现有数量为向量 c。假定这个问题已经被求解出来，则拉格朗日乘子向量 λ 为已知。现在假定经济以外的某个力量，对第一种资源(比如说劳动)投入一个小的增量 dc_1，由其处置。那么最大化问题就可以在新的劳动约束下求解以确定一种新的生产方式。但是无需计算，我们就可以知道由此导致的社会福利的增加值，即 $\lambda_1 dc_1$。于是我们认为，乘子 λ_1 就是这个经济中以社会福利单位来衡量的劳动的边际产出。这显然是经济学上的一个极其重要的信息，也是拉格朗日方法及其乘子在经济学中如此重要的原因。

如果仅有一种稀缺投入，那么第 1 章中的方法的论述就能得到另一种非常有益于理解这个结果的方式。假定我们使用额外的劳动投入来提高某一商品的产量，比如说商品 j，而保持其他所有商品的产量不变。由于我们假定了在两种情况下劳动都被充分利用，那么被选产品产量的增加 $d\bar{x}_j$ 一

定满足：

$$G_j^1(\bar{x})\mathrm{d}\bar{x}_j = \mathrm{d}c_1 \text{ 或 } \mathrm{d}\bar{x}_j = \mathrm{d}c_1/G_j^1(\bar{x})$$

由此带来的社会福利的增加为：

$$F_j(\bar{x})\mathrm{d}\bar{x}_j = [F_j(\bar{x})/G_j^1(\bar{x})]\mathrm{d}c_1$$

最优化时的条件(2.5)式表明，方括号中的比例对于所有的 j 都是相同的。因而劳动供给的边际增加对社会福利的影响与这些额外的劳动如何被使用是无关的。这就是为什么我们可以毫不含糊地提及劳动的边际产量。

现在假定这些额外的劳动必须支付一定的成本才能使用。如果用社会福利单位来衡量，那么这个社会愿为每一边际单位的 c_1 支付的最高成本就是 λ_1。任何低于这个值的支付将使社会从使用额外劳动中获得一个正的净收益；反之，超过这个值，成本就会超过收益。从自然意义上说，拉格朗日乘子就是计划者为劳动服务制定的需求价格。以社会福利单位来衡量的价格看上去或许有些奇怪，但稍做修改它就会变为我们熟悉的情况。考虑其他某种资源，比如土地，并把它记作资源 2。现在假定提供给这个经济额外的劳动 $\mathrm{d}c_1$，但是要求 $\mathrm{d}c_2$ 单位的土地作为回报。在这一交易中的社会福利净收益为$(\lambda_1\mathrm{d}c_1-\lambda_2\mathrm{d}c_2)$，因而社会计划者最多愿意放弃的土地为$(\lambda_1/\lambda_2)\mathrm{d}c_1$。于是同样很自然地，我们可以把比例$(\lambda_1/\lambda_2)$称为以土地单位来衡量的每单位劳动的需求价格。从微观经济学理论中得知，是相对价格而非绝对价格主宰着市场交易；类似地，我们也可以认为，正是各个不同资源对应的拉格朗日乘子的相对大小，决定着计划者用一种资源交换另一种资源的意愿。

如果邻近的经济体因劳动和土地两种资源的相对稀缺程度或技术的差异而面临着不同的权衡，那么这两个经济体之间就有进行互利的要素服务贸易的可能性。(即使要素服务不能贸易，用这些要素生产的商品的贸易也能确保获得这种互利收益的部分甚至全部，不过讨论其中的细节会离题太远以至到国际贸易理论中去了。)

当然，经济体的内部组织并不需要与价格有关，劳动约束的拉格朗日乘子也不必同实际支付给每人每小时的工资相等。在一个指令性的经济中，劳动可能只是简单地被直接分配到各个任务中去。(这一做法有很严重的概念和操作问题，正如绝大部分苏维埃体制的经济体已经出现的问题那样，

不过那又是另一回事了。)不过生产计划隐含地对各类资源赋予了价值,如果计划者能够注意到拉格朗日乘子反映的这些价值,那么他就能更好地理解这个经济体及其可能出现的瓶颈。

现在考虑一个用市场来分配资源的经济体。在均衡时,价格水平使由个人通过求解他们各自的约束条件下的最大化问题而选择的需求和供给,在总量上相等。现在假定有一个经济学家打算用某个给定的准则来评估这个经济体的绩效。为了得到一个可进行比较的标准,他会求解一个计划问题,即在该经济体的资源禀赋、技术和信息传递的约束下,最大化那个准则函数。求出来的解中将包括一个受资源约束的拉格朗日乘子向量。

你也许会觉得有些奇怪,为什么市场经济会复制计划配置,同样奇怪的是,为什么拉格朗日乘子会和市场价格有关。不过确实有一些重要的情形,其最优解可以由市场来复制,其拉格朗日乘子同资源的市场价格成比例:资源的相对价格等于对应的乘子的比例。在这些情形下,该经济学家就会说,这个经济体受一只“看不见的手”引导,进而到达他所计划的最优解。例题4.1会详细地解释这样的情形,但它依赖于很多特别的假设,而这些假设又常常受到质疑。因此,绝大部分的现代经济理论所关心的是,当这些假设无法得到满足时会发生什么。但是作为所有这些理论分析的起点,以及由于许多人信仰市场机制的最优性这样一个现实问题,这个情形又具有极端的重要性。因而它值得深入的研究。

为了引出同价格的联系,同时又为了和市场价格在概念上有所区别,拉格朗日乘子常被称为影子价格。

4.4 不等式约束

现在产生了一个经济学问题:我们认为价格为非负,但是到目前为止我们还没有认为影子价格(拉格朗日乘子)也应该为非负的理由。在上面计划经济的例子中,拉格朗日乘子衡量了稀缺资源禀赋的增加带来的社会福利的增加。拥有更多的某种资源使得以前所有的生产计划依旧可行并增加了新的机会。这使得计划者至少可以取得和原来一样高,而且在绝大多数情况下更高的福利水平。同样地,在更为一般的约束条件下的最大化问题中,放松某个约束应该是一件有利的事情。我们能用数学证明这一直觉吗?

困难在于，在具有等式约束的问题的一般形式中，约束等式右边的数值增加并不必然意味着约束的放松。最简单的例子是，我们可以把约束 $G^i(x)=c_i$ 写作 $-G^i(x)=-c_i$，那么在这种新的形式下，等式右边的数值增加就意味着资源 i 的数量 c_i 的减少。此外，也并非所有的约束都是表示资源禀赋的约束。例如，我们可能想在确保某些消费品的某一最低可接受的供应的基础上，最大化投资的量。现在，这一已规定好的最低水平的上升会使经济预算变紧，因此有一小部分的投资会被挤出，同时拉格朗日乘子为负。在这种情形中，拉格朗日乘子就像一个转换函数(将消费转换成投资)的斜率。我们应该可以预期这样的一条曲线是向下倾斜的，并把负的斜率解释为影子价格。

这些例子都说明，如果我们想要得到非负的影子价格，我们必须小心地将约束写成某种形式，在这种形式下，式子右边数值的增加的确放松了对这个选择变量的限制。

还有另一种更重要的情形需要考虑，即可能存在超过某一点后资源的边际价值为负的情形。例如，太多的人在一起反而互相影响。在这种情况下，劳动资源数量的进一步增加将意味着目标函数的最大值的下降和一个负的拉格朗日乘子。但在这种情形下，不过量地使用这些资源将会更好，即使它们可供使用。习题 3.2 就描述了这样的一种情况，当嫉妒效应呈压倒性时，扔掉一部分商品反而是最优的。从数学上看，等式约束要求用掉所有可供使用的数量。但如果约束是不等式，如 $G^i(x)\leqslant c_i$，那么我们就可以让一部分资源闲置，如果这种闲置有利于最大化目标。

在现实中，让某些资源闲置可能会带来一定的成本。失业从全社会的角度来看可能被认为是不受欢迎的；某些资本，特别是智力资本，如果不用就会生锈。在这样的情况下，我们应该把这些成本也包括到目标函数中去。如果目标函数已包含这类成本，那么从经济学角度来看，就没有理由不让一部分资源禀赋闲置，只要这种闲置会带来更好的结果。

在市场价格中也有一个与以上讨论非常相似的情形。如果有些商品确实是“坏”的，我们就预期它的价格为负。更为一般地，正是自由处置的假定确保了非负的市场价格。

我们在第 3 章中论述的库恩—塔克定理给出了在不等式约束下求解最大化问题的一阶必要条件。它直接证实了这一经济学直觉。条件(3.10)式

表示的是拉格朗日乘子向量非负。它得到了一个非常重要的深入的结论。向量不等式 $\lambda \geqslant 0$ 和 $L_\lambda(\bar{x}) \geqslant 0$ 满足互补松弛条件，而后者为 $G(x) \leqslant c$ 的另一种写法。对于每一个 i，这一对不等式

$$G^i(x) \leqslant c_i,\ \lambda \geqslant 0$$

中至少有一个不等式是以等式的形式成立的。如果资源 i 未被充分利用，那么它的影子价格就是零；有正的影子价格的资源一定被充分利用。

上述论点支持并完成了对影子价格作为资源的边际值产品(marginal value products)的解释。如果某种资源已部分闲置，那么新增的这种资源仍将闲置。此时目标函数的最大值不变，影子价格将为零。而另一方面，正的影子价格意味着现有资源的一单位的边际增加可以得到很好的利用。也就是说，原来的生产计划中这种资源没有被闲置。你可以回到例题 3.2 中对技术性闲置的论述，用上述观点来检验那里的结论。

有一个需要注意的难点。假定 c_i 是这样的一个点，在这点上资源 i 正好处于即将过剩的边际上，即已有的资源恰好被充分利用，但是该资源任何的增量都不会再被使用。互补松弛条件不能告诉我们在这一点处，拉格朗日乘子为正还是为零。事实上，答案视具体的问题而定，它取决于作为 c_i 的函数的最大值 v 的斜率，在边界点处是平滑地变为零还是突然变为零。

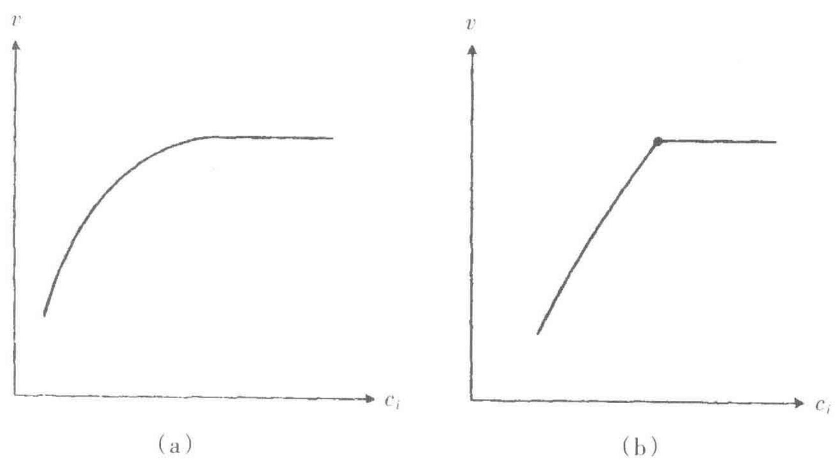

图 4.1　资源数量和影子价格

图 4.1 给出了上述两种可能性。在图(a)中斜率平滑地下降，即在该点处的拉格朗日乘子为零。在图(b)中斜率突然下降，即任何介于左边曲线的

斜率和右边直线的斜率为零之间的 λ_i 都有可能代表影子价格恰好从正转变为零的那一点。这一现象会在线性规划的背景下出现(见例题 7.1)。

例题

例题 4.1:看不到的手——分配

考虑计划的阶段,此时各种商品的产量已经知道,惟一剩下的问题就是如何将这些商品在消费者间进行分配。设有 C 个消费者,标以 $c=1,2,\cdots,C$,并设有 G 种商品,标以 $g=1,2,\cdots,G$。用 X_g 表示商品 g 固定的总量,x_{cg} 表示分配给消费者 c 的这种商品的量。每一个消费者的效用都是一个仅包含他自己分配到的商品的函数:

$$u_c = U^c(x_{c1}, x_{c2}, \cdots, x_{cG}) \tag{4.3}$$

社会福利函数是一个关于这些效用水平的函数:

$$w = W(u_1, u_2, \cdots, u_C)$$

计划者面临的约束是,对于每一种商品,分配给每一个消费者的数量加起来不超过该商品现有的总量。当这些效用函数和社会福利函数都是增函数时,很显然,没有商品会被浪费,所以我们可以把约束写成等式:

$$x_{1g} + x_{2g} + \cdots + x_{cg} = X_g, g = 1, 2, \cdots, G \tag{4.4}$$

应用拉格朗日定理,令 π_g 表示对应于商品 g 的约束条件的拉格朗日乘子,构造拉格朗日函数:

$$L = W(U^1(x_{11}, \cdots, x_{1G}), \cdots U^c(x_{c1}, \cdots, x_{cG})) + \sum_g \pi_g [X_g - \sum_c x_{cg}]$$

为简便起见,L 的自变量以及约束中求和的范围都省略不写了。

在推导一阶条件时,我们必须使用链式法则让 L 对每一个 x_{cg} 求导,从而得到:

$$(\partial W/\partial u_c)(\partial U^c/\partial x_{cg}) - \pi_g = 0 \tag{4.5}$$

同样,所有的偏导数都在最优解处取值;为了方便我也将它们省略了。拉格朗日乘子 π_g 也作为解的一部分而得到。

现在假定 π_g 被作为商品的价格。给予每一个消费者 c 一定的货币收

入 I_c，允许他选择消费向量以使他在预算约束下最大化自己的效用(4.3)式：

$$\pi_1 x_{c1} + \pi_2 x_{c2} + \cdots + \pi_G x_{cG} = I_c \tag{4.6}$$

最优化可以由如下条件描述：

$$\partial U^c / \partial x_{cg} = \lambda_c \pi_g \tag{4.7}$$

对于所有的 g 和 c 都成立。和以往一样，λ_c 为消费者 c 的收入的边际效用。

比较(4.5)式和(4.7)式，我们发现，如果对于所有的 c，令

$$\partial W / \partial u_c = 1/\lambda_c,\ \text{或者}(\partial W / \partial u_c)\lambda_c = 1 \tag{4.8}$$

那么两个式子就是完全一样的。而且，通过调整货币收入 Ic 这可以实现。(4.8)式中第二个等式的左边，正是给消费者 c 一单位收入对社会福利的边际影响，它是对消费者 c 自身效用的边际影响乘以一单位他的效用对社会福利的影响。

换句话说，收入的分配应该被这样安排，以使得在边际上，每一个消费者收入的社会价值都是相等的。只要做到这一点，他们就可以自由地选择自己的消费束。这就是在分配问题中“看不见的手”的作用结果。

仅比较了一阶条件的论证不是十分的严格，而且还存在更好的证明。但重要的是，我们要认识到导出这一结果的关键假设，即每一个消费者的效用仅依赖于他自己消费的数量。如果一个消费者的效用取决于其他人的消费，这种现象被称为“外部效应”或者“外部性”。这种效应会扰乱通过价格机制而实现的简单的分散消费机制。其实质是，我们向每一个消费者索要的价格，不仅包括他自己消费的稀缺价值，还包括他的消费给其他人的效用带来的损失(或者收益，当他的消费给其他人带来好处时，其他人必须向他支付报酬)。这样的价格就要依人而定了，因而市场机制的实施就变得更加复杂。

例题 4.2：免税的购买

让我们从中央计划和收入分配的沉闷情形转到一个到处旅游的富翁的消费决策上来。他可以在他家乡的商店里购买各种不同牌子的酒，也可以在他经过的各个机场的免税商店里购买。免税商店中的价格便宜一些，但是他所能购买的总量受到他的国家海关法规的限制。

假定有 n 种牌子，p 是他家乡的各类牌子的酒的价格行向量，q 是免税商店里的价格行向量，对于所有品牌的酒而言，免税商店里的价格都较低：$q \ll p$。令 x 表示他在家乡购买的各种酒的数量列向量，y 是在免税商店里购买酒的数量列向量。假定他在一年中旅游和享乐的次数足够多，以至于我们可以将数量视作一个连续变量，而不受整瓶数的限制。当然，有整数约束的问题也会给出类似的结果，但需要不同的技巧。

为了简化问题，我们暂不考虑在作为一个整体的酒的消费和所有其他商品消费之间的选择问题。这样，我们将他已决定在这一年中花在酒上的收入看作不变，比如 I。那么预算约束变为：

$$px + qy \leqslant I \tag{4.9}$$

假定在这一年中我们的富翁被允许进口 K 瓶免税酒，这个约束可以写作：

$$y_1 + y_2 + \cdots + y_n \leqslant K$$

为了简化公式，令 e 表示每一个分量都为 1 的行向量，那么上式的左边即为乘积 ey，这个约束可以写作：

$$ey \leqslant K \tag{4.10}$$

除非这个消费者在他的免税限额内就已经满足了（好像是个不太可能的情形），这两个约束才将以等式形式成立。但是我们可以预期，有些品牌仅在免税商店购买而另一些品牌仅在家乡的商店购买才是最优的。因而我们必须记得 x 和 y 受到非负约束。

效用只依赖总的消费数量 $c=x+y$。因而拉格朗日函数为：

$$L = U(c) + \lambda[I - px - qy] + \mu[K - ey]$$

一阶条件由第 3 章中非负变量的拉格朗日定理给出。对于每一个 j，我们都有：

$$\partial L/\partial x_j \equiv \partial U/\partial c_j - \lambda p_j \leqslant 0, x_j \geqslant 0 \tag{4.11}$$

满足互补松弛条件，同时，

$$\partial L/\partial y_j \equiv \partial U/\partial c_j - \lambda q_j - \mu \leqslant 0, y_j \geqslant 0 \tag{4.12}$$

也满足互补松弛条件。

若仅通过这些不等式，在 2^{2n} 种等式和零的情形中，把最优解系统地挑出来似乎不太可能。但是借助经济学直觉的搜寻很快就找到了解。

能有某一种牌子的酒 j 同时在两种商店里以正的数量被购买吗？如果是的话，从(4.11)式和(4.12)式得到的方程为：

$$\partial U/\partial c_j - \lambda p_j = 0 = \partial U/\partial c_j - \lambda q_j - \mu$$

或

$$\lambda p_j = \partial U/\partial c_j = \lambda q_j + \mu \tag{4.13}$$

除了巧合，最多有一个 j 使上式成立。假如有两个 j 使上式成立，比如 $j=1$ 和 2，那么我们会有

$$\lambda(p_1 - q_1) = \mu = \lambda(p_2 - q_2)$$

由于我们假定该消费者没有满足，λ 就是正的，所以

$$p_1 - q_1 = p_2 - q_2$$

对于给定的价格，这种情况偶尔才会发生。上面的分析不仅缩小了我们搜索的范围，而且告诉我们，两组价格之间的绝对差异同最优的购买选择有很大的关系。

现在假定牌子 j 的酒仅在家乡的商店里被购买。由 $x_j>0$，$y_j=0$，我们得到：

$$\partial U/\partial c_j = \lambda p_j \tag{4.14}$$

和

$$\partial U/\partial c_j \leqslant \lambda q_j + \mu \tag{4.15}$$

其中，(4.14)式是我们熟知的，不过即便如此，也有必要对它进行重新解释。等式的左边恰是牌子为 j 的酒的边际效用。等式的右边为在家乡商店购买它的边际机会成本：这样做会占用 p_j 的收入，这部分收入就不能用于其他东西的购买了，而这么多收入的效用价值为 λp_j。

这反过来给解释(4.15)式提供了启示。式子的右边是在免税商店购买一单位 j 的边际机会成本，它要求 q_j 单位的收入具有 λq_j 的效用价值。但它也用掉了一单位的免税限额，而一单位免税限额的影子价格为 μ。因此，总的机会成本就为这两部分的和。如果某个牌子的酒没有在免税商店里被购买，那一定是因为这样做的机会成本超过了消费它得到的边际效用。

现在，消费者的购买原理很清楚：在有较低的机会成本的商店里购买每一种牌子的酒。注意到

$$\lambda q_j + \mu < \lambda p_i，当且仅当\ p_i - q_i > \mu/\lambda$$

因而，富翁会把各个牌子的酒按照它们在两种商店里的绝对价格差进行排序，那些绝对价格差最大的牌子的酒会在免税商店里购买，而那些价格差最小的就会在家乡商店里购买。两者的交会点在恰好用完免税限额的那一点。所以最多只有一种牌子的酒会同时在两种商店里被购买。

顺带说，如果免税限额限制的是购买的总价值 qy 而不是数量 ey，那么解也是相似的，但对牌子排序的依据是相对价格而不是绝对价格。我把这种情况作为练习留给读者。

习题

习题 4.1：看不见的手——生产

继续沿用例题 4.1 中的表示法，不过我们现在允许生产商品。假定有 F 种投入要素，数量固定分别为 Z_f，$f=1,2,\cdots,F$。如果用在商品 g 上的生产要素 f 为 z_{fg}，那么产出 x_g 就由生产函数给出：

$$Xg = \Phi^g(z_{1g}, z_{g2}, \cdots, z_{Fg}) \tag{4.16}$$

把这些约束加到先前的问题中去。验证最优分配的一阶条件和原来一样，但是增加了最优要素配置的新条件。解释拉格朗日乘子的含义。生产是否可以分散化，从而使每个企业都只生产一种产品？证明：分配给消费者的收入 I_c 加起来正好等于总产出的价值。

练习 4.2：看不见的手——要素供给

现在令要素供给 Z_f 也成为最优化的一部分。假定每一个消费者 c 提供的要素 f 为 z_{cf}。这些要素的数量会负面地影响他的效用，即提供要素会有负效用。

找出一阶条件。解释拉格朗日乘子的含义，并讨论如何在市场机制下实施这个最优解。

现在你必须区分消费者的两种收入的来源：他们从提供要素服务获得的收入和从政府获得的一次性金额 I_c。为获得条件(4.8)式，这些一次性金额必须是不同的(因人而异)。证明：分配给消费者的一次性金额的总和等于生产创造的总利润，也就是产值减去支付给要素的报酬。

练习 4.3：借与贷

考虑一个消费者要计划他两年中的消费。他在第一年有收入 I_1，第二年有收入 I_2。在每一年都有两种商品可供消费。在第一年，商品的价格分别是 p_1 和 q_1，相应的消费数量为 x_1 和 y_1。在第二年，类似地，我们有 p_2、q_2 和 x_2、y_2。效用函数为：

$$u_1 = \alpha_1 \ln(x_1) + \beta_1 \ln(y_1) + \alpha_2 \ln(x_2) + \beta_2 \ln(y_2)$$

在两个预算约束(每年有一个)下最大化这个效用函数。

求解这个问题，找出两个约束的拉格朗日乘子。检验它们是如何依赖于货币收入和这个问题中的其他参数的。

如果消费者在第一年放弃了 dI_1 的收入，那么他在第二年要求多获得多少收入呢？换句话说，需要怎样的回报率才能使他在第 1 年储蓄一些收入？你可以期望在这样的消费者组成的经济体中会有借贷机构产生吗？那么由什么理由决定谁会成为借款人，而谁又会成为贷款人呢？

▶ 进一步阅读

一个关于价格在分散化的资源有效配置中的作用的精彩讨论，可以看：Tjalling C. Koopmans, *Three Essays on the State of Economic Science*, New York: McGraw-Hill, 1957。

▶5 最大值函数

在上一章中我们引入了比较静态分析的概念,并利用这个概念,把拉格朗日乘子解释为目标函数可达到的最大值与约束等式的右边参数的变化率。同时,目标函数和约束等式还包括许多其他参数,而目标函数可达到的最大值也依赖于所有的这些参数。在前面章节使用的方法可以被稍加改变,以帮助我们理解这种更为一般的依赖性的本质。这正是本章的目标。和前面一样,我将从所有约束都是等式的情形开始,然后再考虑不等式的情形。

5.1 目标函数中的参数

首先来考虑参数只影响目标函数的情形。一个常见的例子就是生产者选择一组投入要素的组合,以最小的成本生产给定的产量。投入价格是影响他的目标函数的参数。但这一约束,即所选择的投入要素应该生产出想要的产量,仅涉及生产函数而不涉及价格。另一个例子是,考虑一个小国选择其生产方式以最大化用世界价格来衡量的全国的产出,这些价格就是影响目标函数的参数。更为一般地,假定参数向量 θ 进入目标函数,我们需要选择 x,在通常的向量约束 $G(x)=c$ 下最大化目标函数 $F(x,\theta)$。拉格朗日函数的一阶必要条件(3.5)式和(3.6)式稍加修改后依然成立——我们现在发现函数 F 和拉格朗日函数依赖于 θ。因此:

$$L(x,\lambda,\theta)=F(x,\theta)+\lambda[c-G(x)]$$

最优解 $\bar{x}$ 满足一阶条件

$$L_x(\bar{x},\lambda,\theta)=0, L_\lambda(\bar{x},\lambda,\theta)=0$$

再次把 v 记作目标函数的最大值。假定 θ 变动到 $(\theta+\mathrm{d}\theta)$。相应地，让最优解 $\bar{x}$ 变到 $(\bar{x}+\mathrm{d}\bar{x})$，则最大值变到 $(v+\mathrm{d}v)$。采用与第 4 章一样的一阶泰勒近似，我们可以得到 $\mathrm{d}v$ 的一个表达式：

$$\begin{aligned}
\mathrm{d}v &= F(\bar{x}+\mathrm{d}\bar{x},\theta+\mathrm{d}\theta)-F(\bar{x},\theta)\\
&= F_x(\bar{x},\theta)\mathrm{d}\bar{x}+F_\theta(\bar{x},\theta)\mathrm{d}\theta\\
&= \lambda G_x(\bar{x})\mathrm{d}\bar{x}+F_\theta(\bar{x},\theta)\mathrm{d}\theta\\
&= F_\theta(\bar{x},\theta)\mathrm{d}\theta
\end{aligned}\tag{5.1}$$

在上述各步计算中，第二行到第三行的推导使用了拉格朗日一阶条件，最后一步推导利用了在 θ 变化的过程中 G 的值依旧等于 c 的事实。

同样地，当 θ 的变动大到使一阶近似无效，我们就可以把级数展开到更高阶以找到 v 变动的更好近似。不过上述结果因其简洁性而具有重要的意义。它说明了，当一个不影响约束条件的参数发生变化后，如果要找出由此参数的变化导致的目标函数最大值的一阶变化，我们无需担心最优选择 $\bar{x}$ 自身同时发生的变动。我们所需要做的，是在原来的最优选择处计算这个参数变动的偏导数。

上面提到过的成本最小化的例子可以很好地说明这一点。假定 x 是投入要素向量，G 是生产函数，c 是所要求的产出数量。令 θ 表示投入要素的价格行向量。那么生产者希望最小化 θx，也就是，

$$\text{最大化 } F(x,\theta)=-\theta x\text{，同时必然满足 } G(x)=c$$

将由此产生的最大值记为 $(-v)$，那么(5.1)式就变为：

$$\mathrm{d}(-v)=\mathrm{d}\theta F_\theta(\bar{x},\theta)=-\mathrm{d}\theta\bar{x}$$

注意 $\mathrm{d}\theta$ 为行向量，所以我们把 F_θ 解释为一个列向量，并把内积写作上面的形式。结果是

$$\mathrm{d}v=\mathrm{d}\theta\bar{x}\tag{5.2}$$

现在 v 就是当投入要素的价格为 θ 时，生产产出 c 的最小成本。当投入要素的价格发生变动时，生产者就会改变他的投入组合，少用变得相对较贵的要素并多用其他要素。即他会沿着一条等产量曲线进行投入要素的相互

替代。但是(5.2)式中的 $\bar{x}$ 是原来的参数向量 θ 下的最优选择，而不是 $\theta+\mathrm{d}\theta$ 或某个均值下的最优选择。换言之，成本的一阶变动只是原来最优解 $\bar{x}$ 下的成本变动，就好像由固定不变的系数规定的那样。也许，从另一个角度看待这个问题会有所帮助。如果我们写成 $v=\theta\bar{x}$ 并进行微分，得到：

$$\mathrm{d}v=\theta\,\mathrm{d}\bar{x}+\mathrm{d}\theta\,\bar{x}$$

右边的第一项为以原来的价格衡量的投入组合变动的价值。但是在那些价格下，原来的投入组合的选择为最优的，因而任何变动的价值的一阶效应必然为零。所以只留下了第二项，如(5.2)式所示的那样。

5.2 包络定理

图5.1为上面的代数推导给出了一个几何说明。对于 θ 某一特定的值，比如说 θ_1，假定最优的选择是 $\bar{x}^1$。两条曲线代表了两个关于 θ 的函数。一条是 $F(\bar{x}^1,\theta)$，其中当 θ 变动时 x 的值一直固定在 $\bar{x}^1$；另一条是将 v 和 θ 联系起来的最优值函数，其中 x 也可以随着 θ 的变动作出最优的变动。我们把这第二个函数正式定义为：

$$V(\theta)=\max_{x}\{F(x,\theta)\mid G(x)=c\} \tag{5.3}$$

读作"$V(\theta)$是在约束 $G(x)=c$ 下关于 x 的 $F(x,\theta)$的最大值"。接着将最优选择 $\bar{x}$ 本身也写作一个函数 $\bar{x}=X(\theta)$，那么我们就有：

$$V(\theta)=F(X(\theta),\theta)$$

两个函数 $V(\theta)$和 $F(\bar{x}^1,\theta)$在 θ_1 处重合，因为 $\bar{x}^1$ 正好是那里的最优选择。对于 θ 的其他值，除非 $\bar{x}^1$ 仍然为最优选择，否则，代表最优值函数的曲线位置就会比代表 $F(\bar{x}^1,\theta)$的曲线高(无论如何，也不会比它低)。因而这两条曲线应该在 θ_1 处相切，而这正是(5.1)式所表达的含义。

类似地，我们也可以画出 $F(\bar{x}^2,\theta)$的图形，其中 $\bar{x}^2$ 是在 θ_2 处的最优选择。这条曲线将在 θ_2 处碰到最优值函数 $V(\theta)$的图形。事实上，我们可以对 x 的固定值的整个范围画出 $F(x,\theta)$整个族的曲线，而每一个 x 对应于每一个 θ 都是最优的。这族曲线中的任何一条曲线都不会与曲线 $V(\theta)$相交，每一条曲线都会与最优值函数相切于 x 恰为最优选择的 θ 值处。换句话说，

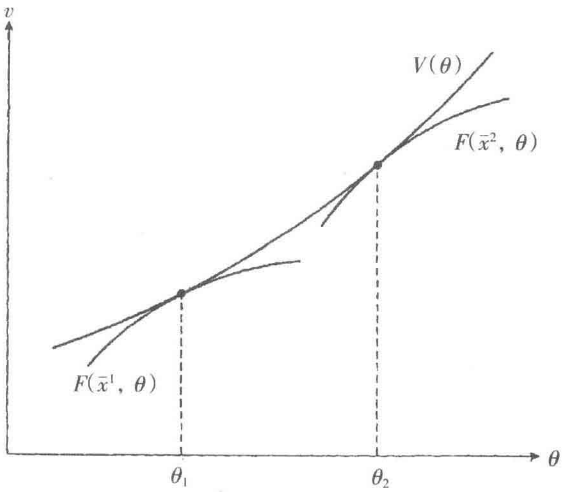

图 5.1　包络定理

最优值函数为值函数族的上包络线，且在每一个值函数中，选择变量都固定不变。这也就是公式(5.1)式经常被称为包络定理的原因。

例如，在成本最小化的应用例子中，令标量参数 θ 仅代表一种投入要素的价格。当投入数量的向量固定不变时，生产成本就是 θ 的一个线性函数。那么，作为 θ 的函数的、已最小化的成本就是所有这些直线的下包络线（不是上包络，因为这是一个最小化而非最大化问题）。图 5.2 给出了说明。

此处的关注点在于一阶或相切的性质：当上包络线碰到这族曲线中的一条时，两者必然相切。在成本最小化的例子中，令 x 代表那个价格 θ 正在

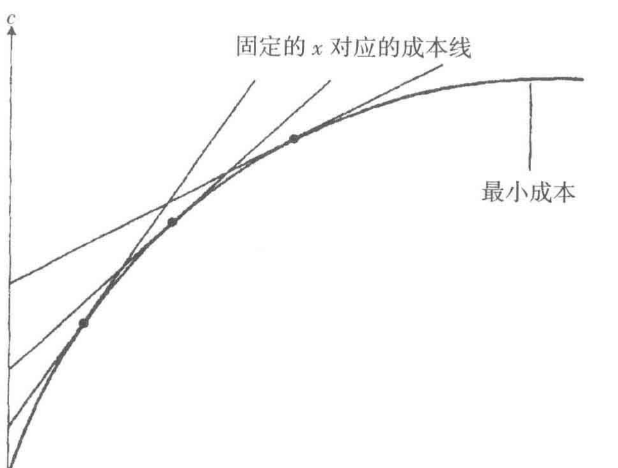

图 5.2　成本最小化函数

变动的投入要素的数量。每一条直线的斜率就等于沿这条直线的固定值 $\bar{x}$。这条直线与成本最小化函数相接触于 $\bar{x}$ 正好为最优选择的 θ 处。因而成本最小化函数在每一点处的斜率就是在那一点处的 x 的最优值。换句话说,成本最小化函数本身就包含了最优投入选择的信息。我们会在例题5.1中进一步说明这一思想。

二阶性质或者说曲率性质也是明显的。图 5.1 中每一个 $F(\bar{x},\theta)$ 为凹曲线而 $V(\theta)$ 为凸曲线。但是更为一般地,包络线必须比其所包络的那族曲线中的任何一条都更凸。这样,任何固定不变投入选择的成本函数是投入要素价格的线性函数,但它们的下包络线(成本最小化曲线)是凹的。第 8 章会更加详细地研究这一性质,并会导出一个重要的、称为勒夏特里艾一萨缪尔森原理的比较静态分析结果。

5.3 影响所有函数的参数

现在假定 G 和 F 一样也含有 θ。计算的过程和前面类似,除了 G 的变动不再为零。假定向量约束为 $G(x,\theta)=c$,其中 θ 和 c 是不同的。那么

$$G_x(x,\theta)\mathrm{d}x + G_\theta(x,\theta)\mathrm{d}\theta = 0$$

在前面一连串等式中应用这个式子,我们有

$$\begin{aligned} \mathrm{d}v &= -\lambda G_\theta(\bar{x},\theta)\mathrm{d}\theta + F_\theta(\bar{x},\theta)\mathrm{d}\theta \\ &= L_\theta(\bar{x},\lambda,\theta)\mathrm{d}\theta \end{aligned} \tag{5.4}$$

上式和(5.1)式的差异有一个直观的解释。当 θ 影响约束时,变动 $\mathrm{d}\theta$ 有直接的影响,即使 G 的值增加了 $G_\theta(\bar{x},\theta)\mathrm{d}\theta$。这相当于 c 的同等量的减少。拉格朗日乘子的解释告诉我们,这等价的 c 的减少会使 v 的值下降 $\lambda G_\theta(\bar{x},\theta)\mathrm{d}\theta$,这就是(5.4)式和(5.1)式相比时额外多出来的那一项。

在前一章中,关于参数 c 变动的一个类似的比较静态分析导出了(4.2)式,这个式子为我们提供了拉格朗日乘子的重要解释。而本章中更为一般的形式可以包含前面的情形。为了更清楚地说明这一点,定义一个更大的参数向量 $\hat{\theta}$,其包括子向量 θ 和 c。并将约束条件写作:

$$\hat{G}(x,\hat{\theta}) \equiv G(x,\theta) - c = 0$$

拉格朗日函数现在可以写为：

$$\hat{L}(x,\lambda,\hat{\theta}) \equiv F(x,\theta) - \lambda\hat{G}(x,\hat{\theta})$$

(5.4)式就变成：

$$\mathrm{d}v = \hat{L}_{\hat{\theta}}(\bar{x},\lambda,\hat{\theta})\mathrm{d}\hat{\theta}$$

将 $\hat{\theta}$ 中的子向量分开，

$$\hat{G}_{\hat{\theta}}(x,\hat{\theta})\mathrm{d}\hat{\theta} = G_{\theta}(x,\theta)\mathrm{d}\theta - dc$$

因而 $\mathrm{d}v$ 的表达式变为

$$\mathrm{d}v = L_{\theta}(\bar{x},\lambda,\theta)\mathrm{d}\theta + \lambda\mathrm{d}c$$

这将(5.4)式和(4.2)式作为特殊的情形包含在其中。

5.4 某些选择变量固定不变

当约束条件下的最优化问题中的参数 θ 发生变动时，最优选择向量 x 和最大值 v 也随之发生变化。但是我们已经知道，对 v 的一阶效应可以通过把 x 固定在原来的最优选择处计算，以得到 θ 对 v 的偏导数效应。如果这些参数的变动还影响到了约束条件，那么我们必须记得将等价的约束等式右边 c 的减少所带来的影响也包括进来，不过我们仍然可以忽略 x 的变化。

这暗含着一个一般化的思想。当只有 x 中的一部分分量调整到它们新的最优水平，而其他分量都固定在原来的最优水平时，对 v 的影响是否一样呢？这一类比较在经济学中很普遍。其中最著名的例子就是短期和长期的区别——在长期中可以最优变动的一些数量在短期中必须固定不变。

这一问题可以采用推导(5.1)式和(5.4)式时使用的同一微分方法来解决。不过，为包络性质提供直觉的几何分析方法可以非常简单地做到这一点。让我们从更加精确地论述这个问题来开始分析。

将向量 x 分解为两个子向量 y 和 z。把约束等式右边的参数也包括进参数向量中，就像上面解释的那样，将这个问题写作：

$$\max F(y,z,\theta), \quad \text{s.t. } G(y, z,\theta) = 0 \tag{5.5}$$

这一问题有两种情形：在长期中，y 和 z 都是选择变量；而在短期中，z 固定不变而仅有 y 可以变化。为了使后一情形有意义，约束条件的数量必须小于 y 的维度。

将长期的最优选择和由此得到的值函数写作 θ 的函数，也就是

$$y = Y(\theta), z = Z(\theta), v = V(\theta) \tag{5.6}$$

在短期中，z 应该被视作和 θ 一样的另一个参数，所以 y 的最优选择和由此得到的值函数 v 就是 (z,θ) 的函数，也就是

$$y = Y(z,\theta), v = V(z,\theta) \tag{5.7}$$

采用相同的符号 Y 和 V 来代表两种情形下不同的两个函数应该不会引起混淆，因为不同的变量会被恰当地表示出来。

由此立即得到最优化的定义：

$$V(\theta) \geqslant V(z,\theta) \text{ 对于所有的}(z,\theta)$$

如果 $z=Z(\theta)$ 为最优选择，上式为等式。那么正如图 5.1 所示，$V(\theta)$ 的图形是在 z 的可能的值的整个范围上作为 θ 的函数 $V(z,\theta)$ 表示的曲线族的上包络线。

如果这些函数是可微的，那么我们可以得出

$$V'(\theta) = V_\theta(Z(\theta),\theta) \tag{5.8}$$

其中，等式的右边就是短期最优值函数 $V(z,\theta)$ 在第一个变量 z 不变时，在点 $z=Z(\theta)$ 处取值的偏导数。

现在几何的方法提醒我们注意到微分方法可能隐藏的一个潜在的问题，即函数可能不可微。即使目标函数和约束函数 F 和 G 是任意平滑的，最优值函数 V 的斜率仍有可能发生突变。将第 4 章最后一部分对拉格朗日乘子的解释同图 4.1 联系起来，我们就会发现一个这样的例子。当对选择变量有不等式约束或者非负约束时，那么这些约束在参数的某些范围中可能为紧，而在另一些范围中可能为松。目标函数对参数变动的反应可能会不同，并且这种反应取决于紧的或松的约束结构。在约束结构发生变动的那一点处，即从紧的变到松的或者反过来，最大值函数的图形就会有一个折拐。

在许多实际应用中，我们可以不关心这样的结构变动而直接应用(5.8)

式，但它失效的可能性还是应该记在心里的。在某些情形下，如线性规划，参数值从一个紧的约束和松的约束的结合部变动到另一个必然会引起斜率的变化。

例题

例题 5.1：短期和长期成本

作为对包络定理的背景的一个说明，考虑如下生产函数的短期和长期成本曲线之间的关系

$$Q = (KL)^{1/\alpha} \tag{5.9}$$

其中 Q 表示产量，K 是短期固定资本，L 是劳动。如果 $\alpha=2$，那么规模报酬不变，如果 $\alpha<2$，那么规模报酬递增，如果 $\alpha>2$，那么规模报酬递减。

令 w 表示工资率，r 表示资本的使用者成本（其为租赁价格，如果资本服务是租赁的；或者为利率和折旧成本之和，如果资本设备是购买的）。长期的成本函数为：

$$C(w,r,Q) = \min_{K,L}\{wL + rK \mid KL = Q^{\alpha}\} \tag{5.10}$$

使用拉格朗日方法，可以很容易地得到使成本最小的投入选择：

$$K = (wQ^{\alpha}/r)^{1/2}, L = (rQ^{\alpha}/w)^{1/2} \tag{5.11}$$

那么

$$C(w,r,Q) = 2(wr)^{1/2}Q^{\alpha/2} \tag{5.12}$$

更一般的情形见下面的习题 5.1。

在短期中没有选择的自由。因此，如果使用资本 K 生产产出 Q，那么必须使用劳动 $L=Q^{\alpha}/K$，成本函数变为：

$$C(w,r,Q,K) = wQ^{\alpha}/K + rK \tag{5.13}$$

如果有第三种投入，比如说原材料，其数量在短期中也可以变化，那么就还有一个短期的成本最小化问题要解。我把它留作一道习题。

长期的边际成本可以通过微分(5.12)式而得到：

$$C_Q(w,r,Q) = \alpha(wr)^{1/2}Q^{\alpha/2-1} \tag{5.14}$$

在短期中(5.13)式则给出

$$C_Q(w,r,Q,K) = \alpha w Q^{\alpha-1}/K \tag{5.15}$$

如果 K 的值恰好是(5.11)式所给出的长期中 K 的最优值,那么短期的边际成本(5.15)式和长期的边际成本(5.14)式正好一致,正如包络定理所要求的那样。

例题 5.2:消费者需求

本章所引入的最重要的新思想,就是把目标函数的最大值看作是关于最大化问题的参数的函数。这样的函数包含了很多经济学上有用的信息,而这些信息在许多实际应用中可以用来简化对最优行为的分析。这个例题讨论了以效用最大化为基础的消费者需求理论的情形。

考虑一个消费者,他在预算 $px=I$ 的约束下最大化效用函数 $U(x)$。其中 p 为价格行向量,x 为数量列向量,I 为货币收入。这个最优化问题的参数为 p 和 I,由此得到的最大效用为函数 $V(p,I)$。这个函数也被称为间接效用函数,以区别于以消费数量定义的直接效用函数。

V 的一些性质显而易见。例如,以相同的比例变动所有价格和收入不会改变预算约束,因而不会改变最优选择和效用水平。因此,V 对(p,I)是零次齐次的。我们在后面还有机会研究 V 的一些其他性质。不过此处的关注点在于包络定理的应用,或者更具体些,在于(5.4)式的应用。注意到拉格朗日函数是:

$$L(x,\lambda,p,I) = U(x) + \lambda(I - px)$$

因而

$$V_I(p,I) = L_I(x,\lambda,p,I) = \lambda \tag{5.16}$$

在最优解处取值。类似地,V 对价格的偏导数列向量为

$$V_p(p,I) = L_p(x,\lambda,p,I) = -\lambda x \tag{5.17}$$

同样在最优解处取值。当然,使效用最大化的数量 x 构成了(向量)需求函数$D(p,I)$。因而我们可以用(5.16)式去除(5.17)式,写成

$$D(p,I) = -V_p(p,I)/V_I(p,I) \tag{5.18}$$

这是一个有用而重要的结论。如果我们知道消费者的(直接)效用函

数，要求找出他的需求函数，我们必须求解整个约束条件下的最大化问题，而这是很复杂的，即使是在最简单的情形下。但是另一方面，如果我们知道了他的间接效用函数，仅仅通过微分就能得到他的需求函数。所以通过间接效用函数总结出有关消费者的信息更为简单。特别是在一般均衡模型中，消费者只是该情形的一个部分，计算和表示法简洁与否会导致很大的差异。在后面的章节中，我们将会看到这方面的一些例子。

接下来考虑一个对称的问题，即消费者如何以最少的支出达到一个既定的效用水平。由此得到的最小的支出是关于价格向量 p 和既定效用水平 u 的函数，将其称为消费者的支出函数，记作 $E(p,u)$。保持 u 不变，而使所有价格以相同比例变动就会使必要的支出同比例地变动；因而对于固定的 u，E 是 p 的一次齐次函数。E 的其他性质将在必要之处再作讨论。同样地，我们的关注点也在包络性质。

将最小化问题的拉格朗日函数写作

$$L(x,\mu,p,u) = px + \mu[u - U(x)]$$

同前面一样地推导，我们有

$$E_u(p,u) = \mu \tag{5.19}$$

这个拉格朗日乘子给出了为实现效用水平的边际增加所要求的支出的增加。因而它正好是前面那个货币的边际效用 λ 的倒数。接下来，

$$E_p(p,u) = x$$

对给定效用水平时使成本最小的商品选择被称为希克斯补偿需求函数 $C(p,u)$。其情形就像是，对于任意的价格变动，消费者通过变动其货币收入得到补偿，而这种变动足以使其仍然处于原来的无差异曲线上。这样我们就得到了：

$$C(p,u) = E_p(p,u) \tag{5.20}$$

这个表达式甚至比前面给出的非补偿需求函数 D 更为简洁。

最后，我们可以把间接效用函数和支出函数联系起来，从而也把非补偿需求函数和补偿需求函数联系起来。假定我们从某个效用水平 u 开始，找到 $I=E(p,u)$。然后我们就把这个 I 作为货币收入，并找出使效用最大化的选择。只要所有的价格都是正的，就像在绝大部分初级经济学应用中那

样，这反过来得到了我们开始时的那个效用水平，也就是 $u=V(p,I)$。使支出最小化的选择 x 也是使效用最大化的选择，即只要 u 和 I 具有上述关系，那么 $C(p,u)=D(p,I)$。取第 j 个分量等式，两边对 p_k 求导。保持 u 不变，但使 $I=E(p,u)$ 为 p_k 的函数。根据链式法则，

$$C_k^j(p,u)=D_k^j(p,I)+D_I^j(p,I)E_k(p,u)$$

而

$$E_k(p,u)=C^k(p,u)=D^k(p,I)$$

表示对商品 k 的需求。因而

$$C_k^j(p,u)=D_k^j(p,I)+D^k(p,I)D_I^j(p,I) \tag{5.21}$$

补偿需求函数的导数和非补偿需求函数的导数之间的这个关系被称为斯拉茨基—希克斯方程。有些读者可能知道它的另一种记法：

$$\left(\frac{\partial x_j}{\partial p_k}\right)_{u \text{ constant}}=\left(\frac{\partial x_j}{\partial p_k}\right)_{I \text{ constant}}+x_k\frac{\partial x_j}{\partial I}$$

习题

习题 5.1：柯布—道格拉斯成本函数

考虑一个生产函数：

$$y=A\prod_{j=1}^{n}X_j^{\alpha_j} \tag{5.22}$$

其中 y 为产出，x_j 为投入，A 和 α_j 为正的常数。令 $w=(w_j)$ 表示投入要素的价格向量，证明：生产给定产出水平 y 的最小成本为：

$$C(w,y)=\beta(y/A)^{1/\beta}\prod_{j=1}^{n}(w_j/\alpha_j)^{\alpha_j/\beta} \tag{5.23}$$

其中 $\beta=\sum_j\alpha_j$，如果 $\beta<1$，那么计算相应的最大利润函数 $\pi(p,w)$，其中 p 表示产出价格。如果 $\beta\geqslant 1$，会出现什么问题？

习题 5.2：CES 支出函数

假定直接效用函数为：

$$U(x,y)=[\alpha x^{\rho}+\beta y^{\rho}]^{1/\rho} \tag{5.24}$$

其中 x 和 y 为两种商品的数量，α、β 和 ρ 是给定的常数，α、β 为正数并且 $\rho<1$。证明支出函数具有如下形式：

$$E(p,q,u)=[ap^{r}+bq^{r}]^{1/r} \tag{5.25}$$

其中 p 和 q 为商品的价格，u 为效用水平，a、b 和 r 为可以用 α、β 和 ρ 表示的常数。

找出补偿需求函数，并证明使支出最小化的商品数量的比例为：

$$x/y=(a/b)(q/p)^{1-r}$$

(x/y) 对 (q/p) 的弹性：

$$\frac{\mathrm{dln}(x/y)}{\mathrm{dln}(q/p)}$$

被称为生产中的替代弹性。证明：在这个例子中，该弹性为常数并且等于 $(1-r)$。我们必须对 ρ 施加怎样的条件，以保证替代弹性为非负，即 $r<1$？

▶ 进一步阅读

关于包络定理的发现的一个有趣的论述见：Jacob Viner, "Cost curves and supply curves"，重印于 George J. Stigler and Kenneth E. Boulding (eds.), *Readings in Price Theory*, Homewood, IL: Irwin, 1952。

Samuelson 澄清了关于 Viner 的发现的混淆之处，他的原著至今仍值得一读：Paul A. Samuelson, *Foundations of Economic Analysis*, Cambridge, MA: Harvard University Press, 1947, pp. 34—6。

对于包络定理更为深入的讨论，可以参考：Varian, *Microeconomic Analysis*（1984，见本书第 3 章中的进一步阅读）中的数学附录；或者 David Kreps, *A Course in Microeconomic Theory*, Princeton, NJ: Princeton University Press, 1990，第 7 章。

对于间接效用函数和支出函数更为细致的讨论可以参考 Varian（同上，第 3 章）和 Kreps（同上，第 2 章），更多的关于企业成本和利润函数的讨论见 Varian（同上，第 1 章）和 Kreps（同上，第 7 章）。

▶6

凸集及其分离

6.1 分离性质

拉格朗日定理和库恩一塔克定理给出了约束条件下最大化问题的一阶必要条件,但是这些条件在一般情况下并不足以确定最优解。如果我们想在同样的约束条件下最小化同样的目标函数,也会得到相同的一阶条件。当我在第 2 章中讨论这一点时,我曾提到最大值和最小值可以通过检验目标函数和约束函数的曲率来加以区别。这一思想将在本章和第 7 章及第 8 章中展开。

让我们从如下的最优化问题开始:即在标量约束 $G(x)\leqslant c$ 下选择向量 x 以最大化 $F(x)$。令 $\bar{x}$ 表示最优选择,$\bar{v}$ 表示最大值。

图 2.1 给出了我们熟悉的相切方法;现在我要用一种新的并且有效的方法来解释最优化问题的解。那张图给出了 $G(x)=c$ 的等值线。而现在我们需要知道所有点 x 的集合,同时,这些点满足更为一般的不等式约束 $G(x)\leqslant c$,或 G 关于 c 的下等值集。在图 6.1 中,它就是等值线 $G(x)=c$ 以下的阴影区域 $\mathscr{A}$。类似地,图 2.1 只给出了等值线 $F(x)=\bar{v}$。而我们现在需要的,是满足 $F(x)\geqslant\bar{v}$ 的所有点的集合,或 F 关于 $\bar{v}$ 的上等值集。在图 6.1 中它就是阴影区域 $\mathscr{B}$。该图假定 F 和 G 都是增函数。这在很多经济学的应用问题中都是成立的,例如当 F 为福利函数,G 为产出向量 x 所要求的投入资源的函数。不过,对于其他情形,我们也可以画出类似的图形。

这两条曲线在 $\bar{x}$ 处相遇,图形给出了它们在这一点处的共切线。我们选择合适的曲率以确保存在一个最大值。在第 2 章中,我们说这些曲率反

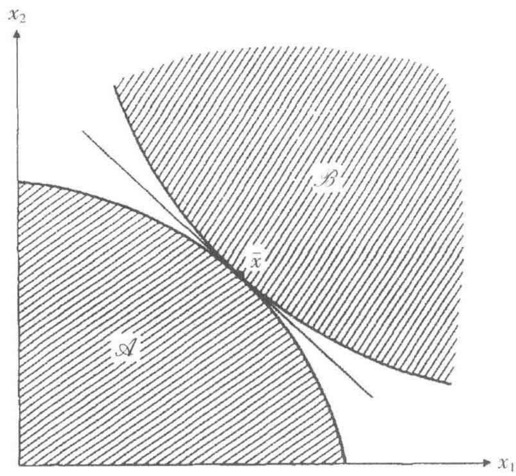

图 6.1　由共切线形成的分离

映了递减的边际替代率和边际转换率，现在我将提供一个不同的解释。

注意到集合 $\mathscr{B}$ 和 $\mathscr{A}$ 分别位于共切线的一边，且仅有一个公共点 $\bar{x}$ 位于共切线上。换言之，共切线把平面 x 分离为两半，每一半包含了其中一个集合。在更高的维度中，共切线将是一个超平面，它也会类似地分离出两个等值集。

分离是一个极其重要的性质，它允许我们区分出最大值和最小值，以获得最大化问题的充分条件。现在，我们必须将关于函数 F 和 G 的隐含条件弄清楚，以保证它们有适当的曲率。

6.2　凸集和凸函数

图 6.1 中的每一个等值集都是向外鼓出的，这也是为什么每一个集合在 $\bar{x}$ 处都会弯离共切线，并且不会弯回来以再次相遇。这种往外鼓的性质被称为凸性。检验凸性的一个几何方法是，对于集合中给定的任意两点，它们的整个连线段应在这个集合之中。从代数上看，对于 n 维空间中的点集 S，如果给定 S 中的任意两点 $x^a=(x_1^a,x_2^a,\cdots x_n^a)$ 和 $x^b=(x_1^b,x_2^b,\cdots x_n^b)$，以及闭区间[0,1]中的任意实数 θ，点 $(\theta x^a+(1-\theta)x^b)$ 或其写成分量的形式 $(\theta x_1^a+(1-\theta)x_1^b,\theta x_2^a+(1-\theta)x_2^b,\cdots,\theta x_n^a+(1-\theta)x_n^b,)$ 也在 S 中，那么 S 被称为是凸的。

应用到 G 的下等值集，其凸性意味着如果 x^a 和 x^b 满足约束条件，那么 $\theta x^a+(1-\theta)x^b$ 也满足约束条件。在经济学的应用问题中，约束往往反映了资源的有限性。所以 $G(x)$ 可能是生产产出向量 x 所需要的劳动数量，而 c 为现有劳动的数量。在这样具体的背景中，集合 $G(x)\leqslant c$ 的凸性就意味着两个生产计划的加权平均不会比其中任何一个生产计划需要更多的劳动数量。这排除了明显的规模经济或专业化情形。类似地，应用到 F 的上等值集，其凸性意味着两个消费计划的加权平均至少和其中的任何一个消费计划一样好；这也是偏好多样性的假定。

从代数上看，上述条件是说：

$$G(x^a)\leqslant c \text{ 和 } G(x^b)\leqslant c\text{，意味着 } G(\theta x^a+(1-\theta)x^b)\leqslant c$$

当其中的一个极端值等于 c 时就出现了对上述表达最严格的检验，因而我们可以把这个条件又表述为：

$$G(\theta x^a+(1-\theta)x^b)\leqslant \max(G(x^a),G(x^b)) \tag{6.1}$$

对于所有的 x^a、x^b 和所有的 $\theta\in[0,1]$ 成立。这种形式的一个额外的优点是，它不必涉及特定的数字 c。由于在实际应用中我们是对一个一般的 c 求解最大化问题的，所以我们必须导出对所有 c 都成立的条件，而像(6.1)式这样的一般形式是实现这一点的最佳途径。

满足(6.1)式的函数 G 被称为拟凸函数。同样，关于 F 的类似的条件为

$$F(\theta x^a+(1-\theta)x^b)\geqslant \min(F(x^a),F(x^b)) \tag{6.2}$$

对于所有的 x^a、x^b 和所有的 $\theta\in[0,1]$ 成立。这样一个函数被称为拟凹函数。

上述定义中的"拟"是为了把它们与经常出现在最优化中的更严格的性质区分开来。在通常的经济学解释中，拟凸性对应于边际转化率沿着约束曲线递减，同时约束函数的凸性对应于规模报酬递减。类似地，目标函数的拟凹性表示边际替代率沿着无差异曲线递减，同时凹性意味着边际效用递减。

正式地，我们定义 G 是凸的，如果对于所有的 x^a、x^b 和所有的 $\theta\in[0,1]$，有

$$G(\theta x^a + (1-\theta)x^b) \leqslant \theta G(x^a) + (1-\theta)G(x^b) \tag{6.3}$$

(6.3)式的右边显然不会大于(6.1)式的右边：

$$\begin{aligned}
&\theta G(x^a) + (1-\theta)G(x^b) \\
&\leqslant \theta \max(G(x^a), G(x^b)) + (1-\theta)\max(G(x^a), G(x^b)) \\
&\leqslant \max(G(x^a), G(x^b))
\end{aligned}$$

因而如果(6.3)式成立，那么(6.1)式也一定成立。换言之，凸函数也是拟凸的，凸性是凸性和拟凸性两者中更强的性质。

类似地，我们将 F 定义为凹的，如果对于所有的 x^a、x^b 和所有的 $\theta \in [0, 1]$，有

$$F(\theta x^a + (1-\theta)x^b) \geqslant \theta F(x^a) + (1-\theta)F(x^b) \tag{6.4}$$

换言之，凹函数的图形位于连接图形上任意两点的弦上或者上方。图 6.2 给出了当 x 为标量时的情形的一个说明。同样，也很容易验证，(6.4)式能导出(6.2)式，即凹函数也是拟凹的。

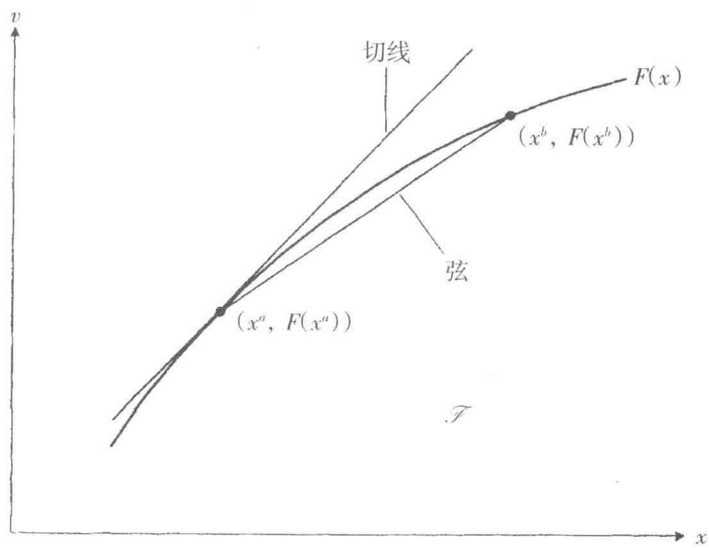

图 6.2　凹函数

注意到(6.3)式和(6.4)式中的不等号是弱的，因而凸函数和凹函数也允许它们的图形中包括直线段，这时弦与图形就重合了。特别要说明的是，线性函数既是凸的又是凹的。我们在后面将有机会来巩固凸性和凹性的

概念。

有时凹函数的另一个定义也是有用的。考虑由点(x, v)组成的$(n+1)$维的空间，其中x为n维的向量，v为标量。由此，我们定义集合：

$$\mathscr{F}=\{(x, v) \mid v \leqslant F(x)\}$$

那么当且仅当$\mathscr{F}$为凸集时F为凹函数，我们可以用(6.4)式来检验这个定义。换言之，凹函数在它的图形下面封闭成一个凸集。在图6.2中我们很容易看到这一点。凸函数的类似的性质也可以很容易地给出，我把它留给读者。最后，如果函数是可微的，那么凹性、拟凹性等性质可以用一阶导数和二阶导数来表示，我将在第7章中继续分析。

我还必须定义另外两个概念，才可以得出我想要得出的基本的数学结论。如果点x^0被集合S中的点以某个距离所包围，那么点$x^0 \in S$被称为是内点，即存在一个$r>0$，使得所有和x^0的距离在r之内的点x都在集合S中。在平面中，这样的点会形成一个圆心为x^0半径为r的圆盘。那么，如果一个点既不是S的内点也不是S补集的内点，我们就称该点为集合S的边界点。这样，如果对于任意的$r>0$，我们可以在x^0的以r为半径的范围内既找到S中的点也找到S外部的点，我们就称x^0为S的边界点。

例如在图6.1中，$\bar{x}$既是$\mathscr{B}$的边界点，同样也是$\mathscr{A}$的边界点。只要F是连续的，那么任何满足$F(x)>\bar{v}$的点x都是$\mathscr{B}$的内点。类似地，只要G是连续的，那么任何满足$G(x)<c$的点x都是$\mathscr{A}$的内点。

这两个集合只有边界点$\bar{x}$这一个公共点，并由共切线将它们分离。令这条共切线的方程为

$$px \equiv p_1 x_1+p_2 x_2=b$$

其中p是系数行向量，所以px为内积，或者说是p和x相应分量的积之和。当然，$p \neq 0$，也就是说p_1和p_2中至少有一个必须为非零。由于$\bar{x}$位于分离线上，我们有：

$$p\bar{x} \equiv p_1 \bar{x}_1+p_2 \bar{x}_2=b$$

对于所有位于直线一边的点x，px大于b，而对于所有位于直线另一边的点x，px小于b。

如果两个集合没有任何公共点，那么它们之间会有一个明显的间隙，我

们当然也可以画出一条把它们分离开的直线，尽管它并不需要同其中的任一集合相切。但是如果两个集合有公共的内点，那么任何一条完全在 $\mathscr{A}$ 以上的直线必定会切入集合 $\mathscr{B}$，而任何一条完全在 $\mathscr{B}$ 以下的直线也必然会切入集合 $\mathscr{A}$；因而两个集合分离将是不可能的。

集合的凸性也很重要。图 6.3 给出了两种情形，在每一种情形中，都有一个集合是非凸的。集合的共切线会切入那个非凸的集合中去，从而使分离失败。

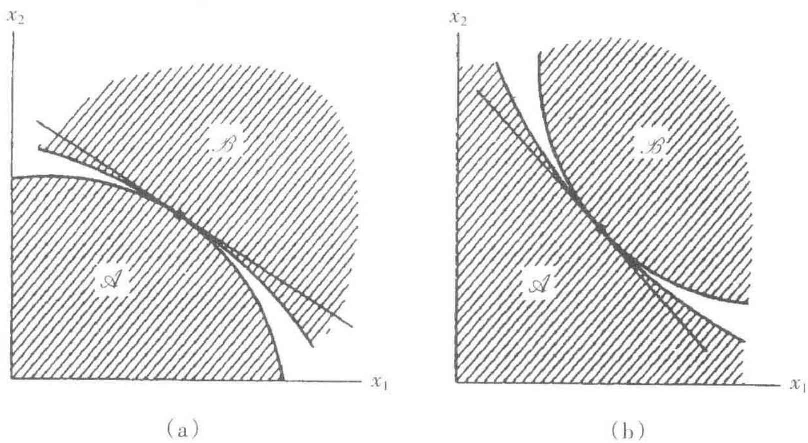

图 6.3　分散化下的局部失灵

在下面的定理中我们将所有这些思想形式化。我希望绝大部分读者觉得图形的说明便足以令人信服，所以我不再给出正式的证明。我只以最简单的形式陈述那些符合我们目的的结论，尽管更为一般的结论也是存在的。

分离定理：如果 $\mathscr{B}$ 和 $\mathscr{A}$ 为两个没有公共内点的凸集，且至少有一个集合有一个非空内点，那么我们总可以找到一个非零向量 p 和一个数 b，使得超平面 $px=b$ 分离这两个集合，或

$$px\begin{cases}\leqslant b, \text{对所有 } x\in\mathscr{A}\\ \geqslant b, \text{对所有 } x\in\mathscr{B}\end{cases}\tag{6.5}$$

"至少有一个集合具有一个非空内点"的要求，排除了集合的维度小于整个空间的维度的一些退化情形。只是出于完整性我才提到这一点，不过在经济学的应用问题中，目标函数的等值集是满维度的，所以不会出现上面论及的困难。

6.3 分离角度的最优化定理

令 F 和 G 都是连续函数，且 F 拟凹 G 拟凸。考虑在约束 $G(x)\leqslant c$ 下 $F(x)$ 的最大化问题。只要约束集是有界的，这个问题就有解。在本书所讨论的绝大部分经济学的应用问题中，除了第 10 章无限时间期界中的一种情形，最优解的存在性不是问题。令 $\bar{x}$ 代表最优解，图 6.1 中假定的所有条件都满足，从而分离定理适用。令 $px=b$ 表示分离共切线方程。如果我们在方程两边都乘以 -1，方程并不会受到影响，但这会改变(6.5)式中不等号的方向。所以为了保证不等式为适当的形式，使得集合 $\mathscr{B}$ 和 $\mathscr{A}$ 以上面论述的形式出现，我们应选择 p_1 和 p_2 为正。

由于点 $\bar{x}$ 位于分离切线上，我们有 $p\bar{x}=b$。因而(6.5)式告诉我们，对于所有 $\mathscr{A}$ 中的点，或所有满足 $G(x)\leqslant c$ 的点，$\bar{x}$ 给出了 px 的最大值。类似地，对于所有在 $\mathscr{B}$ 中的点，或对于所有满足 $F(x)\geqslant\bar{v}$ 的点，$\bar{x}$ 给出了 px 的最小值。这就是这一部分论述的主要结论。

分离角度的最优化定理：*给定拟凹函数 F 和拟凸函数 G，点 $\bar{x}$ 在满足约束 $G(x)\leqslant c$ 下使 $F(x)$ 最大，当且仅当存在一个非零向量 p，使得：*

(i) $\bar{x}$ 在满足 $G(x)\leqslant c$ 下最大化 px，且

(ii) $\bar{x}$ 在满足 $F(x)\geqslant\bar{v}$ 下最小化 px。

将上述结论一般化成有几个约束的情形是非常简单的。如果 G^i 是拟凸的，那么满足 $G^i(x)\leqslant c_i$ 的集合 $\mathscr{B}_i$ 为凸集。如果这对于所有的 i 都成立，那么满足所有约束的点的集合 $\mathscr{B}$，即与凸集 $\mathscr{A}_i$ 的交集，本身也是一个凸集；利用凸集的定义我们可以很容易地证明这一点。论证过程和上面一样。

注意在定理的论述中的“当且仅当”条件，这些条件既是最优化的必要条件，也是最优化的充分条件：如果 $\bar{x}$ 是最优的，那么这些条件就会成立，反之亦然。从两方面来看，这个定理超越了我们前面讲过的拉格朗日定理或者库恩一塔克定理。这是我们第一次遇到了一个充分条件，因而这个定理就其本身而言也是有意义的。不过这些条件在实际应用中并不容易被验证。毕竟，我们用两个最优化的问题替代了一个最优化问题，而且又没有给出有效的标准，来判断什么时候其中一个问题已被求解出来了。在接下去的两章中，我们将会看到一些在此方面更为有用的充分条件。

把一个最优化问题，分解为两个目标函数都为线性函数的最优化问题，其真正的好处在于它经济学上的解释。它给出了利用价格对资源进行**分散化**的最优配置的可能性。为了给出一个最简单的解释，假定 x 为生产—消费向量，约束反映了资源的有限性，目标函数为效用函数。现在把 p 解释为产出的价格行向量，那么上述定理的(i)部分是说，寻求产出价值最大化的企业家会生产出最优产量 $\bar{x}$；(ii)部分是说，一个试图以最小支出达到既定效用的消费者也恰好需要 $\bar{x}$。如果我们假定不考虑由于存在免费物品而出现的技术上的复杂的细节问题，这等价于在预算约束下最大化效用。这样，原来那个社会最优化的问题就可以被分散化的机制实现。让一个企业家选择生产计划，由此得到的产出的价值以通常的收入循环的形式归消费者所有。然后再让消费者求解他的效用最大化问题。这种决策的分离有两个好处。一是信息上的好处：生产者无需知道消费者的品味，消费者也无需知道生产者的技术。无论对谁而言，与对方有关的信息都已经充分概括在价格向量 p 中。另一个好处与激励有关：上述的过程依赖于每一方的自利行为，以保证最优解的有效实施。

注意到在上述问题中，如果我们用一个相同的正数同时乘以价格向量 p 和实数 b，经济的实质不会发生任何变化。换句话说，只有相对价格如(p_1/p_2)才会影响经济决策。当生产者让他们在生产中的边际转化率等于相对价格，而消费者让他们在消费中的边际替代率等于相对价格时，我们就解出了上述定理中的分离最优化问题。

当然，现实中的分散化机制问题要复杂得多。当有多个生产者和消费者时，就必须讨论外部效应和收入分配的问题。同时还有如何确定正确的价格向量的问题。通常，人们并没有积极性去披露他们的私人信息，而这些私人信息对计算正确的价格、税收等又是需要的。他们这样的行为会给社会最优问题添加额外的约束。这些问题在微观经济学的教科书中有长篇的讨论，并且仍然是研究中非常活跃的论题。不过上文中这个简单的描述仍然为我们提供了一个有用的起点。

如果我们不假定 $\mathscr{B}$ 和 $\mathscr{A}$ 都是凸集，那么充分的分散化机制是不可能实现的。图 6.3 说明了这一点。在情形(a)中，$\mathscr{B}$ 不是一个凸集，因此，$\bar{x}$ 就不能在这个集合上最小化 px。此时，与多样化的消费束相比，消费者更偏好极端点。在情形(b)中，$\mathscr{A}$ 不是一个凸集，因此，$\bar{x}$ 就不能在这个集合上最大

化 px。此时，生产技术具有规模经济或专业化经济的效应。不过在这两种情况下，$\bar{x}$ 确实在约束 $G(x)\leqslant c$ 下均使 $F(x)$ 取得了最大值。因此，假定 F 为拟凹函数 G 为拟凸函数而得到的分离角度的最优化定理，并没有涵盖所有经济学上有趣的情形。真正起作用的是 F 和 G 等值线的相对曲率。在第 8 章中，我将用微积分的方法讨论符合这一思想的条件。

6.4 惟一性

在图 6.1 中，两个集合 $\mathscr{B}$ 和 $\mathscr{A}$ 的边界都画成了光滑的曲线，但根据凸集的定义这并不是必然的。在一般情况下，一凸集在其边界上可以有线段，而这仍然可以让连接集合中任意两点的线段还是包括在集合中。此外，集合的边界上还可以有折拐，只要这个转折点是朝外的。

这些可能性对分离和最优化有一定的意义。图 6.4 给出了一些此类情形。在(a)中，两个折拐正好在最优解 $\bar{x}$ 处相遇。现在我们可以找出很多条经过 $\bar{x}$ 的直线来分离这两个集合，即分散机制中的价格向量 p 不是惟一的。没有一条分离直线可以被称为通常意义上的共切线，但这对于此问题的经济学解释并不是必要的。分散化机制只依赖于分离的性质，即两个集合分别位于直线 $px=b$ 的两边。因此分离是比共切线更为一般的概念。这种一般化和分散化的性质，甚至在函数 F 和 G 不可导时仍然成立。在这样的情况下，那种通常平滑的边际替代率和边际转化率都没有定义，而且也不能像平时那样等于价格的比率。实际的情况是，在无差异曲线上，折拐处的左边和右边具有不同的边际替代率，而价格比率就位于这两个值之间。转换曲线也是类似的。

在情形(b)中，两个集合具有一段公共的平坦部分。现在这一段上的任意一点都可以作为最优解 $\bar{x}$。所有这些点都会得到目标函数 $F(x)$ 的相同的值，所以不惟一性并不是关注的原因。不过，分散化机制确实存在一个问题。给定价格 p，所有位于 $\mathscr{A}$ 平坦部分的点都会给生产者带来相同的产出的值，所有位于 $\mathscr{B}$ 平坦部分的点也都会给消费者带来相等的效用。因此，在这个程度上他们的选择将是任意的，当然也没有理由保证这些独立的选择会恰好一致。我们只能作出相对较弱的论断，如果两者正好作出了一致的选择，那么没有哪一方有任何正的激励去偏离这一选择。这是在任何经济

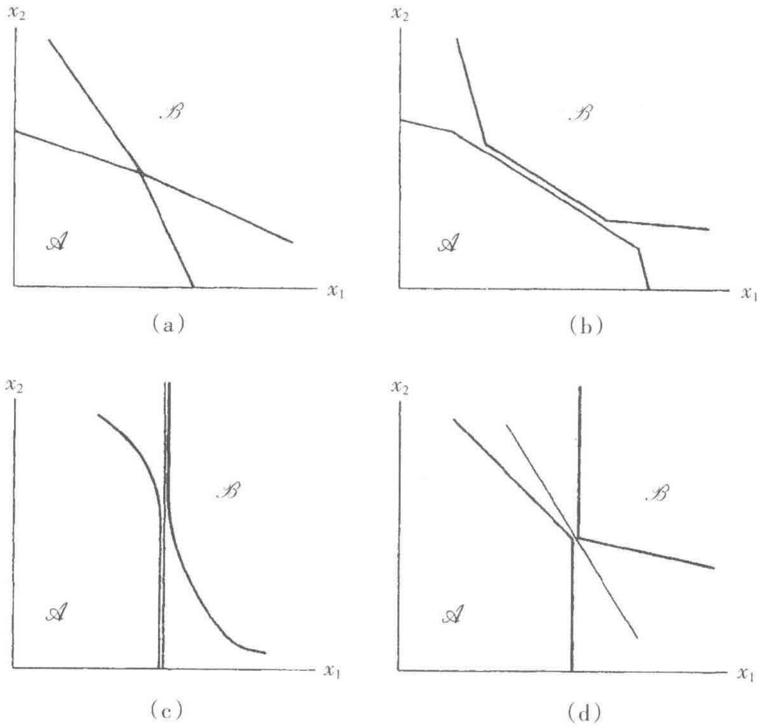

图 6.4　在折拐点和平坦部分的最优解

均衡理论的谨慎论述中一个典型的局限性。

如果两个集合的边界有垂直的公共部分，如在情形(c)中所画的那样，那么我们就有一条垂直的分离线。用方程表示，即 $p_2=0$，从分散化机制的角度来解释，即商品 2 为免费物品。类似地，水平的分离线则意味着 $p_1=0$。在情形(d)中也有一个折拐，有垂直的分离线，也有不垂直的分离线。因此，如果没有更强的假设，保证价格为正是不可能的。事实上，如果在最优处两个边界都是向上倾斜的，那么共切线会有一个正的斜率，故其中有一个价格将为负。通常可以通过如下假定来避免这个问题：要么可以自由处置两种商品，此时 $\mathscr{A}$ 的边界不可能向上倾斜；要么两种商品都是想要的，此时 $\mathscr{B}$ 的边界不能向上倾斜。上面的图示都隐含了这两个假设。

总结一下，出现折拐的问题并不严重；事实上它们允许我们将相切的概念一般化，并保留分散化的性质。出现平坦部分的问题更为严重，因为最优选择可能不惟一从而难以保证生产者和消费者们作出相互一致的选择。因而，知道函数 F 和 G 的哪些性质可以解决此困难是有帮助的。强调拟凹性

和拟凸性的概念也是需要的。

重温凸集的定义：如果给定集合 S 中的两点 x^a 和 x^b，以及满足 $0\leqslant\theta\leqslant 1$ 的任意 θ，中间点 $\theta x^a+(1-\theta)x^b$ 也在 S 中，那么 S 为凸的。连接 x^a 到 x^b 的整个线段可能位于 S 的边界上，那是因为边界上可能有平坦的部分。我们可以通过要求这条线段上除了端点以外的所有点都是 S 的内点，来强化凸集的定义：只要 x^a 和 x^b 是不同的两个点，并且 θ 不等于 0 或者 1，那么点 $\theta x^a+(1-\theta)x^b$ 就应该是 S 的内点。如果情况确实如此，那么集合 S 就被称为严格凸的。如果函数 F 所有的上等值集为严格的凸集，那么它被称为是严格拟凹的。

现在考虑在约束 $G(x)\leqslant c$ 下最大化 $F(x)$ 的问题，其中 F 为严格拟凹的，G 为严格拟凸的。假定 $\bar{x}$ 满足分离角度的最优化定理的条件，那么它是此最优化问题的惟一解。

为了理解这一点，假定 $\hat{x}$ 为另一个解。两个解对分散化机制下的最优化问题中的第二部分（消费者最优化问题）都必须是可行的，因此有：

$$p\bar{x}=p\hat{x}=b, F(\bar{x})=F(\hat{x})=\bar{v}$$

取它们的中点 $\tilde{x}=\frac{1}{2}(\bar{x}+\hat{x})$。由于 F 为严格拟凹的，这个中点就一定是上等值集 $\mathscr{B}$ 的一个内点，即 $F(\tilde{x})>\bar{v}$。我们可以找到一个仍然在 $\mathscr{B}$ 中并且与它相邻的点，使它有一个比 $\bar{x}$ 更小的 px 值。更正式地，取正数 ε，并把 x' 定义为：

$$x'_j=\tilde{x}_j-\varepsilon p_j,\ j=1,2,\cdots,n$$

因为 $\tilde{x}$ 为 $\mathscr{B}$ 的内点，故只要 ε 足够小，x' 就在 $\mathscr{B}$ 的内部，而且

$$\begin{aligned}px' &= p\tilde{x}-\varepsilon p^2\\ &=\frac{1}{2}(p\bar{x}+p\hat{x})-\varepsilon p^2\\ &=\frac{1}{2}(b+b)-\varepsilon p^2\\ &<b\end{aligned}$$

其中 p^2 代表价格向量 p 与其自身的内积，也就是 $\sum_j p_j^2>0$。这样我们就找到了一个点 $x\in\mathscr{B}$ 满足 $px<b$，而这与分离的性质相矛盾。这一矛盾来源于

最初的 $\bar{x}$ 不是惟一解的假定。因而那个假定必定是错误的。因此 F 的严格拟凹性就意味着最优解的惟一性。

G 的严格拟凸性将发挥同样的作用。当存在几个约束条件时，我们必须假定每一个分量约束 G^i 都是严格拟凸的，以保证约束集 $\mathscr{A}$ 的严格凸性。

例题

例题 6.1：分离的一个图示

假定有两个非负选择变量，用 x 和 y 来表示，函数 F 和 G 由下式给出：

$$F(x,y) = xy,\ G(x,y) = x^2 + y^2$$

约束条件为 $G(x,y) \leqslant 25$。

图 6.5 给出了一个图示。可行集 $\mathscr{A}$ 由四分之一圆周及其以下所有的点组成。函数 F 关于 v 的不同值的上等值集的边界在图中表示为一族直角双曲线。最优解位于 $(\bar{x},\bar{y})=(5/\sqrt{2},5/\sqrt{2})$，$F(x,y)$ 的最大值为 $\bar{v}=12\frac{1}{2}$。

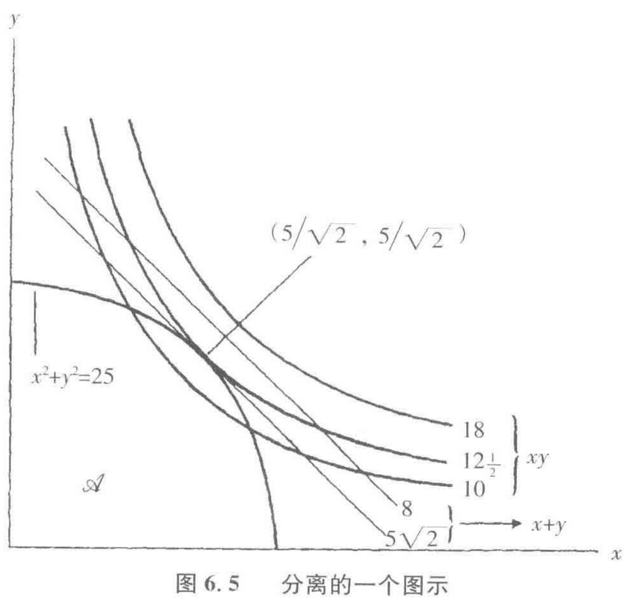

图 6.5　分离的一个图示

对应于 $\bar{v}$ 的上等值集 $\mathscr{B}$ 与可行集在最优解处相碰，它们被共切线 $x+y=5\sqrt{2}$ 相分离。当取更大的值 18 时，F 的上等值集和集合 $\mathscr{A}$ 就没有公

共点了，我们可以在它们中间清晰的空隙中画一条分离直线 $x+y=8$。对于小于 $\bar{v}$ 的值，比如说 10，F 的上等值集和可行集就有公共的内点（它们交集的透镜形状的区域），因而两者不能被分离。

例题 6.2：间接效用函数和支出函数

例题 5.2 定义了一个效用最大化的消费者的间接效用函数和支出函数。在此，我将检验它们的凸性性质。

先从如下支出函数开始：

$$E(p,u)=\min_{x}\{px \mid U(x)\geqslant u\} \tag{6.6}$$

对于每一个固定的 u，它是关于 p 的凹函数。为了明白这一点，取任意两个价格向量 p^a 和 p^b，以及[0,1]中任意的一个 θ。令 $p^c=\theta p^a+(1-\theta)p^b$，那么我们需要证明：

$$E(p^c,u)\geqslant \theta E(p^a,u)+(1-\theta)E(p^b,u) \tag{6.7}$$

令 x^c 表示在价格 p^c 下使支出最小的消费束。当然 x^c 必须满足约束，所以$U(x^c)\geqslant u$。约束条件并不涉及价格向量，因此当价格为 p^a 或 p^b 时 x^c 也是可行的。在每一种情况下，它不可能比最小值有更小的支出：

$$p^a x^c\geqslant E(p^a,u),\ p^b x^c\geqslant E(p^b,u) \tag{6.8}$$

用 θ 乘以上面的第一式，用$(1-\theta)$乘以第二式。由于这两个乘数都是正的，所以相乘后不会改变不等号的方向。将两个式子相加，有：

$$p^c x^c=(\theta p^a+(1-\theta)p^b)x^c\geqslant \theta E(p^a,u)+(1-\theta)E(p^b,u)$$

上式左边正好为 $E(p^c,u)$。上式就证明了(6.7)式。

其中的经济学直觉为，当价格向量变动时，一个人可以保持他的消费数量向量不变，那么支出函数会随价格线性地变动。同时，沿着无差异曲线各个消费束之间存在着替代关系，从这个意义上说，消费选择可以随着价格的变动而进行相应的调整。这将使支出的变动慢于线性变动，即最小化了的支出将是价格的凹函数。另一种理解方法就是考虑最糟糕的情形，即消费中**不存在**替代关系，无差异曲线是 L 形的。且 x^c 是实现效用水平 u 的最优解，与价格水平无关。故(6.8)式中的两个不等式以等式的形式成立，于是(6.7)式也成了等式：支出是价格的线性函数。

接下来考虑间接效用函数：

$$V(p,I)=\max_x\{U(x)\mid px\leqslant I\} \tag{6.9}$$

它关于(p,I)是拟凸的。证明方法也和上文类似。令(p^a,I^a)和(p^b,I^b)表示两个价格—收入向量，θ为$[0,1]$中的任意数。令

$$(p^c,I^c)=\theta(p^a,I^a)+(1-\theta)(p^b,I^b)$$

假定x^c是价格—收入为(p^c,I^c)时效用最大化问题的解。它满足预算约束，所以$p^cx^c\leqslant I^c$。

现在我断定，至少对于(p^a,I^a)和(p^b,I^b)这两个价格—收入向量中的一个，x^c是可行的。如果它不可行，我们有：

$$p^ax^c>I^a,p^bx^c>I^b$$

用θ乘以上面的第一式，用$(1-\theta)$乘以第二式，然后将这两项相加，我们就会得到$p^cx^c\geqslant I^c$，这和x^c对(p^c,I^c)是可行的相矛盾。

只要x^c是可行的，无论哪一种情形，它的效用$U(x^c)$不可能超过那种情形下所能取得的最大效用。因而，

$$U(x^c)\leqslant V(p^a,I^a)\text{ 和 }U(x^c)\leqslant V(p^b,I^b)$$

至少有一个必定成立。于是有

$$V(p^c,I^c)\leqslant\max(V(p^a,I^a),V(p^b,I^b))$$

这正是V为拟凸函数的表述。

换句话说，间接效用函数的下等值集是凸集。这导致了一个非常不幸的结果。当政府选择最优的间接税时，它实际上是在选择某个价格以最大化间接效用函数。我们的结果表明，对于最大化问题而言这个目标函数有错误的曲率。因而最优税收问题的充分条件往往是难以验证的。

习题

习题 6.1：引起负效用的商品

以下两种情形下，图 6.1 会有怎样的变化？

(a)选择变量之一是劳动,它给消费者带来负效用,并且是生产的一种投入要素;

(b)商品之一是污染,它给消费者带来负效用,同时又是作为另一个选择变量的一种经济上有用的商品的副产品。

解释这两种情形下相应的价格的含义。

习题 6.2:企业利润函数的凸性

一个企业在生产可能性约束 $G(x,y)\leqslant 0$ 下选择投入向量 x 和产出向量 y,以最大化利润($qy-px$),其中 q 代表产出的价格行向量,p 为投入的价格行向量。令 $\Pi(q,p)$ 表示关于价格的最大化利润函数。证明:Π 为 (q,p)的凸函数。

习题 6.3:角点解

考虑一个有劳动禀赋 L 的经济。它可以生产两种商品 x_1 和 x_2。一单位商品 j 需要固定数量的 a_j 单位的劳动,所以生产可能性约束为:

$$a_1x_1+a_2x_2\leqslant L$$

这两种商品的世界价格分别是 p_1 和 p_2,且与这个国家选择的产出水平无关。目标为最大化国民产值($p_1x_1+p_2x_2$)。

画出图形,并通过分离两个凸集来求解这个问题。你需要根据(p_1/p_2)和(a_1/a_2)的大小问题,分别考虑两种情形。

当你在图形上找到答案之后,验证拉格朗日条件。找出并解释拉格朗日乘子。证明:用价格函数(收入函数或者 GNP 函数)来表示的最大化的国民产值为

$$R(p_1,p_2)=\max(p_1/a_1,p_2/a_2)$$

▶ 进一步阅读

关于凸集、凸函数、拟凸函数等等的定义,见 Varian, *Microeconomic Analysis*(1984,见本书第 3 章中的进一步阅读)第 311 页到 315 页。在此书

的第3章中，他更加详细地考虑了支出函数和间接效用函数，并证明了在给定这些最大值函数中的一个情形中，在什么条件下和怎样做，回到消费者的偏好才是可能的，这样就实现了这两种使消费者的选择模型化的方法之间的对偶性。另一个使用对偶的消费者理论的精彩讨论见：Angus Deaton and John Muellbauer，*Economics and Consumer Behavior*，Cambridge，UK：Cambridge University Press，1980。

对于最优税收的一个先驱性的分析，见：Peter Diamond and James Mirrlees，"Optimal Taxation and Public Production"，*American Economic Review*，61，March and June 1971，pp. 8—27，261—278。

▶7

凹规划

7.1 凹函数及其导数

在上一章我定义了凸集、拟凹函数和凹函数，并且在两个凸集的可分离性质的基础上导出了约束条件下最优化的一个几何解法。该方法有概念上的优点，即给出了社会的经济最优化问题的一个分散化的实施机制。但对于求解实际问题它的价值有限。在本章中我将凸性的思想和更传统的微积分方法结合起来，这样做得到的结果就是：拉格朗日或者库恩一塔克条件，连同目标函数和约束函数的凸性，是实现最优化的充分条件。

第一步是用函数的导数形式来表述它们的凸性等性质。在第 6 章中，凹函数是由性质"函数图形上连接任意两点的弦完全在该图形以下(或最多重合)"来定义的，见图 6.2。如果我们让点 x^b 越来越靠近 x^a，那么弦就逼近了 $F(x)$在 x^a 处的切线。于是，凹性的要求表明函数的图形应落在这条切线上，或者位于切线的下方。凹性所允许的最极端的情形，就是当 $F(x)$ 有直线段时，切线可以和这一线段重合。

更正式地，对于 $\theta\in[0,1]$，我们有：

$$F(x^a+\theta(x^b-x^a))=F(\theta x^b+(1-\theta)x^a)\geqslant\theta F(x^b)+(1-\theta)F(x^a)$$

于是，

$$\frac{F(x^a+\theta(x^b-x^a))-F(x^a)}{\theta}\geqslant F(x^b)-F(x^a)$$

现在让 θ 趋近于零。如果 F 是可微的，则链式法则表明，上式左边会逐渐趋

近于 $F_x(x^a)(x^b-x^a)$，其中 F_x 是偏导数行向量，乘积为内积。于是有

$$F_x(x^a)(x^b-x^a) \geqslant F(x^b)-F(x^a) \tag{7.1}$$

当 x 是一维的时候，$F_x(x^a)$ 就是 $F(x)$ 在 x^a 处切线的斜率，因此上式左边就是 x^a 向 x^b 移动时沿此切线的垂直距离。因而(7.1)式表明，凹函数值的变动用它的切线来估计是过大的，也就是说，切线位于曲线之上。更为一般的，上式左边就是以 x^a 作为初始点，$F(x)$ 值变动的泰勒近似的线性项，而这个近似又过大地估计了 $F(x)$ 的变动。

类似地，如果 G 是一个可微的凸函数，则有，

$$G_x(x^a)(x^b-x^a) \leqslant G(x^b)-G(x^a) \tag{7.2}$$

有一类特别重要的最优化问题有凹的目标函数和凸的约束函数，我们经常用术语凹规划用来描述这类一般的问题，而这正是下一部分的主题。

7.2 凹规划

考虑在向量约束 $G(x)\leqslant c$ 下最大化 $F(x)$ 的问题，其中 F 是可微和凹的，并且约束函数的每一个分量 G^i 是可微和凸的。这被称为凹规划的一般问题。为了使问题具体化并在经济学上有意义，我将使用生产问题中的术语，其中 x 为产出向量，$F(x)$ 为出售产品获得的收入，c 为固定的投入要素的供给向量，$G(x)$ 为生产 x 所需的投入向量。不过，数学本身是独立于这个经济解释的。

我们可以通过先考虑 c 为一般值的最大化问题来找出 c 为特定值的最优化条件。于是 x 的最优选择 $\bar{x}$ 与最大值 $\bar{v}=F(\bar{x})$ 都成了 c 的函数。令 $X(c)$ 表示最优选择函数，$V(c)$ 为最大值函数。

第一个结论是 V 为非减函数，这是因为，如果 x 在给定的 c 下是可行的，那么当 c 的任何分量增加后 x 仍然可行，因此最大值不可能下降。

接下来的结论是 V 为凹的。令 c 和 c' 为任意两个投入要素的供给向量，有

$$\bar{x}=X(c),\bar{x}'=X(c'),\bar{v}=V(c),\bar{v}'=V(c')$$

由于最优选择一定是可行的，那么我们有 $G(\bar{x})\leqslant c$ 和 $G(\bar{x}')\leqslant c'$。

现在令 θ 是[0,1]内的任意数。要使 V 为凹函数，必须要求，当投入要

素供给向量为 $\theta c+(1-\theta)c'$ 时，获得大于等于 $\theta V(c)+(1-\theta)V(c')$ 的收入是可能的。对这个产出向量的一个自然候选就是 $\theta\bar{x}+(1-\theta)\bar{x}'$。第一个要验证的是看它是否可行。对每一个 i，G^i 为凸的意味着

$$G^i(\theta\bar{x}+(1-\theta)\bar{x}')\leqslant\theta G^i(\bar{x})+(1-\theta)G^i(\bar{x}')\leqslant\theta c_i+(1-\theta)c_i'$$

可行性得到了证明。接下来要找出由此产生的收入。利用 F 的凹性，我们有

$$F(\theta\bar{x}+(1-\theta)\bar{x}')\geqslant\theta F(\bar{x})+(1-\theta)F(\bar{x}')\geqslant\theta\bar{v}+(1-\theta)\bar{v}'$$

这样我们已经找到了一个可行的产出向量，由此产生的收入大于等于这一串不等式最右边的表达式，而最大值 $V(\theta c+(1-\theta)c')$ 只会更大。这就是我们想要证明的结论。

这背后的经济学含义是这样的：G 为凸的排除了生产中的规模经济或专业化，确保了一个加权平均的产量能够通过相同权重的平均投入生产出来。然后，F 为凹的保证了由此产生的收入大于等于各个收入的相同权重的平均值。

由于 V 是凹函数，位于它的函数图形上和以下部分的点的集合为凸集。这是一个 $(m+1)$ 维的集合，即是所有满足 $v\leqslant V(c)$ 的点 (c,v) 的集合。也就是说，使用投入向量 c 进行生产，生产出的产品至少能带来 v 的收入。因而，很自然地，我们可以把它看作是收入的生产可能性集合。图 7.1 中用阴影区域 $\mathscr{A}$ 表示了这个集合，此处 c 是一个标量。由于 V 是非减和凹的，所以沿着这个集合的边界，投入对于收入有正的、但是递减的边际产品。

凸集意味着它能和其他的凸集相分离。有目的地实现这一点的最有效的途径，是在 $\mathscr{A}$ 中选择一点 (c^*,v^*)，使得 $v^*=V(c^*)$。这必定是一个边界点，因为对于任意的 $r>0$，点 (c^*,v^*-r) 在 $\mathscr{A}$ 中，但点 (c^*,v^*+r) 却不在 $\mathscr{A}$ 中。现在将 $\mathscr{B}$ 定义为所有满足 $c\leqslant c^*$ 和 $v\geqslant v^*$ 的点 (c,v) 的集合，它表示除了当 $c=c^*$ 和 $v=v^*$ 时，投入 c 是不可能获得收入 v 的。这样，集合 $\mathscr{B}$ 就起到了与第 6 章中相应集合同样的作用。显然，$\mathscr{B}$ 是一个有非空内点的凸集，并且 $\mathscr{A}$ 和 $\mathscr{B}$ 只有边界点是共同拥有的。因而，我们可以应用分离定理。出于接下来马上就很显然的理由，我把分离超平面的方程写作：

$$\iota v-\lambda c=b=\iota v^*-\lambda c^*$$

其中 ι 为标量，λ 是一个 m 维的行向量，符号的选择满足

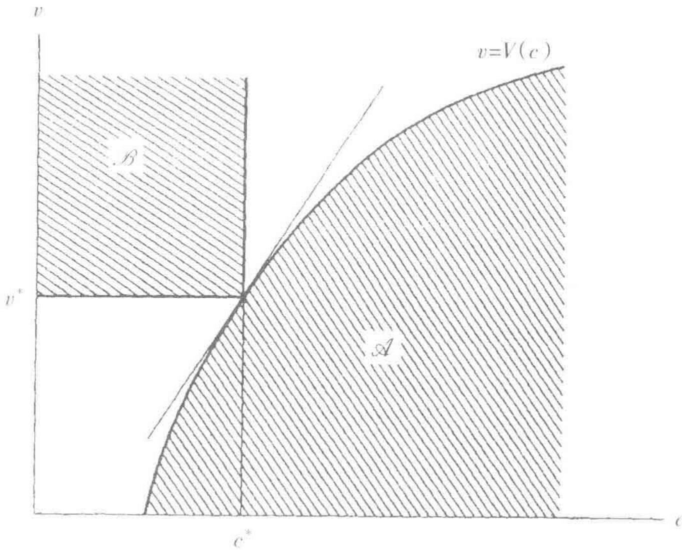

图 7.1 凹规划中的值函数

$$\iota v-\lambda c\begin{cases}\leqslant b, \text{对所有的}(c,v)\in\mathscr{A}\\ \geqslant b, \text{对所有的}(c,v)\in\mathscr{B}\end{cases}\tag{7.3}$$

首先要注明的是，ι 和 λ 必须都是非负的。举个例子，假如 ι 是负的。现在考虑点(c^*, v^*+1)，很显然这个点在 $\mathscr{B}$ 中。我们有：

$$\iota(v^*+1)-\lambda c^*=b+\iota<b$$

与分离性质相矛盾。类似地，对每一个 $i=1,2,\cdots,m$，考虑点(c^*-e^i, v^*)，其中 e^i 为第 i 个分量为1、其他分量为0的向量，我们看到λ^i 不可能为负。

现在一个更微妙(subtle)的问题是：ι 可以为零吗？让我们来看看如果 ι 为零的结果。为了使超平面方程有意义，组合向量(ι,λ)必须为非零。如果 $\iota=0$，那么至少 λ 的一个分量为非零，即为正的。分离超平面的方程就变成$-\lambda c=-\lambda c^*$，或 $\lambda(c-c^*)=0$。对所有 $\mathscr{A}$ 中的点(c,v)，我们有$-\lambda c\leqslant-\lambda c^*$，或者 $\lambda(c-c^*)\geqslant 0$。在约束为标量的情形下，分离线在点 c^* 处是垂直的，集合 $\mathscr{A}$ 整个都位于它的右边。

图 7.2 给出了导致这一结果的两种情形。在这两种情形中，c^* 的左边都没有可行的点；如果要素投入低于这个水平，生产就变得不可能了。在一些实际情况中，不可分性会导致这一结果的出现。这两种情形的区别在于，当 c 从右边向 c^* 趋近时 $V(c)$的表现不同。在情形(a)中，资源的边际收入

产品趋向无穷大，且只有垂直的分离线。而在情形(b)中，资源的边际收入产品的极限是有限的，尽管有垂直的分离线存在，但也存在许多其他的斜率为有限值的分离线，从而有正的 ι。这表明很快将被找到的保证 ι 为正的条件，只是充分的而非必要的。

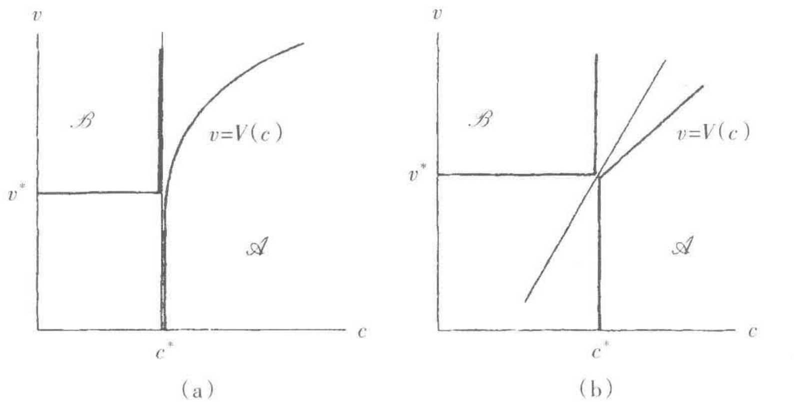

图 7.2　不满足约束规格的情形

很自然地，这个条件要排除不可分性。如果集合 $\mathscr{A}$ 有任何在 c^* 左边的点，那么它在 c^* 点不可能有无穷大的斜率。为此，必然存在一点 x^0，满足 $G(x^0)<c^*$，且 $F(x^0)$ 是有定义的；所以我们可以选择 $(G(x^0), F(x^0))$ 作为 $\mathscr{A}$ 中所要的点。如果有几个约束条件，我们就要求相应的向量不等式 $G(x^0)\ll c^*$。这就是凹规划问题的**约束规格**。它有时也被称为**斯拉特**(Slater)**条件**。

为了证明斯拉特条件意味着一个正的 ι，假定该条件成立但 $\iota=0$，那么至少 λ 的一个分量必定为正。现在 $G(x^0)-c^*$ 的每一个分量都严格为负。那么，

$$\lambda(G(x^0)-c^*)=\sum_{i=1}^{m}\lambda_i(G^i(x^0)-c_i^*)<0$$

因为等式右边至少有一个分量乘积为负的，而所有分量都是非正的。但点 $(G(x^0), F(x^0))$ 在 $\mathscr{A}$ 中，所以根据分离性质，

$$-\lambda G(x^0)=\iota F(x^0)-\lambda G(x^0)\leqslant \iota v^*-\lambda c^*=-\lambda c^*$$

或 $\lambda(G(x^0)-c^*)\geqslant 0$。我们已经证明了相同的表达式同时是负的和非负的，这个矛盾迫使我们得出这样的结论：假设 $\iota=0$ 肯定是错的。

如果我们让 b、ι 和 λ 的每一个分量都乘以同一个正数，分离性质(7.3)

式不变。一旦我们能确信 $\iota \neq 0$,我们就能选择一个比例使得 $\iota=1$。用经济学的术语来说,ι 和 λ 组成了一个影子价格的体系,ι 为收入的影子价格同时 λ 为投入的影子价格。在经济决策中只有相对价格是重要的,将 ι 设为 $\iota=1$,我们选择收入作为计价物。这似乎是一个很显然的选择,我将在稍后采用这一设定。不过有时我们也可能有其他的做法,例如,$F(x)$可能给出了以外国货币衡量的收入,那么 ι 将会是把它转换成本国货币单位的汇率。其实选择计价物的关键在于建立一定的条件,在这些条件下投入的边际收入产品在最优解处是有限的,从而确保我们想用的计价物并非免费商品。

下面,我们观察到,根据分离性质,在所有点$(c,v)\in \mathscr{A}$中,点(c^*,v^*)使$(v-\lambda c)$取得了最大值。这具有一个非常重要的经济含义。如果我们把 λ 解释为投入要素的影子价格向量,那么$(v-\lambda c)$就是当一个生产者用投入 c 生产收入 v 时产生的利润。由于 $\mathscr{A}$ 中所有的点都代表了这类可行的生产计划,这个结果表明了利润最大化的生产者会选择(c^*,v^*)。他并不一定要意识到实际中可供使用的投入限于 c^*,他可以认为自己可以自由地选择任何一个 c,只不过最后的结果恰好选择了 c^*。而价格 λ 使他认识到了资源的稀缺性。这一解释有些特别,但是其中的原理却是非常重要的而且具有普遍的意义:只要恰当的稀缺成本或者约束的影子价值从准则函数中扣除掉,那么约束下的选择可以转换成无约束下的选择。对于经济学家而言,这是凹规划中拉格朗日方法最重要的特征。

我们可以更加正式地肯定 λ 作为影子价格的解释。对于任意的 c,点$(c,V(c))$在 $\mathscr{A}$ 中。所以根据分离性质,我们有,

$$V(c)-\lambda c \leqslant V(c^*)-\lambda c^*$$

或者

$$V(c)-V(c^*) \leqslant \lambda(c-c^*) \tag{7.4}$$

右边的线性函数过高地估计了 V 值的变动。这看上去和凹性性质(7.1)式非常相似,并且意味着 λ 应该等于 $V_c(c^*)$,即 V 在 c^* 处的偏导数向量。这使得 λ 等于最优解处投入要素的边际收入产品,或各个约束的影子价格向量。不过有一个问题依然存在:我们不能确保 V 是可微的。本章到这里为止,甚至还没有假定 F 和 G 是可微的,不过即使我们做了这样的假定,V 也有可能不可微。其中的原因和我们在第 4 章及图 4.1 中看到的

一样。对于参数的不同值域，各个不等式约束可能会以等式的形式成立，并且当解从一种情形转变到另一种情形时，V 的斜率可能会突然改变。

即使在 V 的斜率存在这样的不连续性时，报酬递减概念的一个非常自然的一般化仍然成立。当 c 的某个分量值增加时，V 相应的偏导数会下降，而不是上升。这源于 V 的凹性。

出于分离的考虑，我们在 (c,v) 空间中确定了一个特定的点，并用星号表示，以示区别。而现在星号可以不用了。让我们考虑一个一般的点 $(c,V(c))$ 及其相应的乘数向量 λ。将它和一个邻近点相比较，这个邻近点与原来的点相比较只有第 i 个要素有所增加，表示为 $(c+he^i,V(c+he^i))$，其中 h 是一个正的标量，e^i 是第 i 个元素为 1 而所有其他元素都为零的向量。那么(7.4)式就变为：

$$V(c+he^i)-V(c)\leqslant \lambda he^i=h\lambda_i$$

由于 h 为正，我们可以两边同除以 h，把上式写作：

$$[V(c+he^i)-V(c)]/h\leqslant \lambda_i$$

很容易证明，由于 V 是凹的，上式左边是 h 的一个非递增函数，因而当 h 从正值趋向零时，它必定有极限。在横坐标为 c_i 和纵坐标为 $V(c)$ 的图中，上式左边的表达式就是连接点 $(c,V(c))$ 和其右(因 $h>0$)邻近点的弦的斜率。其极限被定义为 V 关于 c 的第 i 个分量的“右”偏导数，并记作 $V_i^+(c)$。这样我们就已经证明了 $V_i^+(c)\leqslant\lambda_i$。

接下来重复上述过程，只是令 h 为负数。现在两边同除 h 会使不等号的方向反过来，从 h 的负值方向取极限就会得到“左”偏导数 $V_i^-(c)$。这证明了 $V_i^-(c)\geqslant\lambda_i$。把这两者相合并，我们得到了最终结果，它把边际回报递减这个概念一般化了，并将乘子同这些一般化的边际产品联系起来：

$$V_i^-(c)\geqslant\lambda_i\geqslant V_i^+(c) \tag{7.5}$$

图 7.3 给出了当 c 为标量时的情形。

到目前为止，我们一直没有明确讨论选择变量 x 向量，现在就让我们来讨论它。回想起 $\bar{x}$ 在约束 $G(x)\leqslant c$ 下最大化 $F(x)$。令 λ 表示由如图 7.1 中那样的分离超平面得到的投入要素的影子价格向量。点 $(F(\bar{x}),G(\bar{x}))$ 在 $\mathscr{A}$ 中，所以由分离定理得：

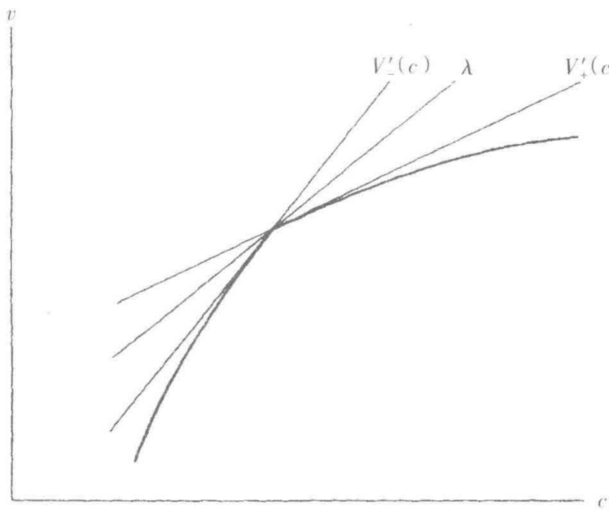

图 7.3　一般化的边际产品

$$F(\bar{x}) - \lambda G(\bar{x}) \leqslant V(c) - \lambda c \tag{7.6}$$

当然 $F(\bar{x})=V(c)$，所以

$$\lambda[c - G(\bar{x})] \leqslant 0 \tag{7.7}$$

现在行向量 λ 的每一个分量都是非负的，并且列向量$[c-G(\bar{x})]$的每一个分量也都是非负的。因此，对于每一个 i，乘积 $\lambda_i[c_i-G^i(\bar{x})]$是非负的。(7.7)式中的内积就是这些项的和。只有当每一项都为零时它才可能是非正的。因此，对于每一个 i，要么 $\lambda_i=0$，要么 $c_i-G^i(\bar{x})=0$。这正是第 3 章中导出的互补松弛条件的概念，它与 λ_i 是第 i 个约束条件的影子价格这一解释完全一致。任何松的约束条件的影子价格必然为零，任何具有正的影子价格的约束一定是紧的。注意到互补松弛条件的含义：(7.6)式和(7.7)式中的不等式事实上必定以等式形式成立。

最后，因为对任意的 x，$(F(x),G(x))$都在 $\mathscr{A}$ 中，并且因为(7.6)式以等式形式成立，由分离性质得到，对所有的 x 有：

$$F(x) - \lambda G(x) \leqslant F(\bar{x}) - \lambda G(\bar{x}) \tag{7.8}$$

也就是说，在没有任何约束的情况下，$\bar{x}$ 最大化 $F(x)-\lambda G(x)$。这是另一种用选择变量表达的论述，即影子价格是怎样把原来的有约束的收入最大化问题转换为一个无约束的利润最大化问题的。

上述所有的推理现在可以总结为这一部分的基本定理：

凹规划的必要条件：假定 F 是一个凹函数而 G 是一个凸函数向量，并且存在一个 x^o 满足 $G(x^o) \ll c$。如果 $\bar{x}$ 在约束 $G(x) \leqslant c$ 下最大化 $F(x)$，那么存在一个行向量 λ，使得：

(i) $\bar{x}$ 在无任何约束的情况下最大化 $F(x)-\lambda G(x)$，且

(ii) $\lambda \geqslant 0$，$G(\bar{x}) \leqslant c$，并且满足互补松弛条件。

这并不要求 F 和 G 有导数。但是如果这两个函数是可微的，那么在(i)中我们有最大化的一阶必要条件，即：

$$F_x(\bar{x}) - \lambda G_x(\bar{x}) = 0 \tag{7.9}$$

如果用第 3 章中的拉格朗日函数 $L(x,\lambda)$ 来表达，(7.9)式就变成 $L_x(\bar{x},\lambda)=0$。这正是有不等式约束情形下的拉格朗日定理的条件(3.5)式。在此，我并没有对选择变量 x 施加任何的非负约束，不过那只是为了使代数的表达式尽量简单。这样的非负约束条件可以被加进来而不会带来任何新的问题，于是我们就得到了库恩—塔克定理相应的条件(3.7)式。我把它作为一个习题留给读者去验证。

凹规划超越了一般的拉格朗日或库恩—塔克条件。第 3 章中给出的条件仅仅要求拉格朗日函数对选择变量的一阶导数为零。这并不足以确保最大化，而且一般也不存在 $\bar{x}$ 最大化了拉格朗日函数的论断。当 F 是凹函数而 G 是凸函数时，上述关于凹规划的定理的(i)部分可以很容易地转化成：对于所有的 x，$L(x,\lambda) \leqslant L(\bar{x},\lambda)$，所以 $\bar{x}$ 确实最大化了拉格朗日函数。这一区别确实有经济学上的意义：众所周知，当存在规模经济时，在给定价格下的利润最大化可能会有问题。在这一情形下，在最优解处，价格必定仍然等于边际成本，但是即使和它相邻的点相比，利润未必已经取得了最大值。同样地，当我们把拉格朗日方法解释为把有约束的收入最大化问题转化为无约束的利润最大化问题时，这一解释也必须限制在凹规划的情形下。

上一段内容似乎表明，在凹规划问题中一阶条件是产生一个最大值的充分条件。事实确实如此。论证可以从两方面进行。

首先，假定 $\bar{x}$ 满足必要条件中的条件(i)和(ii)，那么对于任意 x，我们有：

$$F(\bar{x}) - \lambda G(\bar{x}) \geqslant F(x) - \lambda G(x)$$

或者使用互补松弛条件，

$$F(\bar{x})-\lambda c \geqslant F(x)-\lambda G(x)$$

如果 x 是可行的,$G(x)\leqslant c$,那么

$$F(\bar{x}) \geqslant F(x)+\lambda[c-G(x)] \geqslant F(x)$$

接下来假定 $\bar{x}$ 满足一阶条件(7.9)式。由于 F 是凹函数,G 是凸函数,并且 $\lambda\geqslant 0$,那么 $F-\lambda G$ 就是凹函数。于是,把(7.1)式应用到这个函数,有

$$[F(x)-\lambda G(x)]-[F(\bar{x})-\lambda G(\bar{x})] \leqslant [F_x(\bar{x})-\lambda G_x(\bar{x})](x-\bar{x})$$

但根据(7.9)式,上式的右边等于零。因而有

$$F(x)-\lambda G(x) \leqslant F(\bar{x})-\lambda G(\bar{x})$$

或 $\bar{x}$ 在无任何约束的情况下最大化了 $F(x)-\lambda G(x)$。

这可以总结为如下定理:

凹规划的充分条件:如果 $\bar{x}$ 和 λ 满足以下条件

(i)$\bar{x}$ 在无任何约束的情况下最大化了 $F(x)-\lambda G(x)$,并且

(ii)$\lambda\geqslant 0$ 和 $G(\bar{x})\leqslant c$,满足互补松弛条件。

那么 $\bar{x}$ 在约束 $G(x)\leqslant c$ 下最大化 $F(x)$。如果($F-\lambda G$)是凹函数(反过来它对 F 为凹函数而 G 为凸函数是充分的),那么(7.9)式也意味着上述的条件(i)。

注意到充分条件中没有出现任何的约束规格,它与必要条件的有效性有关。

7.3 拟凹规划

在第 6 章中的分离方法中,F 只是拟凹函数,且 G 的每一个分量约束函数都是拟凸函数,而在本章中到目前为止我们对凹性和凸性都做了更强的假定。事实上,拟凹性和拟凸性的较弱假定对必要条件而言几乎没有什么差异。它们也会产生与上述凹规划一样的充分条件,但是只在一些更技术性的条件下成立,而导出这些技术性条件是非常复杂的。因而我将只讨论拟凹规划的一个限制性的情形,即目标函数为拟凹的、约束函数为线性的情形。当然,与之对称的情形,即线性的目标函数和拟凸的约束函数可以同样地进行处理。第 6 章中的分离方法把一般化的拟凹规划问题分解为一对分散化机制下的决策问题,因而我的分析可以涵盖每一对这样的决策问题。

首先，我们必须建立拟凹函数的一个性质，这一性质同凹函数的"用切线近似会高估"的性质相类似。先看定义：如果 F 为拟凹函数，那么对于任意的 x^a 和 x^b，以及[0,1]中任意的 θ，我们将有：

$$F((1-\theta)x^a+\theta x^b)\geqslant \min(F(x^a),F(x^b))$$

假定 $F(x^b)\geqslant F(x^a)$，那么

$$F(x^a+\theta(x^b-x^a))\geqslant F(x^a)$$

固定 x^a 和 x^b，将上式左边看作一个关于 θ 的函数，比如 $h(\theta)$。那么这个不等式就是：

$$h(\theta)\geqslant h(0)\text{，对于所有的 }\theta\geqslant 0\text{（且}\leqslant 1\text{）}$$

因而 $h'(0)\geqslant 0$。但是根据链式法则，

$$h'(\theta)=F_x(x^a+\theta(x^b-x^a))(x^b-x^a)$$

在 $\theta=0$ 处取值，我们有：

$$F_x(x^a)(x^b-x^a)\geqslant 0 \tag{7.10}$$

上式对所有满足 $F(x^a)\geqslant F(x^b)$ 的 x^a 和 x^b 都成立。注意到 F_x 为行向量，所以上式左边的表达式为内积。

现在考虑在约束 $px\leqslant b$ 下最大化 $F(x)$ 的问题，其中 p 为行向量，b 为一个常数。$\bar{x}$ 为最优解的必要条件是：

$$F_x(\bar{x})-\lambda p=0 \tag{7.11}$$

如果 $\lambda>0$，约束是紧的，并且这是我们需要考虑的惟一情形。(如果取 F 的立方，那么我们就有可能得到一个虚假的稳定点，在该点处(7.11)式以 $\lambda=0$ 成立。不过对于初级的内容而言，这不是一个在经济学上很有意义的情形。)我们的目标是要证明，如果 F 是连续和拟凹的，那么条件(7.11)式也是充分的。也就是说，如果 $\bar{x}$ 和 $\lambda>0$ 满足该条件，那么 $\bar{x}$ 就是这个拟凹规划问题的解。

为了证明这一点，考虑任意满足 $F(x)>F(\bar{x})\equiv\bar{v}$ 的 x，我将证明 x 是不可行的，即会有 $px>b$。把 $x^a=\bar{x}$ 和 $x^b=x$ 代入(7.10)式，那么 $F(x)>F(\bar{x})$ 意味着

$$F_x(\bar{x})(x-\bar{x})\geqslant 0$$

把(7.11)式代入上式并除以正数 λ，得到

$$p(x-\bar{x}) \geqslant 0 \text{ 或 } px \geqslant p\bar{x}$$

换言之，$F(x)$关于 v 的上等值集被包含在位于约束线上或者约束线上方的一半的 x一空间中。

图 7.4 说明了这一点。从几何上看，向量 $F_x(\bar{x})$在 $\bar{x}$ 处正交(垂直)于 $F(x)$的等值线。向量 p 在约束直线 $px=b$ 上的任何一点都与之正交。两条曲线相切的通常条件等价于它们的法向量(normal vector)相平行。这正是(7.11)式所表达的含义，此时这两个向量之间的比例为常数 λ。

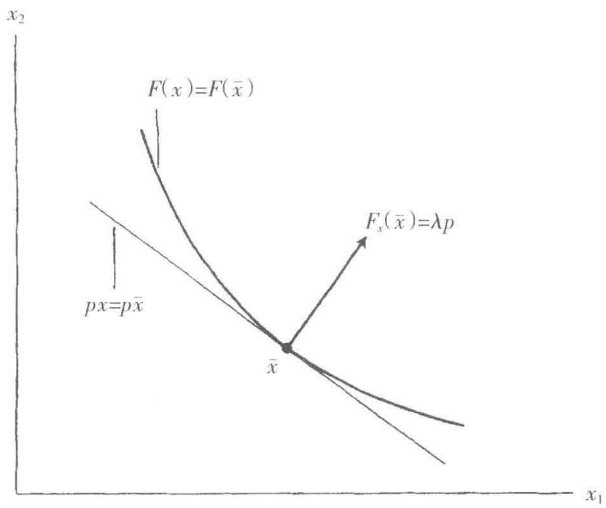

图 7.4　拟凹的目标函数和直线约束

由于 F 是连续的并且 $F(x)>F(\bar{x})$，那么事实上 x 就是上等值集中的一个内点。因而它也是集合 $px\geqslant b$ 的一个内点。换句话说，它满足 $px>b$。所以，任何产生比 $F(\bar{x})$更大的值的 x 都是不可行的，或者说任何可行的 x 产生的值都不会超过 $F(\bar{x})$。证毕。

7.4　惟一性

上述关于凹规划和拟凹规划的充分条件在某种意义上都是弱的，即它们导出了没有其他可行的选择 x 会做得比 $\bar{x}$ 更好。但是，它们没有排除其他的能导出 $F(x)=F(\bar{x})$的可行选择。换句话说，它们没有确立最优解的惟一性。和第 6 章一样，我们需要一个额外的条件，即加强凹函数和拟凹函

数的概念来得出惟一性。

例如，严格凹函数的定义，要求连接函数图形上任意两点的弦上的内点都严格位于函数图形的下方。这排除了所有会使函数图形与弦相重合的直线段。正式地，称 F 为严格凹函数，如果对于不同的两个点 x^a 和 x^b，以及任意的 $\theta\in(0,1)$，我们有

$$F(\theta x^a+(1-\theta)x^b)>\theta F(x^a)+(1-\theta)F(x^b) \tag{7.12}$$

如果凹规划中的目标函数 F 是严格凹的，那么最大化的解 $\bar{x}$ 就是惟一的。其证明可以通过假定存在另一个一样好的选择 $\widetilde{x}$，然后指出$\frac{1}{2}(\bar{x}+\widetilde{x})$可以做得更好。我把这些简单的细节留给读者推导。

例题

例题 7.1：线性规划

凹规划的一个重要的特例是线性规划理论。此处，目标函数和约束函数都是线性的：

$$F(x)=a\,x,\ G(x)=B\,x$$

其中 a 为 n 维行向量，B 为 $m\times n$ 矩阵。
现在

$$F_x(x)=a, G_x(x)=B$$

还需要加上符号约束 $x\geqslant 0$。当约束函数是线性时，就不需要约束规格了。有兴趣的读者可以从附录中库恩—塔克定理的正式推导中看到这个结论的理由。

凹规划的所有条件都是满足的，恰当形式的库恩—塔克条件(3.7)式和(3.10)式既是必要的也是充分的。我们有一个小小的新的关于表示法的说明，即在这个问题中，我们有时需要把拉格朗日乘子也看作变量。所以，它们对应于目前问题的最优解的特定值，都被加上横线以示区别。

拉格朗日函数为

$$L(x,\lambda)=ax+\lambda[c-Bx] \tag{7.13}$$

最优解 $\bar{x}$、$\bar{\lambda}$ 满足条件

$$a-\bar{\lambda}B \leqslant 0, \bar{x} \geqslant 0, \text{满足互补松弛条件} \tag{7.14}$$

$$c-B\bar{x} \geqslant 0, \bar{\lambda} \geqslant 0, \text{满足互补松弛条件} \tag{7.15}$$

在这里,(7.14)式和(7.15)式一共包含了 2^{m+n} 种等式和不等式的组合情形。但并非所有情形都是成立的。一般来说,如果(7.15)式中有 k 个约束以等式形式成立,那么这就对 n 维向量 x 施加了 k 个限制。为了决定 x 的值,我们还需要从(7.14)式中获得 $(n-k)$ 个条件,所以应该恰好有 $(n-k)$ 个非负约束是紧的。在这种情形下,$\bar{\lambda}$ 的相应计数也是正确的。每一对这样的 $(\bar{x}, \bar{\lambda})$ 被称为一个“基础解”。参数空间 (a, c, B) 被分成几个区域,在每一个区域中有一个基础解。在两个这样的区域相交的边界处存在额外的非基础解,此时满足互补松弛条件的不等式对中的一些不等式都以等式形式成立。

从一个区域转换到另一个区域会引起拉格朗日乘子的突然变动,从而会出现我们在图 4.1 中看到的最大值函数的折拐。

现在考虑一个新的线性规划问题:找出一个行向量 y,在约束 $yB \geqslant a$、$y \geqslant 0$ 下使得 yc 取得最小值。其中向量 a、c 和矩阵 B 与前面完全一样。用我们的最大化的术语来说,这个问题可以表述为:

$$\max(-yc), \text{ s. t. } -yB \leqslant -a, y \geqslant 0$$

除了行向量和列向量互换之外,它和上面用 x 表述的问题具有相同的形式。所以,我们可以引入一个列向量乘子 μ,并定义拉格朗日函数

$$M(y, \mu) = -yc + [-a + yB]\mu \tag{7.16}$$

最优解 $\bar{y}$ 和 $\bar{\mu}$ 由充分且必要的库恩一塔克条件来定义:

$$-c + B\bar{\mu} \leqslant 0, \bar{y} \geqslant 0, \text{满足互补松弛条件} \tag{7.17}$$

$$-a + \bar{y}B \geqslant 0, \bar{\mu} \geqslant 0, \text{满足互补松弛条件} \tag{7.18}$$

现在如果我们用 $\bar{\lambda}$ 代替 $\bar{y}$,用 $\bar{x}$ 代替 $\bar{\mu}$,则(7.17)式和(7.15)式完全相同,(7.18)式和(7.14)式完全相同。换句话说,原来那个问题的最优的 $\bar{x}$ 和 $\bar{\lambda}$ 也解出了这个新的问题的最优解,不过它们的角色互换了:$\bar{\lambda}$ 是新问题中的最优选择变量向量,而 $\bar{x}$ 是相应的乘子向量。

这个新的问题被称为原来那个问题的对偶,而原来的问题被称为这对

问题中的原问题。这抓住了经济学中价格和数量之间一个非常重要的经济关系。原问题有标准的解释。令 x 代表产出数量向量，a 代表产出的价格向量或者单位价值向量。矩阵 B 包含了单位投入系数，所以 Bx 就是生产 x 所需的投入向量。最后，c 为要素供给向量。

当我们找到最优解 $\bar{x}$ 之后，相应的 $\bar{\lambda}$ 就是投入要素的影子价格向量。我们刚才看到，在所有满足约束 $\lambda B \geqslant a$ 和 $\lambda \geqslant 0$ 的向量中，$\bar{\lambda}$ 使 λ_c 取得了最小值。这样，影子价格使投入 c 的成本最小。注意到 λB 的第 j 个分量为 $\sum_i \lambda_i B_{ij}$，正是用投入的影子价格来计算的生产一单位商品 j 所需的投入要素组合的成本。所以，约束条件为这样的投入成本向量要大于或等于产出的单位价值向量。换句话说，投入要素的影子价格保证了没有一种商品会产生严格正的利润 —— 经济学中一个标准的"竞争性"条件。

(7.14)式中的互补松弛条件保证了，如果生产商品 j 的单位成本实际上超过了它的价格，即如果以恰当的投入的影子价格来计算时生产会导致损失，那么这种商品的产量为零。反之，如果商品 j 有一个正的产量，那么单位成本正好等于价格，即利润正好为零。

通过下面的观察我们可以给上述结论作一总结：(7.14)式中的互补松弛条件意味着

$$a\bar{x} = \bar{\lambda}B\bar{x}$$

(7.15)式中的互补松弛条件意味着

$$\bar{\lambda}c = \bar{\lambda}B\bar{x}$$

将两者结合起来，我们有

$$a\bar{x} = \bar{\lambda}c$$

这说明了最优产出的价值等于用影子价格来计算的要素供给的成本。这一结果也可以解释为我们熟悉的收入循环，即国民产出等于国民收入。

最后，可以很容易地检验，如果我们把对偶问题作为起始点，并经历一些机械的步骤找到它的对偶问题，那么我们又会回到原问题。换句话说，对偶是"自反的"。

除了一点以外，这大体上就是线性规划的对偶理论。我们把最优解 $\bar{x}$ 作为起始点，而没有考虑最优解是否存在的问题。这可能会导致一些问题，

或者因为约束条件彼此互不相容，或者因为约束条件定义了一个无界的可行集，并且当我们在这个无界集合上计算时目标函数会趋于无穷大。我把这个问题的分析留给更加高深的教科书。

例题 7.2：利润最大化失效

对于标量 x，考虑在约束 $G(x)\equiv x\leqslant 1$ 下最大化 $F(x)=e^x$。由于 F 是增函数，最优解在 $x=1$ 处，应用库恩—塔克条件，得出 $\lambda=e$。

但 $x=1$ 并没有使无约束情形下的目标函数 $F(x)-\lambda G(x)$ 取得最大值。事实上，只要让 x 大于 1，e^x-ex 就可以取得任意大。拉格朗日方法并没有把原来那个约束条件下的最大化问题转化为一个无约束的最大化问题。其问题在于 F 不是凹函数。

习题

习题 7.1：最小化

使用与本章推导凹函数的最大化问题相类似的方法，推导出凸函数的最小化问题理论。

习题 7.2：最大值函数的凸性

令 θ 表示一个参数向量，考虑在约束 $G(x)\leqslant c$ 下选择 x 使 $F(x,\theta)$ 取得最大值的问题。令 $V(\theta)$ 表示作为参数的函数的最大值。证明：如果对于每一个固定的 x，F 是 θ 的一个凸函数，则 V 为凸函数。

在第 5 章中我们从图形（图 5.2）上看到，生产给定产量的最小成本，被看作是投入价格的函数，且它为凹函数。把它作为上述一般结论的推论加以正式推导。

习题 7.3：更多关于线性规划的内容

证明：例题 7.1 的线性规划问题中的最优解 $\bar{x}$ 和相应的乘子向量 $\bar{\lambda}$ 满足

$$L(x,\bar{\lambda}) \leqslant L(\bar{x},\bar{\lambda}) \leqslant L(\bar{x},\lambda)$$

对所有非负的 x 和 λ 都成立。换句话说，当 $\lambda=\bar{\lambda}$ 时，$\bar{x}$ 使拉格朗日函数取得最大值，而当 $x=\bar{x}$ 时，$\bar{\lambda}$ 使拉格朗日函数取得最小值。换言之，在 (x,λ) 空间中拉格朗日函数的图形看上去像一个马鞍。因而 $(\bar{x},\bar{\lambda})$ 被称为拉格朗日函数的鞍点。

令 $V(a,c)$ 代表线性规划问题的最大值函数。证明：对每一个固定不变的 c，V 为 a 的凸函数，而对每一个固定不变的 a，V 为 c 的凹函数。

▶ 进一步阅读

最好的线性规划分析可以继续看：Robert Dorfman, Paul A. Sumuelson, and Robert M. Solow, *Linear Programming and Economic Analysis*, New York: McGraw-Hill, 1958。

关于拟凹规划，最初的那篇论文仍然具有最高的价值：Kenneth Arrow and Alain Enthoven, "Quasi-concave Programming", *Econometrica*, 29(4), 1961, pp. 779—800。

▶8

二阶条件

8.1 局部和全局最大值

利用像凹函数和拟凹函数这样的性质，前一章已经导出了最优解的充分条件。这些性质是定义在全局上的，即函数的整个定义域上。例如，如果一个函数在任意一点的切线都位于该函数的图形上或者图形上方，或者如果函数图形位于连接其任意两点的弦上或者弦的上方，那么这个函数被称作凹的。相应地，这些条件对全局最优解是充分的，满足这些条件的 $\bar{x}$ 至少和任意其他可行的 x 一样好。从某种意义上说这些条件是理想的；如果满足这些条件，那么我们无需担心，在某个较远的地方可能有某个点会比我们正关注的这个点更好。但在许多的应用中，函数在整个定义域上并没有我们所要求的性质。

本章的关注点在于，目标函数和约束函数在所考察的最优值的一个小的邻域内的曲率。这些条件可以用这些函数在该点的二阶导数来表示。这些导数的条件对局部最优解而言是充分的；如果它们成立的话，所考虑的点在其充分小的邻域内就优于任何其他的点。

当全局条件不满足时，这是一个有用的性质。此外，它还有一个很有价值的副产品。在第 4 章中我引入了比较静态分析，作为比较参数发生微小变动时对应的最优解的一般方法。到目前为止，我们只考虑了最大值函数 v 的比较静态分析，还没有考虑最优选择变量 x 的比较静态分析。现在我们重点考虑后者。我们会发现，二阶条件将有助于决定最优选择变量的比较静态反应。因而，我将在这章中展开二阶条件的理论及其在比较静态中

的应用。我们先从无约束的最大化问题这一简单情形开始，然后讨论更复杂的约束最优化理论。

8.2 无约束最大化问题

首先假定我们要选择一个标量变量 x 使得 $F(x)$ 最大，令 $\bar{x}$ 表示最优选择的候选解，并在 $\bar{x}$ 处附近将 $F(x)$ 按泰勒级数展开：

$$F(x) = F(\bar{x}) + F'(\bar{x})(x-\bar{x}) + \frac{1}{2}F''(\bar{x})(x-\bar{x})^2 + \cdots \tag{8.1}$$

最优化的一阶必要条件是 $F'(\bar{x})=0$。利用这一点，我们可以把(8.1)式写作

$$F(x) - F(\bar{x}) = \frac{1}{2}F''(\bar{x})(x-\bar{x})^2 + \cdots \tag{8.2}$$

对足够接近于 $\bar{x}$ 的 x，二阶项就支配了泰勒展开式中的高阶项(隐藏在省略号中)。因而，如果 $F''(x)$ 为正，我们就可以找到一个足够接近于 $\bar{x}$ 的 x，使得 $F(x)>F(\bar{x})$。换句话说，$\bar{x}$ 在一个小的邻域内不会产生 $F(x)$ 的最大值。当然，它也不可能在 F 的整个定义域内产生最大值。上述推导给出了 $\bar{x}$ 产生 $F(x)$ 的局部的或全局的最大值的二阶必要条件，即

$$F''(\bar{x}) \leqslant 0 \tag{8.3}$$

如果这一导数是负的，那么(8.2)式中的二次项为负。在 $\bar{x}$ 周围足够小的区间内，不管高阶项的符号是什么，我们都有 $F(x)<F(\bar{x})$。所以，假如(8.1)式成立，那么

$$F''(\bar{x}) < 0 \tag{8.4}$$

就是 $\bar{x}$ 产生 $F(x)$ 的一个局部最大值的二阶充分条件。

注意到(8.3)式和(8.4)式之间有两个不同之处。其一，前者是一个弱的不等式而后者是相应的严格的不等式。其二，前者是局部或者全局最大值的一个必要条件，而后者仅是局部最大值的一个充分条件。类似的陈述也适用于更一般情形下的二阶条件。因而我将集中讨论二阶条件的局部充分性作用，而把相应的必要性的推导留给读者。

一个满足二阶条件的局部最大值被称为正则最大值。如果一个最大值是"非正则的",即 $F''(\bar{x})=0$,我们就得看更高阶的导数。因而 $F'''(\bar{x})=0$ 就是一个必要条件,而 $F''''(\bar{x})<0$ 是充分条件。但我在此处并不讨论这样复杂的情形。

现假定此最大化问题包含一个参数 θ。一阶条件为

$$F_x(\bar{x},\theta)=0 \tag{8.5}$$

该式隐含地把 $\bar{x}$ 定义为 θ 的一个函数。我们需要知道最优选择将如何随着 θ 的变动而变动。将(8.5)式全微分,我们有

$$F_{xx}(\bar{x},\theta)\mathrm{d}\bar{x}+F_{x\theta}(\bar{x},\theta)\mathrm{d}\theta=0$$

或

$$\mathrm{d}\bar{x}/\mathrm{d}\theta=-F_{x\theta}(\bar{x},\theta)/F_{xx}(\bar{x},\theta) \tag{8.6}$$

在正则最优解处,等式右边的分母为负。因此 $\mathrm{d}\bar{x}/\mathrm{d}\theta$ 的符号与交叉偏导数 $F_{x\theta}$ 在最优解处的符号相同。这表明了二阶条件如何帮助我们评价参数变化对最优选择的数量影响。

作为一个简单的经济学例子,考虑一个利润最大化的企业的需求曲线及收入曲线发生移动时的情形。令 $R(x,\theta)$ 表示一个关于产出 x 和移动参数 θ 的收入函数。设计适当的函数形式使得 R_θ 总是正的:θ 的增加使需求曲线和收入曲线都向上移动。与前面类似的计算表明,如果 $R_{x\theta}$ 为正,即如果 θ 的增加使边际收入向上移动,那么 θ 的增加就会使利润最大化的产出水平 $\bar{x}$ 上升。现在,也完全有可能出现当 θ 增加时,平均收入(需求曲线)上移但边际收入曲线下移的情况;所需要的是一个减小需求弹性的扭曲。那么一个有利于企业的需求的移动将会引起产出水平的下降。这就是初级微观经济学课程中许多有趣的悖论和带小窍门的问题的背后机制。

让我们转到选择变量为向量的最优化情形,但仍然没有约束条件。现在泰勒展开为:

$$F(x)=F(\bar{x})+F_x(\bar{x})(x-\bar{x})+\frac{1}{2}(x-\bar{x})^TF_{xx}(\bar{x})(x-\bar{x})+\cdots \tag{8.7}$$

此处 F_{xx} 是由二阶偏导数 $F_{jk}\equiv\partial^2F/\partial x_j\partial x_k$ 组成的对称方阵,上标 T 代表

把列向量转换成行向量的转置操作，二阶项就是二次型

$$(x-\bar{x})^T F_{xx}(\bar{x})(x-\bar{x}) = \sum_{j=1}^{n}\sum_{k=1}^{n} F_{jk}(\bar{x})(x_j-\bar{x}_j)(x_k-\bar{x}_k) \quad (8.8)$$

在转到最大化问题之前，先看看(8.7)式是如何给出新的凹函数的描述很有帮助。如果 F 是凹的，它就会位于它的切线上或切线以下，或它的值小于等于泰勒展开到一阶项的值。当 x 足够接近 $\bar{x}$ 时，此时由二阶项支配这个差额，要使上述性质为真，我们必须有

$$(x-\bar{x})^T F_{xx}(\bar{x})(x-\bar{x}) \leqslant 0$$

对所有的 x 成立。

二次型 $y^T My$，其中 M 为对称矩阵，如果对于所有的 $y\neq 0$ 它的值为负，那么，该二次型被称为负定的。它对应于 M 的以下条件。考虑 M 的 k 阶主子式，即由任意的 k 行和同行数相应的 k 列元素组成的子矩阵。这样的主子式的行列式的符号为$(-1)^k$，当 k 为奇数时为负，k 为偶数时为正。此外，这对于任意选定的 k 行和相应的列都成立，其中 k 从 1 到 n(矩阵的维数)。事实上，只要检查前主子式，即由最前面的 k 行和 k 列形成的主子式，就足够验证上述性质了。如果对于所有的 y，二次型的值是非正的，那么它被称为半负定的。它的条件就是关于主子式的相应的弱不等式。有了这些定义后，如果 $F_{xx}(\bar{x})$是半负定的二次型的矩阵，我们就说 F 在 $\bar{x}$ 处是凹的。

现在考虑最大化问题。一阶必要条件为 $F_x(\bar{x})=0$，而二阶充分条件为：对于所有的 $x\neq\bar{x}$，(8.7)式右边的二次型是负的，即负定的。相应的必要条件则要求它为半负定的。目标函数并不需要在第 6 章所定义的一般意义上是凹的，但是它必须在点 $\bar{x}$ 处刚才所定义的意义上是凹的。

所有这些都有助于进行比较静态分析。假定有一个参数向量 θ 进入 F 的定义。一阶条件为 $F_x(\bar{x},\theta)=0$。对它全微分，我们有

$$F_{xx}(\bar{x},\theta)\mathrm{d}\bar{x} + F_{x\theta}(\bar{x},\theta)\mathrm{d}\theta = 0$$

它看上去和一个变量的情形时相应的方程非常相像，只不过现在 $\mathrm{d}\bar{x}$ 和 $\mathrm{d}\theta$ 是向量，F_{xx} 和 $F_{x\theta}$是矩阵。$\mathrm{d}\bar{x}$ 的解是

$$\mathrm{d}\bar{x} = -F_{xx}(\bar{x},\theta)^{-1}F_{x\theta}(\bar{x},\theta)\mathrm{d}\theta \quad (8.9)$$

负定矩阵的逆矩阵也是负定的，而且关于其子式符号的信息可以和特定问题中 $F_{x\theta}$的信息相结合，以找出当某些参数变动时选择变量变动的符号。

因此，有多个选择变量和多个参数情形下的结论不像前面那些情形一样简单或一般。不过我在例题 8.4 中会提供一个应用的例子。

8.3 约束最优化

我将从有两个选择变量和一个等式约束的最简单的情形开始，然后论述它的结论如何扩展到更为一般的情形。考虑在约束 $G(x_1, x_2)=c$ 下最大化 $F(x_1, x_2)$，其中 F 和 G 都是自变量的增函数。图 2.1、图 6.1 和图 6.3 已考虑了这一问题的不同方面。关键之处在于这两条等值线穿过最优解 $\bar{x}$ 时的相对曲率，即 F 的等值线应该比 G 的等值线更凸。为了在代数上表达这一点，我们应该把 x_2 看作沿着每一条等值线的关于 x_1 的函数，并找出这个函数的二阶导数。让我们从 $F(x_1, x_2)$ 的等值线开始分析。

一阶导数的表达式可以结合图 2.1 的相切方法得到。方程(2.8)式为

$$\mathrm{d}x_2/\mathrm{d}x_1 = -F_1(x_1, x_2)/F_2(x_1, x_2)$$

现在我们必须把这个式子再次微分，记住等式右边的 x_2 是关于 x_1 的函数。所以

$$\begin{aligned}\frac{\mathrm{d}^2 x_2}{\mathrm{d}x_1^2} &= \frac{\mathrm{d}[-F_1/F_2]}{\mathrm{d}x_1} \\ &= -\frac{F_2(F_{11}+F_{12}\mathrm{d}x_2/\mathrm{d}x_1)-F_1(F_{21}+F_{22}\mathrm{d}x_2/\mathrm{d}x_1)}{F_2^2} \\ &= -\frac{F_2^2F_{11}-2F_1F_2F_{12}+F_1^2F_{22}}{F_2^3}\end{aligned}$$

为求简洁，F 的所有导数中变量 (x_1, x_2) 都被省略了。

类似的表达式也可以沿着约束曲线求二阶导数而得到。$\bar{x}$ 为局部最优解的二阶充分条件为，$\mathrm{d}^2x_2/\mathrm{d}x_1^2$ 沿着 F 的等值线的值应该比它沿着 G 的等值线的值更大。利用一阶必要条件

$$F_j(\bar{x}) = \lambda G_j(\bar{x}), j = 1,2$$

记得我们假定 F_j 和 G_j 都是正的，将上式化简，二阶条件变为

$$G_2^2(F_{11}-\lambda G_{11})-2G_1G_2(F_{12}-\lambda G_{12})+G_1^2(F_{22}-\lambda G_{22})<0 \tag{8.10}$$

用矩阵表示法可以把上式表达得更加简洁：

$$\det\begin{bmatrix} F_{11}-\lambda G_{11} & F_{12}-\lambda G_{12} & -G_1 \\ F_{21}-\lambda G_{21} & F_{22}-\lambda G_{22} & -G_2 \\ -G_1 & -G_2 & 0 \end{bmatrix} > 0 \tag{8.11}$$

其值在 $\bar{x}$ 处计算。

有 n 个选择变量和 $m<n$ 个等式约束的一般问题的二阶条件就是上式的一个直接扩展。在已经建立的矩阵表示法中，我们形成分块矩阵

$$\begin{bmatrix} F_{xx}-\lambda G_{xx} & -G_x^T \\ -G_x & 0 \end{bmatrix}$$

其值当然也是在 $\bar{x}$ 处计算。左上方的一块是一个 $n\times n$ 的矩阵，右下方的一块是一个 $m\times m$ 的零矩阵，其他两个分块分别是 $m\times n$ 和 $n\times m$ 的矩阵。

考虑由这个矩阵的最后 k 行和相应的列组成的方的子矩阵。我们可以让 k 从 1 变化到$(m+n)$，$k=m+n$ 时的子矩阵就是整个矩阵。当 k 很小时，子矩阵就会是奇异的，因为它的右下角有大量的零值。但是那些由 $k=2m$ 或者更大的 k 形成的子矩阵就未必是奇异的了。于是，局部最大值的二阶充分条件对这些矩阵的行列式的符号施加了限制，即要求符号依此变化，且第一个子矩阵（由最后 $2m$ 行和相应的列组成）的符号为$(-1)^m$。

我省略了对此的证明，但是可以把早先的结果(8.11)式作为一个特例来验证。当 $n=2$，$m=1$ 时，我们只需考虑两个子矩阵。由最后的两行两列组成的子矩阵的行列式是 $-G_2^2$，其值为负，从而自动满足其符号应该为 $(-1)^1$ 的要求。接下去的行列式就是整个矩阵的行列式，交替规则要求它为正，而这正是(8.11)式所表达的。

注意到连续的子矩阵是从右下角而不是从左上方开始的。这样，$(F_{xx}-\lambda G_{xx})$就没有被涉及。它并不需要有任何特定符号的行列式，更不用说必须是负定的行列式了。这样，$(F-\lambda G)$不需要是凹的，而且当一阶必要条件表明 $\bar{x}$ 给出了$(F-\lambda G)$的一个稳定点时，$\bar{x}$ 并不需要使它取得最大值。第 7 章中提到了这种可能性，并在例题7.2中进行了说明。现在我们可以更清楚地看到为什么会产生这种情况。

和通常一样，二阶条件和比较静态分析问题紧密相连。在上述问题中

的函数 F 和 G 中加入一个 s 维的参数向量 θ。那么一阶条件为

$$F_x(\bar{x},\theta)-\lambda G_x(\bar{x},\theta)=0$$

其也满足约束条件，所以

$$G(\bar{x},\theta)=c$$

最优选择向量 $\bar{x}$ 以及乘子向量 λ 可以随着 θ 的变化而变化。对一阶条件全微分，我们得到

$$\begin{aligned}&\sum_{k=1}^{n}(\partial^2F/\partial x_j\partial x_k)\mathrm{d}\bar{x}_k+\sum_{r=1}^{s}(\partial^2F/\partial x_j\partial\theta_r)\mathrm{d}\theta_r\\&-\sum_{i=1}^{m}\lambda_i\Big\{\sum_{k=1}^{n}(\partial^2G^i/\partial x_j\partial x_k)\mathrm{d}\bar{x}_k+\sum_{r=1}^{s}(\partial^2G^i/\partial x_j\partial\theta_r)\mathrm{d}\theta_r\Big\}\\&-\sum_{i=1}^{m}\mathrm{d}\lambda_i\partial G^i/\partial x_j=0\end{aligned}$$

这个复杂的表达式，以及约束条件全微分后得到的一个较简单的表达式，都可以用矩阵的表示法表达成紧凑得多的形式：

$$\begin{bmatrix}F_{xx}-\lambda G_{xx} & -G_x^T\\ -G_x & 0\end{bmatrix}\begin{bmatrix}\mathrm{d}\bar{x}\\ \mathrm{d}\lambda^T\end{bmatrix}=-\begin{bmatrix}F_{x\theta}-\lambda G_{x\theta}\\ -G_\theta\end{bmatrix}\mathrm{d}\theta \tag{8.12}$$

当然所有导数的值都在$(\bar{x},\theta)$处计算。

等式右边的那个分块矩阵与二阶条件中涉及的那个矩阵相同，这并不是让人吃惊的事。这些条件再一次给了我们关于解的一些信息。它们的作用在具体的例子中可以得到最好的证明，我会在例题 8.4 中给出证明。

最后，考虑不等式约束的情形：在约束 $G(x)\leqslant c$ 下选择 x 使 $F(x)$ 取得最大值。在第 3 章中我们看到，进入函数 F 和 G 的定义的参数 θ 的空间可以划分为几个区域，在每一个区域中，约束的一个子集是紧的（以等式形式成立），而其余的是松的。不同的区域有不同的松紧约束的情形。只要最初的参数结构是位于这些区域中某个区域的内部，我们就可以考虑对它采取一个小的偏离，把紧的约束就当作上面的等式约束理论一样来处理，而忽略那些在整个过程中仍然为松的约束。但是，如果最初的点是位于两个具有不同松紧约束情形的区域的边界上，那么偏向一边时就得用一组等式来分析，而偏向另一边时就得用另一组等式来分析。我们这种以普通的泰勒展开为基础的方法将无法处理这一情况。事实上，在这样一些情形下很难说

有什么系统的方法；每一种情形都必须视具体情况来处理。所幸的是，只有在极其特殊的参数结构下，最优解才会落在两个区域的边界上，所以我们无需太担心这些情况。

8.4 包络性质

在第 5 章中，我们建立了最优化问题的最大值函数的包络性质：

$$V(\theta)=\max_{x}\{F(x,\theta)\mid G(x)\leqslant c\}$$

它是一族函数 $F(x,\theta)$的上包络线，其中每一个函数中 x 都是固定的。如果 x^1 对于 θ_1 恰是最优解，那么 $V(\theta)$和 $F(x^1,\theta)$在 θ_1 处相切。图 5.1 给出了这种情形。这两个函数的曲率的关系从那张图上看也很明显：V 比每一个 F 都更凸。这个二阶包络性质就是这个部分的主题。

第 5 章随后考虑了一个更一般的问题，即向量 x 划分成子向量(y,z)的情形。在被解释为短期的子问题中，变量 y 总是随着 θ 的变动而最优地变动，而 z 则保持不变。我们看到，假如短期和长期最大值函数都是可微的，当短期中固定的变量正好是它们的长期最优值时，它们的斜率相同。现在一个相关的二阶性质本身表明：固定不变的变量数越少，最大值函数越凸。这就是我将推导出来的结论的形式。

继续采用本章的表示法，令 $Z(\theta)$为变量 z 的长期最优值，$V(\theta)$为长期最大值函数，$V(z,\theta)$为短期最大值函数，当 θ 变化到 θ'时，我们有

$$V(Z(\theta),\theta')\leqslant V(Z(\theta'),\theta')=V(\theta')$$

等式两边都在 θ 附近用泰勒级数展开，我们有

$$V(Z(\theta),\theta)+V_\theta(Z(\theta),\theta)(\theta'-\theta)+\frac{1}{2}V_{\theta\theta}(Z(\theta),\theta)(\theta'-\theta)^2+\cdots$$

$$\leqslant V(\theta)+V_\theta(\theta)(\theta'-\theta)+\frac{1}{2}V_{\theta\theta}(\theta)(\theta'-\theta)^2+\cdots$$

一阶包络性质允许我们同时删去等式两边的第一项和第二项。这剩下

$$(V_{\theta\theta}(Z(\theta),\theta)-V_{\theta\theta}(\theta))(\theta'-\theta)^2+\cdots\leqslant 0$$

取 θ'足够接近于 θ，二次项就支配左边余下的展开项（隐藏在省略号中）。为了使这个不等式在此情形下成立，我们必须有

$$V_{\theta\theta}(Z(\theta),\theta) \leqslant V_{\theta\theta}(\theta) \tag{8.13}$$

该式表明，在长期最大值函数和短期最大值函数相切的那一点处，前者至少和后者一样凸。对于合适的"正则"最大解，我们有像(8.13)式那样的严格不等式。我不再进一步地分析，但会在例题 8.2 中给出(8.13)式的一个重要的应用。

例题

例题 8.1：消费者理论

在例题 5.2 中，消费者的支出函数 $E(p, u)$被定义为价格为 p 时，为了实现效用水平 u 所需的最小支出。补偿需求函数 $C(p, u)$是这个成本最小化问题的解的数量向量。包络定理意味着

$$C(p,u) = E_p(p,u)$$

即支出函数对价格的偏导数向量。

在例题 6.2 中我们证明了支出函数是凹的。现在我们知道可以用二阶导数来刻画凹函数，并可以以此得到补偿需求函数的有用的性质。

将上式微分，我们有

$$C_p(p,u) = E_{pp}(p,u) \tag{8.14}$$

由于等式右边的二阶导数矩阵是对称的，我们推知价格变动的替代效应具有对称性：

$$\partial C^j/\partial p_k = \partial C^k/\partial p_j = E_{jk}$$

接下来，由于 E 是凹函数，等式右边的矩阵就是负半定的。特别地，它的对角项，即 1×1 维的子式，必定小于或等于 0，所以，对于所有的 j，我们有

$$\partial C^j/\partial p_j \leqslant 0 \tag{8.15}$$

换句话说，价格变化的自替代效应为非正。

相同的结论甚至可以更简单地从最大值概念本身得到。"显示性偏好"不等式的运算表明，最优选择至少和其他可行的选择一样好，这会给我们带来想要的结论。假定 p^a 和 p^b 为两个价格向量，x^a 和 x^b 为相应的补偿需

求。两者都能获得相同的效用水平 u。当价格为 p^a 时，x^a 支出得更少，同样，当价格为 p^b 时，x^b 支出得更少，因此我们有

$$p^a x^a \leqslant p^a x^b, p^b x^b \leqslant p^b x^a$$

两不等式相加并化简得到，

$$(p^b - p^a)(x^b - x^a) \leqslant 0 \tag{8.16}$$

这是(8.15)式的更一般的形式：如果 p^a 和 p^b 只在第 j 个分量上不同，那么(8.16)式中的乘积就缩减为：

$$(p_j^b - p_j^a)(x_j^b - x_j^a) \leqslant 0$$

这表明任何价格变动的自替代效应都是非正的。这个方法在另一层意义上更具一般性：它不要求可微性和拟凹性等假定。因而，凡是在可能的情况下，你都应该使用"显示性偏好"的方法。

例题 8.2：勒夏特里埃—萨缪尔森原理

再次考虑消费者的支出最小化问题，这次我们关注的是二阶包络性质。考虑任何一个价格的变动，比如说 p_1。比较如下两种情形。在第一种情形中，所有商品的数量都可以自由地作出最优的变动。在第二种情形中，有一种商品的数量，比如说 x_2，必须固定在它原来的最优水平上。每一种情形都有各自的支出函数。一阶包络性质表明，作为 p_1 的函数，这两个函数会在原来的点处相切。二阶性质则表明，第一个问题(此时的选择更加自由)的支出函数会更加凹(记得这是一个最小化问题)：它对 p_1 的二阶导数会更负一些。但是，每一个问题中的一阶导数都是在那个情形下对 x_1 的补偿需求。因而

$$\left|\frac{\partial x_1}{\partial p_1}\right|_{x_2^{\text{free}}} \geqslant \left|\frac{\partial x_1}{\partial p_1}\right|_{x_2^{\text{fixed}}} \tag{8.17}$$

换言之，固定某个其他的商品，如商品 2 的数量会使商品 1 的补偿需求对它自身价格的反应较不敏感。粗略地说，任何施加给经济中某一个部门的刚性约束会引起其他部门对价格反应的敏感性下降。不论商品 1 和商品 2 是替代的还是互补的，这一结论都是成立的。这就是著名的勒夏特里埃—萨缪尔森原理。

例题 8.3：派生需求

生产者的成本函数可以用与消费者的支出函数相类似的方法来定义。如果 y 为由投入向量 x 和生产函数 $f(x)$ 生产出来的标量产出，用 w 表示投入要素价格的行向量，同时定义

$$C(w, y) = \min_x \{wx \mid f(x) \geqslant y\} \tag{8.18}$$

这个成本函数的性质可以通过与支出函数的性质相类比而得到（见例题 5.2、6.2 和 8.1）。它是所有自变量的增函数，对每一个固定的 y，它是 w 的一次齐次的凹函数。使成本最小化的投入选择向量 x，可以通过成本函数对投入价格求导得到：

$$x = C_w(w, y) \tag{8.19}$$

在生产问题的情形下，还有一个有趣之处。即产出有一个天然的标尺而效用没有，因而规模报酬的概念与生产者的成本函数相关，但与消费者的支出函数无关。特别地，如果规模报酬不变，那么成本与产出成比例，并且成本函数具有相乘可分的形式：

$$C(w, y) = yc(w) \tag{8.20}$$

函数 $c(w)$ 现在是生产一单位产出的最小成本。在初级微观经济学中，当成本曲线为水平时，就说平均成本等于边际成本；这里我们可以更进一步，认识到这条成本曲线的高度依赖于投入的价格。

现在考虑一个行业的竞争均衡。该行业有上述的成本曲线和一个需求曲线 $D(p)$。当每一个企业的平均成本或边际成本在 $c(w)$ 处为水平时，该行业的平均成本也为 $c(w)$。它与需求曲线的交点决定了行业的均衡。即产品价格等于平均成本（边际成本）：

$$p = c(w)$$

从需求曲线我们可以得出产量：

$$y = D(p)$$

最后，通过在一般结果(8.19)式中使用特殊形式(8.20)式得到投入的需求：

$$x = yc_w(w)$$

当产品市场处于均衡时，我们可以通过连续替代，得出该行业对投入的需求，它是关于投入价格的函数：

$$x = D(c(w))c_w(w) \tag{8.21}$$

这被称为"派生需求",接下来的问题就是找出它的导数。

由链式法则,得:

$$\partial x_j / \partial w_k = Dc_{jk} + D' c_k c_j$$

其中我省略了函数中的变量以求简洁。把它写作弹性的形式更好:

$$\frac{w_k}{x_j} \frac{\partial x_j}{\partial w_k} = \theta_k (\sigma_{jk} - \eta) \tag{8.22}$$

其中 η 为行业需求曲线的弹性:

$$\eta = - pD'(p)/D(p)$$

θ_k 是第 k 种投入在平均成本中的份额:

$$\theta_k = w_k x_k / (yc(w))$$

σ_{jk} 是投入 j 和 k 之间的替代弹性:

$$\sigma_{jk} = cc_{jk} / (c_j c_k)$$

我们在习题 5.2 中已考察了它的一种特例。

为了解释(8.22)式,把 w_k 对 x_j 的影响分成两部分是有帮助的。第一部分是替代效应:当要素的相对价格变动时,使成本最小化的要素比例也会发生变化。和在消费者理论中一样,当 $j=k$ 时,替代效应的符号是确定的。C 为凹函数意味着 $c_{kk} \leqslant 0$,因而 $\sigma_{kk} \leqslant 0$。当 $j \neq k$ 时,该效应就取决于投入 j 和 k 是替代的还是互补的。另一个为产出效应。w_k 的增加提高了整个的平均成本(在此例中也就是边际成本),沿需求曲线减少了均衡产出,从而减少了对所有要素的需求。

例题 8.4:二阶条件的使用

考虑一个企业在 w 的价格下购买了投入向量 x,生产了产出 $y=f(x)$,最后将其出售获得收入 $R(y)$。它的利润可以表达为选择变量 x 和投入价格(参数)w 的函数:

$$F(x, w) = R(f(x)) - wx$$

利用通用的公式(8.9)式我们可以找出 w 的变动对最优选择 x 的影响。现在用 w 替代 θ,有

$$F_x = -w, F_{xw} = -I$$

其中 I 为单位矩阵。故

$$\mathrm{d}\bar{x} = F_{xx}(\bar{x}, w)^{-1}\mathrm{d}w^T$$

上式中出现转置是因为 $\mathrm{d}w$ 是一个和 w 一样的行向量。于是

$$\mathrm{d}w\mathrm{d}\bar{x} = \mathrm{d}wF_{xx}(\bar{x}, w)^{-1}\mathrm{d}w^T$$

利用二阶必要条件,这个二次型是负半定的,因而有 $\mathrm{d}w\mathrm{d}\bar{x} \leqslant 0$。如果这个最大值是正则的,也就是说如果满足二阶充分条件,那么这个二次型是负定的,我们有更强的结论,即 $\mathrm{d}w\mathrm{d}\bar{x} < 0$。

现在考虑消费者在预算约束 $px \leqslant I$ 下最大化效用 $U(x)$。一阶条件为

$$U_x(\bar{x}) = \lambda p$$

只要效用函数是递增的,预算约束就以等式形式成立,即

$$p\bar{x} = I$$

我们想找出价格变动的纯替代效应。所以让价格变动 $\mathrm{d}p$,同时让收入变动 $\mathrm{d}I = \mathrm{d}p\bar{x}$ 以补偿消费者,其中 $\bar{x}$ 是在原先的 (p, I) 处的最优选择。那么全微分一阶条件和预算等式,就得到了一般结论(8.12)式的一个具体情形:

$$\begin{bmatrix} U_{xx} & -p^T \\ -p & 0 \end{bmatrix} \begin{bmatrix} \mathrm{d}\bar{x} \\ \mathrm{d}\lambda \end{bmatrix} \begin{bmatrix} \lambda\mathrm{d}p^T \\ 0 \end{bmatrix}$$

于是

$$\mathrm{d}p\mathrm{d}\bar{x} = \frac{1}{\lambda}\begin{bmatrix} \mathrm{d}\bar{x}^T & \mathrm{d}\lambda \end{bmatrix} \begin{bmatrix} U_{xx} & -p^T \\ -p & 0 \end{bmatrix} \begin{bmatrix} \mathrm{d}\bar{x} \\ \mathrm{d}\lambda \end{bmatrix}$$

当正则最大值的二阶充分条件成立时,该式为负。这是另一种确定消费中自替代效应符号的方法。

习题

习题 8.1:生产者理论

在习题 6.2 中我们考察了企业的利润函数

$$\prod(q,p)=\max_{x,y}\{qy-px \mid G(x,y)\leqslant 0\}$$

其中 q 和 p 分别代表产出和投入的价格向量，y 和 x 是相应的数量向量。而约束则反映了技术上的可行性。在那里我们证明了Π是(q,p)的凸函数。

现在，请证明 y 和 x 的最优选择可以用Π的偏导数形式给出：

$$y=\prod{}_q(q,p),x=-\prod{}_p(q,p)$$

进而证明，对于所有的 j 和 k，产出供给曲线是向上倾斜的，要素需求曲线是向下倾斜的：

$$\partial y_j/\partial q_j\geqslant 0,\ \partial x_k/\partial p_k\leqslant 0$$

习题 8.2：关于派生需求的更多内容

考虑一个类似于例题 8.3 中的企业的竞争性企业，但是没有规模报酬不变的假定。假定它的总成本函数为 $C(w,y)$，从而边际成本为 $C_y(w,y)$。假定边际成本曲线是递增的，或 $C_{yy}>0$。给定产出价格向量 p 和投入价格向量 w，企业最大化其利润。找出决定其投入需求向量 x 的一组方程。考察交叉偏导数$\partial^2 C/\partial y\partial w_k$ 在决定$\partial x_j/\partial w_k$ 的符号中的作用，并解释你的结论。

习题 8.3：最小化

类似于正文中的最大化问题，推导无约束和约束最小化问题的二阶条件。你需要定义正定和正半定的二次型，以及它们矩阵的主子式的符号。

▶ 进一步阅读

关于消费者和生产者理论的一个更详细的讨论见 Varian，*Microeconomic Analysis*（1984，见本书第 3 章中的进一步阅读），第 1 和第 3 章，或者 James M. Henderson and Richard E. Quandt，*Microeconomic Theory：A Mathematical Approach*，3rd edn.，New York：Mcgraw-Hill，1980，第 2—5 章。

比较静态分析中使用二阶条件是由萨缪尔森首先提出来的，他那篇经典论文仍然值得一读。

Paul A. Samuelson, *Foundation of Economic Analysis*, Cambridge, MA: Harvard University Press, 1947。

对这些问题的进一步思考，包括对勒夏特里埃—萨缪尔森原理的讨论可以见他的诺贝尔奖获奖演说，Paul A. Samuelson, "Maximum Principles in Analytical Economics", *American Economic Review*, 62(3), 1972, pp. 249—62。

该原理的扩展和应用，见：Eugene Silberberg, "The LeChatelier Principle as a Corollary to a Generalized Envelope Theorem", *Journal of Economic Theory*, 3(2), 1971, pp. 146—55。

在使用最优化理论时，下面这篇文献可以帮助你避开某些陷阱，Knut Sydseter, "Letter to the Editor on some Frequently Occurring Errors in the Economic Literature Pertaining to Problems of Maxima and Minima", *Journal of Economic Theory*, 9(4), 1974, pp. 464—66。

▶9 不确定性

9.1 期望效用

从正式的意义上说,不确定性下的最优化理论并不需要任何像前面那样的新的数学理论。选择变量以客观或主观的概率分布产生随机的结果。目标函数是适当的概率加权平均值(期望值)。这些选择变量也受到某些其他的约束。在前面八章中导出的一般理论在本章中可以继续应用。事实上,不确定性下的选择问题的特殊结构会产生特殊的结论,而这些结论在约束最优化的一般数学理论中无法获得。

让我们将这些想法说得更精确些。假设在手头的决策已经做出来了,这个世界可以按照许多不同方式中的任何一种方式演进。这被称为世界的不同状态或者基本事件。假设这些状态首先是离散的和数量有限的,表示为 $i=1,2\cdots,m$。记 p_i 为它们的概率,客观概率或主观概率,视考虑的问题而定,且这些概率非负而且相加为一。在各种情形中,与经济相关的实现结果通常是收入水平、财富或决策者应得的利润,用 Y_i 表示这些结果。在绝大部分时候,我将 Y_i 看作是标量,但在通常意义上它们也可以是向量。由此我们可以写出一个一般的目标函数:

$$F(Y_1,Y_2,\cdots,Y_m;p_1,p_2,\cdots,p_m)$$

选择变量或控制变量 x 将影响部分的或者全部的 Y_i 和 p_i;变量 x 也可能受到一些另外的约束。因此最大化问题能够用我们熟知的方法来解决。

在对偏好施加某些限制后,目标函数就能被表达成一个特殊的形式,即

在不同状态下各实现结果的效用函数 U 的数学期望(概率加权平均):

$$\sum_{i=1}^{m} p_i U(Y_i) \tag{9.1}$$

函数 U 被称为(收入、财富或利润的,视情况而定)冯·诺伊曼一摩根斯坦效用函数,而整个表达式(9.1)式被称为期望效用。

这个表述非常有用,不仅因为它的简洁,而且因为它能够抓住人们的诸多行为中那些在经济学上有意义的行为,例如,风险规避行为。在实际应用中,它几乎无处不在。因此,我讨论的所有内容都以之为基础。由于我关注的是最优化的技术问题,因此我将不去陈述或者讨论在怎样的条件下期望效用最大化是有效的,但我向读者推荐在本章末所引用的阿罗(Arrow)和克瑞普斯(Kreps)的书。近来的研究已经开始发展出更加一般且易于处理的其他方法;参见本章末引用的莫克那(Machina)所作的综述。

在大部分的情形下,人们不希望决策者对得到 Y 的整个值域中的某个值所包含的风险漠不关心。假定现在仅存在两个状态,有不同的实现结果 Y_1 和 Y_2,以及相应的正的概率 p 和 $(1-p)$。将此与一个在保险精算上等价的结果进行比较,即可以确定性地获得 Y 的数学期望$[pY_1+(1-p)Y_2]$。风险规避的决策者将偏好确定的值。即,

$$U(pY_1+(1-p)Y_2) > pU(Y_1)+(1-p)U(Y_2)$$

这恰恰表明了 U 在范围$[Y_1, Y_2]$内是(严格)凹的。更为一般地,严格凹性意味着,当概率是正的并且实现结果不同时,有

$$U(\sum_{i=1}^{m} p_1 Y_i) > \sum_{i=1}^{m} p_i U(Y_i) \tag{9.2}$$

如果 U 是二次可微的,$U''<0$ 对应于风险规避。

再次,决策变量 x 会影响某些或者全部的产出和概率,于是,我们就有了一个最优化问题。几个简洁的例子就能说明其中的经济学直觉。

首先假设 $Y_1<Y_2$,那么第一个状态会引起一些损失,至少相对于第二个状态而言。一个自然的反应就是购买保险。假设如果状态 1 不幸发生,预付保费 x 会让你得到 X。如果此保险业是完全竞争的,并且每个企业能够分散大量的独立风险,则保险从精算意义上说就是公平的。因此 $pX=x$,或 $X=x/p$。因而 Y_1 变为(Y_1-x+x/p),Y_2变为 Y_2-x;记住保费是提前支付的,而且在两个状态下都会支付。目标函数的值变为

$$pU(Y_1 - x + x/p) + (1-p)U(Y_2 - x)$$

用链式法则可以找到 x 为最优解的一阶条件：

$$pU'(Y_1 - x + x/p)(1/p - 1) - (1-p)U'(Y_2 - x) = 0$$

即

$$U'(Y_1 - x + x/p) = U'(Y_2 - x)$$

如果 $U''<0$，那么该条件也是充分的，并且意味着 $Y_1 - x + x/p = Y_2 - x$。这样一个风险规避的人将购买保险精算上公平的保险直到在各个不同的状态下收益相等为止。他将完全保险或保值。

其次，假设通过一个预先的支出 z 可以降低坏的结果 1 的概率。在具体的背景下，这可能意味着使用一个更可靠但更昂贵的产品，或在有风险的活动中更加谨慎小心，而这种小心翼翼会带来负效用。现在我们令 p 为 z 的一个函数；这是一个递减的函数，又由于 p 是有下界的，这个函数一般为凸的。这时，假定不能获得保险。目标函数可以写成

$$\emptyset(z) \equiv p(z)U(Y_1 - z) + [1 - p(z)]U(Y_2 - z)$$

则

$$\begin{aligned}\emptyset'(z) =& - p'(z)[U(Y_2 - z) - U(y_1 - z)] \\ & - \{p(z)U'(Y_1 - z) + [1 - p(z)]U'(Y_2 - z)\} \qquad (9.3)\end{aligned}$$

等式右边的第一项是谨慎行为或使用高质量产品的期望边际收益，它是坏结果的概率的边际减少和两种结果的效用的差额的乘积。另一项为谨慎行为或使用高质量产品的期望边际成本。最优的 z 值可以用一阶条件来定义。

最后，假定保险和谨慎行为都是存在的。保险公司不能辨别人们是否谨慎行事，而只能观察到结果。如果存在保险精算上公平的保险，目标函数为

$$\emptyset(x,z) \equiv p(z)U(Y_1 - z - x + x/p(z)) + [1 - p(z)]U(Y_2 - z - x)$$

根据和前面相同的推理，关于 x 的一阶条件为：

$$Y_1 - z - x + x/p(z) = Y_2 - z - x$$

令这两个共同的值为 Y_0。于是，类似于(9.3)式，在 x 为最优解时，我们有

$$
\begin{aligned}
\varnothing_z(x,z) =& -p'(z)[U(Y_2-z-x)-U(Y_1-z-x+x/p(z))] \\
& -\{p(z)U'(Y_1-z-x+x/p(z)) \\
& +[1-p(z)]U'(Y_2-z-x)\} \\
=& -p'(z)[U(Y_0)-U(Y_0)]-U'(Y_0) \\
=& -U'(Y_0)<0
\end{aligned}
$$

综上所述，当存在完全保险时，谨慎行为所获得的边际收益消失了，而边际成本仍然为正。因而谨慎行为的最优值位于角点 $z=0$ 处。换言之，完全保险的存在破坏了采取有成本的谨慎行为的激励。用保险业的术语来说，这就是大家熟知的“道德风险”。实际上，当出现道德风险时，只存在部分的保险。

这是信息不对称此类一般问题的一个例子：一方拥有另一方所没有的关于某些相关变量，如风险、努力或质量的信息。因而在古典的竞争性市场中的交易就不可能发生了；人们必须设计出一个合约或者激励计划以尽可能地绕过信息不对称问题。在过去的十年中，大量新的研究领域已经在有这种情形的经济理论中展开了。我在例题 9.1 和 9.2 中讨论了这类问题中的两个简单的例子，并在本章末提供了可供进一步阅读的参考文献。

本章余下的内容是研究在不确定性下进行选择的传统情形，特别是投资组合的选择。在某种程度上，连续随机变量比有限数量状态更易于处理。因而我们用一个定义在范围 $[\underline{r},\bar{r}]$上的随机变量 r 代替指数 i，用密度函数 $f(r)$代替概率 p_i，并用

$$
E[U(Y)]=\int_{\underline{r}}^{\bar{r}}U(Y(r)f(r))\mathrm{d}r \tag{9.4}
$$

代替期望效用表达式(9.1)式。选择变量 x 就变成函数 Y 和 f 中一个额外的变量。风险规避的解释与(9.2)式相类似。Y 的数学期望为

$$
E[Y]=\int_{\underline{r}}^{\bar{r}}Y(r)f(r)\mathrm{d}r
$$

因此，$U''<0$ 意味着

$$
U(E[Y])>E[U(Y)] \tag{9.5}
$$

这是詹森不等式(Jensen's Inequality)的一个应用。

9.2 一种无风险资产和一种风险资产

一个投资者拥有初始财富 W_0。投资于风险资产 x 产生总收益(本金加利息)$x(1+r)$,其中 r 为一个随机变量,其密度函数为 $f(r)$。无风险资产支付零利息;但这个假设可以被一般化,例如支付一个固定的利息,但那样会使得表示法有一点混乱。现在,最终的(随机)财富为

$$W = (W_0 - x) + x(1+r) = W_0 + xr \tag{9.6}$$

x 的数量必定在范围$[0,W_0]$内;卖空和以无风险利率贷款投资于风险资产是不允许的。投资者有严格凹的冯·诺伊曼—摩根斯坦效用函数 U,选择 x 以最大化

$$E[U(W)] = \int_{\underline{r}}^{\bar{r}} U(W_0 + xr) f(r) \mathrm{d}r \tag{9.7}$$

令 $\varnothing(x)$标记此表达式,并将其视为关于 x 的一个函数。那么

$$\varnothing'(x) = \int_{\underline{r}}^{\bar{r}} rU'(W_0 + xr) f(r) \mathrm{d}r = 0$$

特别地,

$$\varnothing'(0) = U'(W_0)\int_{\underline{r}}^{\bar{r}} rf(r)\mathrm{d}r = U'(W_0)E[r]$$

注意到 $U'(W_0)$为非随机变量,因而能够将其拿到积分号(求和符号)之外。如果风险资产的利率的数学期望(概率性平均值)是正的,则 $\varnothing'(0)$为正,那么 x 的最优值就不可能为零。换言之,如果有一个保险精算上公平的投资,风险规避的投资者至少会购买其中的一部分。

如果$\underline{r}>0$,那么对所有的 x,$\varnothing'(x)>0$,而且将所有的 W_0 投资于风险资产是最优的。更典型的情形是,投资者对每一种资产都会持有一些。关于 x 的一阶条件为

$$\int_{\underline{r}}^{\bar{r}} rU'(W_0 + xr) f(r) \mathrm{d}r = 0 \tag{9.8}$$

如果存在一个满足上述条件的 $x<W_0$,则 U 的凹性保证了它是全局的最优值。

接下来的一步显然是要对这个选择进行比较静态分析。首先假设 W_0

变动，注意到 W_0 为 $\varnothing$ 的一个参数，因而，最大值为 $\varnothing(x,W_0)$。将它关于 x 的偏导数写为 $\varnothing_x$，关于 W_0 的偏导数记为 $\varnothing_w$，那么一阶条件就是 $\varnothing_x(x,W_0)=0$。像在第八章那样对它全微分，我们得到

$$\mathrm{d}x/\mathrm{d}W_0=-\varnothing_{xw}(x,W_0)/\varnothing_{xx}(x,W_0)$$

由二阶条件可知，上式分母为负，因而比较静态导数的符号与分子的符号相同。现在

$$\varnothing_{xw}(x,W_0)=\int_{\underline{r}}^{\bar{r}} rU''(W_0+xr)f(r)\mathrm{d}r \tag{9.9}$$

由于仅当$\underline{r}<0<\bar{r}$ 时，一个内点解 x 才可能是最优的，所以如果没有进一步的研究，我们尚不能确定(9.9)式的符号。这个问题的答案事实上取决于对风险规避的度量的性质。回想起 $U''<0$ 意味着风险规避；一个有用的数量度量意味着

$$A(W)=-U''(W)/U'(W) \tag{9.10}$$

要解释此含义，我们考虑略微降低 W 分布的方差，看看它的均值要降低多少才能使投资者无差异。这个边际替代率为$\frac{1}{2}A(W)$。一个较高的 $A(W)$ 表示，为使方差降低给定小的量，投资者愿意放弃更多的均值回报。因此 $A(W)$被称作投资者的绝对风险规避。我们可以预期较富有的投资者更能容忍给定的边际风险，即 $A(W)$应该是递减函数。

于是结果就是，如果绝对风险规避 $A(W)$是财富的递减函数，那么 $\varnothing_{xw}$ 为正；即富有的投资者将持有更多的风险资产。

要证明这一点，注意到 $r<0$，我们有

$$-U''(W_0+xr)/U'(W_0+xr)>-U''(W_0)/U'(W_0)=A(W_0)$$

上式乘以$-r$，得

$$rU''(W_0+xr)/U'(W_0+xr)>-rA(W_0)$$

即

$$rU''(W_0+xr)>-A(W_0)rU'(W_0+xr) \tag{9.11}$$

对 $r>0$，

$$-U''(W_0+xr)/U'(W_0+xr)<-U''(W_0)/U'(W_0)=A(W_0)$$

上式乘以 r，即

$$-rU''(W_0+xr)/U'(W_0+xr)<rA(W_0)$$

或再次得到

$$rU''(W_0+xr)>-A(W_0)rU'(W_0+xr)$$

我们已经证明不等式(9.11)式对所有的 r，不论正负，都成立。对它进行积分，

$$\int_{\underline{r}}^{\bar{r}} rU''(W_0+xr)f(r)\mathrm{d}r>-A(W_0)\int_{\underline{r}}^{\bar{r}} rU'(W_0+xr)f(r)\mathrm{d}r$$

由(9.8)式可知，不等式右边为零。这就证明了刚才的结论。

接下来的问题自然是密度函数的变动，特别是风险增加对 x 的影响；以及效用函数的变动，尤其是风险规避程度增加对 x 的影响。但这些论题会占去太多的篇幅，我将它们留给了本章末所给出的恰当的微观经济学教科书和研究论文。

9.3 投资组合选择

此处我们允许任何数目的风险资产。金融经济学中对此情形的处理通常是，假设投资者的目标函数可以表示成财富的均值 M 和标准差 S 的一个函数。这是一个限制性非常强的假设。在期望效用的框架中，它对应于两个特殊情形之一：(1)如果冯·诺伊曼—摩根斯坦效用函数是二次的，

$$U(W)=W-\frac{1}{2}aW^2$$

其中 $a>0$ 且是常数，那么

$$E[U(W)]=M-\frac{1}{2}a(M^2+S^2$$

(2)如果每一种资产都有正态分布的回报，那么财富也是正态分布的，并且任何冯·诺伊曼—摩根斯坦效用函数的期望都可以用它的均值和方差来表示。在这种情形下，值得关注的一个具体的函数是

$$U(W)=-\exp(-aW)$$

其中 $a>0$，其期望为

$$E[U(W)] = -\exp(-a[M - \frac{1}{2}aS^2])$$

因而当$(M-\frac{1}{2}aS^2)$达到最大值时期望效用也达到最大。利用(9.10)式，我们会发现，对这个函数而言绝对风险规避为常数并等于a。

在应用均值一方差分析的两种情形中的任何一种时，我们都可以在(M,S)空间中画出期望效用的无差异曲线。通过构建M和S之间的可行边界，我们可以用图形的方法来研究投资组合选择的问题。

在整个推导过程中初始财富都保持不变，因而我们将其正规化为1。假设存在n种资产，将它们的总收益（本金加利息）写成一个向量$r=(r_1,r_2,\cdots,r_n)$，且它为随机变数。令它们的数学期望组成一个向量$\mu=(\mu_1,\mu_2,\cdots,\mu_n)$，总收益的（对称正定）方差一协方差矩阵为$\sum=(\sigma_{ij})$。投资组合即为投资于各种不同资产的财富比例的向量，$x=(x_1,x_2,\cdots,x_n)$。

随机的最终财富为

$$W = \sum_{i=1}^{n} x_i r_i = x^T r$$

最终财富的均值和方差分别为

$$M = \sum_{i=1}^{n} x_i \mu_i = x^T \mu$$

和

$$S^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j \sigma_{ij} = x^T \sum x$$

注意M和S^2都是x的函数。

为找到M和S之间的可行转换边界，对给定的均值我们要最小化标准差，即

$$\min S = (x^T \sum x)^{1/2} \text{ s. t. } x^T\mu = M, x^T e = 1$$

其中e为由1组成的向量。

最小化的S是M的一个凸函数。它可能有递减的部分，但对于很大的M值它是递增的，因而显示出收益和风险之间的权衡。两种资产的情形足以展现问题的要点，因而这里我将注意力限制在此种情形。然而一般的情形可以用来展示前几章中的最优化方法的一些技巧的运用；我将在例题9.3

中简要说明。

在两种资产的情况下，令 x 代表 x_1，那么 $x_2=1-x$。因而

$$M = \mu_2 + (\mu_1 - \mu_2)x \tag{9.12}$$

且

$$S^2 = \sigma_2^2 - 2K_2 x + (K_1 + K_2)x^2 \tag{9.13}$$

其中我定义了

$$K_1 \equiv \sigma_1^2 - \rho\sigma_1\sigma_2, \quad K_2 \equiv \sigma_2^2 - \rho\sigma_1\sigma_2$$

ρ 是随机收益 r_1 和 r_2 之间的相关系数，即协方差除以它们标准差的积。

将资产排序使得 $\mu_1>\mu_2$。那么方程(9.12)式和(9.13)式通过参数 x 在 (M,S)空间中定义了边界。当 x 从 0 增加到 1，M 从 μ_2 增加到 μ_1，且 S 从 σ_2 变到 σ_1 时，沿此方向，有

$$S\,\mathrm{d}S/\mathrm{d}x = -K_2 + (K_1 + K_2)x \tag{9.14}$$

因此

$$\text{在 } x = 0 \text{ 处}, \mathrm{d}S/\mathrm{d}x = -(\sigma_2 - \rho\sigma_1)$$

及

$$\text{在 } x = 1 \text{ 处}, \mathrm{d}S/\mathrm{d}x = \sigma_1 - \rho\sigma_2$$

如果 $\sigma_1>\sigma_2$，那么在两个完全专门化的投资组合之间存在风险—收益的权衡，因而在 $x=1$ 附近 $\mathrm{d}S/\mathrm{d}x$ 肯定为正，即在收益高到一定范围后必然存在一个权衡。即使 $\sigma_1<\sigma_2$，即资产 1 完全优于资产 2，只要 ρ 足够小于 1，也可能通过混合两种资产而得到分散化的收益，从而也存在一个权衡。

更为一般地，方差最小的投资组合由下式给出：

$$x = \frac{K_2}{K_1 + K_2} = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{(\sigma_1^2 - \rho\sigma_1\sigma_2) + (\sigma_2^2 - \rho\sigma_1\sigma_2)}$$

从上面等式看，使 x 位于[0,1]间，即不含有任何卖空行为的条件是显然的。

将(9.14)式再次微分，

$$S\,\mathrm{d}^2S/\mathrm{d}x^2 + (\mathrm{d}S/\mathrm{d}x)^2 = C$$

化简上式，我们得到

$$\operatorname{sgn} d^2 S/dx^2 = \operatorname{sgn}(1-\rho^2)\sigma_1^2\sigma_2^2 > 0$$

因此，在(M,S)空间中该函数的边界为凸的。

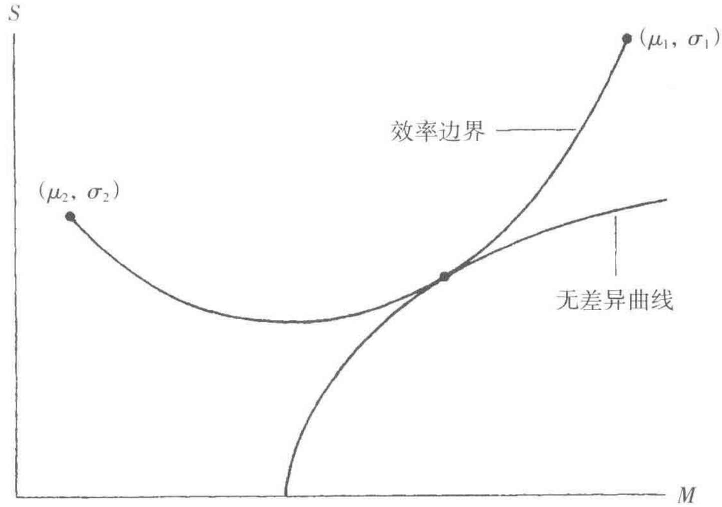

图 9.1　在均值—方差框架下的投资组合选择

图 9.1 汇总了到目前为止的所有结论。其中有两个极端的情形值得指出。如果 $\rho=1$，该边界就是连接(μ_1,σ_1)到(μ_2,σ_2)的直线。如果 $\rho=-1$，则最小方差为零，边界由分别连接此无风险投资组合点与(μ_1,σ_1)和(μ_2,σ_2)的两条直线组成。

下面假设存在一种有确定的总收益 μ_0 的无风险资产和如前述一样的多种风险资产。如果财富中的 x_0 比例用来持有无风险资产，则

$$M = x_0\mu_0 + x^T\mu = x_0\mu_0 + (1-x_0)\xi^T\mu$$

其中向量 ξ 给出了投资组合中有风险部分的各种不同资产的比例。同样地，

$$S^2 = x^T \sum x = (1-x_0)^2 \xi^T \sum \xi$$

或

$$S = (1-x_0)(\xi^T \sum \xi)^{1/2}$$

因此通过连接$(\mu_0,0)$和前面得到的边界上的每一个点可以找到现在的可行点。在图 9.2 中，如果不允许卖空，这将产生效率边界 ABP_1，如果允许卖空将产生直线边界 ABC。在此情形中，风险资产仅以点 B 所代表的比例被

持有。每个人都混合持有确定性资产和投资组合 B,这个混合的比例依赖于个人对风险的态度。因而 B 变成了一个对冲基金(a mutual fund)。

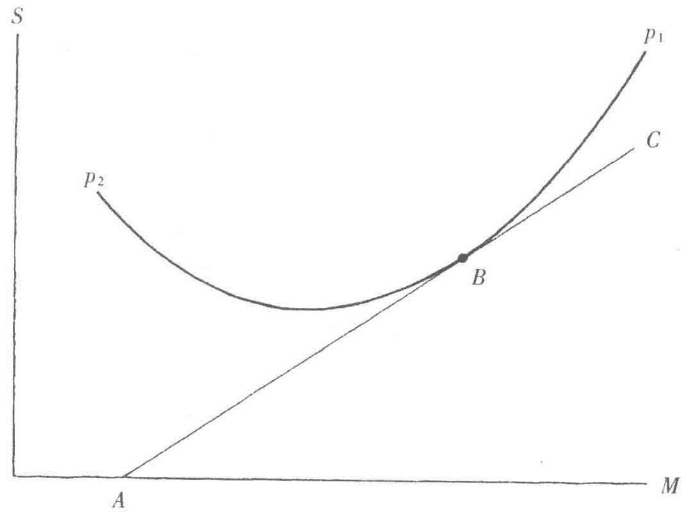

图 9.2　有一种无风险资产时的投资组合选择

例题

例题 9.1:管理者的激励

诱使下属提供一个恰当的努力水平对资本主义的企业所有者和社会主义的计划者是一个相同的问题。这里我们举一个简单的例子。企业的所有者必须雇用一个管理者来经营一个项目。如果项目成功了,它将产生价值 V。成功的概率依赖于管理者工作的质量。假定工作为高质量,项目成功的概率为 p,但工作低质量将使成功的概率降为 q。吸引一个管理者的基本薪水为 w。但为实现高质量他必须更加努力,并且只有当他得到奖金 e 时他才会这么做。所有者和管理者都是风险中性的,即最大化各自的货币收益的数学期望(在管理者的情形中减去与货币等价的努力成本)。

首先假定所有者可以观察到管理者工作的质量,并且能直接补偿管理者的努力。所有者诱使高质量的工作而带来的期望利润为 $pV-(w+e)$,而低质量的工作将只能给他带来 $qV-w$ 的利润。一种值得考虑的情形满足如下两个条件:(a)高质量工作比低质量工作产生更多的利润,即

$$pV-(w+e)>qV-w \text{ 或 } (p-q)V>e \tag{9.15}$$

(b)高质量工作的利润为正，即

$$pV>w+e \tag{9.16}$$

现在假设所有者不能观察到管理者努力的质量。如果所有者给管理者奖金 e，以获得管理者会提供高质量工作这一不可验证的保证，那么管理者可能会欺骗所有者而提供低质量工作。这样的行为将降低项目成功的概率，但是只要 $1>p>q>0$，所有者并不能通过观察单个项目的成败来推断管理者努力的质量，从而也无法惩罚管理者的违约行为。

因此，所有者必须将他的支付方案建立在他惟一能够观察到的事项之上，也就是成功或失败的基础之上。假定当项目成功时他支付给管理者 x，失败时支付 y。给定这个方案，管理者将选择高质量的努力，只要这将给他带来更多的扣除努力成本后的收益，即

$$px+(1-p)y-e>qx+(1-q)y$$

如果这个不等式的两边相等，则管理者付出高质量的努力和低质量的努力是无差异的。通常假定管理者是一个好人，当这两者对他而言无差异时，他将站在所有者的这一边，也就是说他会选择付出高质量的努力。因而不等式是弱的而不是严格的。它可化简为

$$(p-q)(x-y)\geqslant e \tag{9.17}$$

其次，管理者将同意为所有者工作，如果满足

$$px+(1-p)y\geqslant w+e$$

或

$$y+p(x-y)\geqslant w+e \tag{9.18}$$

(9.17)和(9.18)这两个不等式就是给管理者提供恰当激励的约束。

所有者的期望利润为

$$\pi=pV-[px+(1-p)y]=pV-y-p(x-y) \tag{9.19}$$

在满足上面两个激励约束的情况下他要使期望利润最大。显然，他应使 y 和 $(x-y)$ 尽可能小。于是约束以等式形式成立，并有

$$x-y=e/(p-q),\quad y=w+e-ep/(p-q)$$

或

$$y = w - eq/(p-q), \quad x = w + e(1-q)/(p-q) \qquad (9.20)$$

这两个等式有一个简单的解释：管理者的补偿计划包括基本的薪水加上对成功的奖金或者减去对失败的罚金。

有了这些值之后，所有者的期望利润为

$$\pi = pV - w - e$$

与他可以直接观察到管理者的努力水平时的期望利润相同；因此，不能观察努力水平并未产生差别。

然而，有一个困难，即没有什么条件可以保证 $y \geqslant 0$。因而如果项目失败了，最优补偿方案可能包含了管理者支付给所有者的一个罚金。但在现实中这样的罚金是很难实施的。如果排除罚金，这个问题必须增加一个额外的约束 $y \geqslant 0$ 再进行求解。解将尽可能地靠近边界，即令 $y=0$。那么 x 必须满足余下的两个约束，所以

$$x \geqslant (w+e)/p, \quad x \geqslant e/(p-q)$$

但当 y 在无约束解中为负时，$w+e$ 小于 $ep/(p-q)$，而后者为紧的约束。因此 $x=e/(p-q)$，且所有者的期望利润为

$$\pi = pV - ep/(p-q)$$

由开头假设的条件(9.15)式和(9.16)式，这个值小于 $pV-(w+e)$，但仍然为正。这样，基于产出的激励机制所要求的支付水平就意味着较小的期望利润，但是这一支付水平并不会多到让整个企业无利可图的程度。

这个例子只是近年来关于激励合约设计的大量研究中的一个起点。这里考虑的问题同保险业中道德风险的问题相似，保险公司也不能观察到投保方为降低风险采取的谨慎行为。我们为单个项目设计了最优合约，但如果这种关系一直进行下去，管理者在一段时间中的平均成功业绩就能提供有关他的努力情况的统计信息，从而使那些能够带来更高期望利润的更好的合约得以实现。类似地，如果计划者有几个面临相同风险的类似的管理者，那么通过观察他们的相对表现就可以获得他们相对努力的信息。最后，在这个例子中存在清偿约束，即管理者不能被处以罚金，但这并不是风险规避。允许管理者风险规避会带来如何在计划者和管理者之间有效地分配风

险这一额外的问题。本章末提供了一些新兴的关于这类问题的参考文献。

下面的例子将关注信息的一个不同的问题。计划者可能不知道管理者固有的禀赋。这在保险业中也有相类似的情形，投保者比保险公司拥有更多的关于他自己的风险等级的信息。这就导致了“逆向选择”；一份保险合约特别会吸引那些知道自己的风险高于保单隐含的风险的客户，而这些客户对保险公司来说是不受欢迎的。所以保险公司要设计一个能够处理这一问题的合约。我将在一个不同的背景中构建此例以求变化。

例题 9.2：成本加成合约

政府在国防、健康甚至教育方面的支出通常以成本加成为基础。也就是说，政府补偿供货商的成本并加上一个正常的利润。而且，政府采购部门通常不知道生产这些物品或服务的真实成本。如果一个能以低成本生产的供货商假装有一个较高的生产成本，那他就会获得较高的补偿。这些补偿可用以支付给他自己和其他工人很高的薪水，并可供在办公室及类似设施方面挥霍。如果政府希望避免过多的支付，它就必须设计出一个能消除谎报成本企图的方案。下面是一个非常简单的例子。

假定真实的平均生产成本只能取 c_1 和 c_2 两个值中的一个，且 $c_1<c_2$，每个值都已经包含了正常的利润。问题是高的值可能是真实的也可能是伪装的，而政府无法进行区分。换言之，供货商可能为两种类型中的任何一种，低成本或者高成本，而采购部门不能断定它面对的是一个真实的高成本供货商，还是一个假装成 c_2 类型并谋划着享受额外支付的 c_1 类型供货商。

政府能够决定它将购买多少单位，将支付多少，而这些都取决于供货商宣称的成本。假定它宣布：如果供货商声称具有平均成本 c_i，$i=1,2$，它将购买 q_i 个单位，并共为它们支付 R_i。

什么是这些选择的约束呢？首先，对于两个成本值中的任何一个而言，如果它正好是真实的，供货商应该愿意提供产品或服务。也就是说，他的成本应该被抵消：

$$R_i \geqslant c_i q_i, \quad i=1,2 \tag{9.21}$$

第二，对供货商来说，披露他的真实成本应该是最优的。这两点被称为激励相容约束。如果供货商的真实平均成本为 c_1，他应该不希望伪装成他的成

本为 c_2,反过来也是如此。经济直觉表明只有前一种情况会是紧的(一个真正高成本的企业将不会伪装成低成本的),但应该有一个恰当的理论分析能证明这一点,而不是在一开头就假设它成立。现在,如果成本为 c_1 的企业伪装成有成本 c_2,它将销售 q_2 个单位而不是 q_1,获得收入 R_2 而不是 R_1,但在它计算利润时它的实际成本仍然为 c_1。因此,对成本为 c_1 的企业来说,真实报告其成本的约束为

$$R_1 - c_1 q_1 \geqslant R_2 - c_2 q_2 \tag{9.22}$$

类似地,对真实成本为 c_2 的企业而言,其约束为

$$R_2 - c_2 q_2 \geqslant R_1 - c_2 q_1 \tag{9.23}$$

假定政府从数量 q 中获得收益 $B(q)$,其中 B 为一个递增且严格凹的函数。假定它对真实成本为 c_1 的概率的估计为 θ_1,对真实成本为 c_2 的概率的估计为 $\theta_2 = 1 - \theta_1$,那么在扣除给供货商的支付之后它的期望利润为

$$\theta_1[B(q_1) - R_1] + \theta_2[B(q_2) - R_2] \tag{9.24}$$

最优方案在满足参与约束和激励相容约束(9.21—9.23 式)下将最大化这个值。而且现在我将忽略对 q_i 和 R_i 的非负约束。

构建拉格朗日函数

$$\begin{aligned} L = {} & \theta_1[B(q_1) - R_1] + \theta_2[B(q_2) - R_2] + \mu_1[R_1 - c_1 q_1] + \mu_2[R_2 - c_2 q_2] \\ & + \lambda_1[R_1 - R_2 - c_1 q_1 + c_1 q_2] + \lambda_2[R_2 - R_1 - c_2 q_2 + c_2 q_1] \end{aligned}$$

不用求解整个问题,就可以得到大多数在经济学意义上有趣的结论。首先我们将激励相容约束相加在一起,化简而写成

$$(c_2 - c_1)(q_1 - q_2) \geqslant 0 \tag{9.25}$$

因此如果供货商宣称低成本,他将至少和高成本类型销售一样多,而且一般而言要销售得更多。这是使低成本类型企业作出真实反应的一部分激励。事实上解可以是 $q_2 = 0$ 和 $q_1 > 0$。那将有效消除供货商伪装成拥有高成本的激励。但这也带来了风险:如果供货商被证明是真地拥有高成本,那么已经承诺执行这个方案的政府将不能从他那里购买物品,尽管这样的交易也许是事后值得的。因此,只有当这样的风险或者其后果足够小,即当 θ_2 很小,或者 $B'(0)$ 有限且很小时,这样的解才会出现。我将把细节留给读者,并在此后假定 q_2 和 q_1 都是正的。

现在考虑参与约束(9.21)式,并旨在证明两个约束不可能同时为松。我们注意到,如果类型2企业获得了正的利润,则类型1企业也将获得正的利润,以此开始我们的分析。为证明这一点,假定 $R_2-c_2q_2>0$,则从类型1企业的激励相容约束(9.22)式我们得到

$$R_1-c_1q_1>(c_2-c_1)q_2>0$$

其次注意到对 R_1 和 R_2 的一阶条件:

$$-\theta_1+\lambda_1-\lambda_2+\mu_1=0$$
$$-\theta_2-\lambda_1+\lambda_2+\mu_2=0$$

如果两种类型的企业都获得了正的利润,两个参与约束都是松的,且互补松弛意味着 $\mu_1=0=\mu_2$,那么

$$\theta_1=\lambda_1-\lambda_2=-\theta_2$$

但 θ_1 和 θ_2 都是正的,因此上式是不可能的。

因而最优方案必须有

$$R_2-c_2q_2=0,\ R_1-c_1q_1>0 \tag{9.26}$$

正的纯利润是类型1企业真实披露其低成本的另外一部分激励。它的平均收入 R_1/q_1 超过它的平均成本 c_1,但根据类型2企业的激励相容约束,它不可能超过 c_2。

当然,在满足类型1企业的激励相容约束(9.22)式下政府将使 R_1 尽可能小。在此约束中替代 R_2,我们得到

$$R_1=c_1q_1+(c_2-c_1)q_2 \tag{9.27}$$

由此,容易验证(9.23)式自动满足,并且只要 $q_1>q_2$ 它就是严格的不等式,而在最优解处的确有 $q_1>q_2$。这样我们已经证明了前文中那个出于直觉的结论,即高成本企业不会想低报它的成本,但真正的低成本企业位于想高报它成本的边际上。

现在政府的目标函数可以写成

$$\theta_1[B(q_1)-c_1q_1-(c_2-c_1)q_2]+\theta_2[B(q_2)-c_2q_2]$$

关于 q_1 和$_2$ 的一阶条件为

$$B'(q_1)=c_1 \tag{9.28}$$

和

$$B'(q_2) = c_2 + (\theta_1/\theta_2)(c_2 - c_1) \tag{9.29}$$

这两个等式也在激励方案中起作用，其中，(9.28)式是直观的。如果供应商宣称自己是低成本类型的，则政府从他那里购买了使其边际收益等于边际成本(等于平均成本)的数量的商品或服务。但(9.29)式更复杂。如果供应商宣称自己是高成本类型的，从政府角度看这也将是购买较少物品的一个理由；且边际收益等于边际成本意味着 $B'(q_2)=c_2$，因而 $q_2<q_1$。事实上等式(9.29)式的右边大于 c_2，因此政府从高成本类型企业中购买得更少。这再一次降低了低成本类型企业高报成本的诱惑。当供应商确实为类型 2 时，这样的政策的确意味着放弃了一部分社会净边际收益，但付出这样的成本以获得激励方面的补偿收益是值得的。表达这一观点的另一种方式是：注意到成本较高的可能性为低成本企业提供了伪装成高成本企业的诱惑，这就像由高成本企业引起的外部不经济。等式(9.29)式的右边在 c_2 上又增加了这种外部性的影子价格以得到类型 2 企业的真实的社会成本。因此从该企业购得的数量 q_2 使边际收益等于此真实的社会成本。

例题 9.3：均值—方差边界

我将在此题中分析有 n 种资产的投资组合选择中转换边界的凸性。这填补了被正文省略掉的技术细节，并且附带地提供了使用凸性和最大值函数的一个很好的例证。论证的推导用到了两个中间结果或引理。

引理 1：$\varnothing(x)=(x^T \sum x)^{\frac{1}{2}}$ 是 x 的凸函数。

证明：令 $\varnothing^2=x^T \sum x$，则两边微分得，

$$2\varnothing\varnothing_x = 2\sum x$$

且

$$\varnothing\varnothing_{xx} + \varnothing_x\varnothing_x^T = \sum$$

因此

$$\varnothing\varnothing_{xx} = \sum - \frac{\sum xx^T \sum}{\varnothing^2}$$

我们证明矩阵 $\varnothing_{xx}$ 是正半定的。对任意的向量 z，

$$\emptyset z^T \emptyset_{xx} z = z^T \sum z - z^T \sum x x^T \sum z / \emptyset^2 = z^T \sum z - (z^T \sum x)^2 / (x^T \sum x)$$

由柯西一施瓦茨(Cauchy-Schwartz)不等式可得,这个表达式为非负。

引理 2: 如果 $\emptyset(x)$ 是凸的,且 $S=\min\{\emptyset(x) \mid \mu^T x = M, e^T = 1\}$,则 S 是 M 的凸函数。

证明: 令 M_1、M_2 为 M 的任意两个值,且令 x^1、x^2 为相应的最小值。那么对任意的 $\theta \in [0,1]$,

$$\mu^T(\theta x^1 + (1-\theta)x^2) = \theta M_1 + (1-\theta)M_2$$
$$e^T(\theta x^1 + (1-\theta)x^2) = 1$$

因而平均投资组合是可行的。因此

$$\begin{aligned} S(\theta M_1 + (1-\theta)M_2 &\leqslant \emptyset(\theta x^1 + (1-\theta)x^2 \\ &\leqslant \theta\emptyset(x^1) + (1-\theta)\emptyset(x^2) \\ &= \theta S(M_1) - (1-\theta)S(M_2) \end{aligned}$$

这就证明了想要得到的凸性。现在我简要地说明怎样得到边界 $S(M)$ 的显性解。这会使 S^2 最小更容易。令 α、β 为相应约束的乘子。一阶条件为

$$\sum x = \alpha\mu + \beta e$$

求解出 x 并代入约束中,

$$\alpha(\mu' \sum{}^{-1} \mu) + \beta(e' \sum{}^{-1} \mu) = M$$
$$\alpha(\mu' \sum{}^{-1} e) + \beta(e' \sum{}^{-1} e) = 1$$

这两个等式可以求解出 α 和 β,从而给出用 M 表示的最优投资组合 x 和最小化的 S 的表达式。

习题

习题 9.1:风险收入的征税

假定在正文中一种无风险资产和一种风险资产的模型中,我们引入对

风险资产的利息收入课以税率为 τ 的税收(若为损失则减少了损失),其中 $0<\tau<1$。因而最终的随机收入为

$$W = W_0 + (1+\tau)xr$$

证明内点最优解 x 的一阶条件为

$$\int_{\underline{r}}^{\bar{r}} rU'(W_0 + (1-\tau)xr) f(r)\mathrm{d}r = 0$$

推导:如果 τ 变动,最优的 x 也将变动,以保持 $(1-\tau)x$ 不变。因此如果风险收入的税率增加,则以风险资产持有的财富的数量也将增加。指出这一看上去矛盾的结论的经济学直觉。

例题 9.2:有不确定性的储蓄

一个消费者生活两期。在第一期的收入是确定的,等于 Y_1。第二期的收入 Y_2 可以是随机的。如果他从第一期收入中储蓄 S,他在第二期获得总收益(本金加利息)rS,其中 r 可以是随机的。他的目标是最大化两期内消费的期望效用现值:

$$U(Y_1 - S) + \delta E[U(Y_2 + rS)]$$

其中 $U'>0$ 和 $U''<0$。写出一阶和二阶条件。证明当 Y_1 增加时,S 也增加但以一个更小的比例增加,即边际储蓄倾向在 0 和 1 之间。

下面假定 Y_2 是确定的但 r 为随机的,分析 Y_2 增加的影响。最后,假定 r 为确定的,但 Y_2 是随机的,分析 r 上升的影响。

▶ 进一步阅读

不确定性经济学的一个基本阐述可以看 Varian, *Microeconomic Analysis*,第 3 章以及第 18—20 章的部分小节。更完整的分析,包括期望效用最大化的有效性的讨论,参见 David M. Kreps, *Notes on the Theory of Choice*, Boulder, CO: Westview Press, 1988。

经典的而且仍然经常被推荐的有关期望效用理论的分析和应用,可参

见 Kenneth J. Arrow, *Essays in the Theory of Risk-Bearing*, Amsterdam: North-Holland, 1971, 第 1—3 章。

下面的论文综述了关于比期望效用更一般的准则函数的近期研究: Mark Machina, " Choice under Uncertainty: Problems Solved and Unsolved", *Journal of Economic Perspectives*, 1(1), 1987, pp. 121—54。

有关风险和风险规避上升的比较静态效应的一般理论可以在下面二文中找到:Michael Rothschild and Joseph Stiglitz, "Increasing Risk: (1) A Definition, and (2) Economic Consequences", *Journal of Economic Theory*, 2,1970,pp. 225—43 and 3, 1971, pp. 66—84;Peter Diamond and Joseph Stiglitz, "Increases in Risk and in Risk Aversion", *Journal of Economic Theory*, 8, 1974, pp. 337—60。

投资组合选择的理论在每一本金融经济学的书中都有阐述。关于初级水平的理论,有兴趣的读者可以参考 Richard A. Brealey and Stewart Myers, *Principles of Corporate Finance*, 2nd edn., New York: McGraw-Hill, 1984。

更高级的分析参见 Chi-Fu Huang and Robert H. Litzenberger, *Foundations of Financial Economics*, Amsterdam: North-Holland, 1988。

Varian (1984,参见第 3 章的进一步阅读,第 8 章)给出了信息不对称的简要分析。更详细的阐述见 David Kreps, *A Course in Microeconomic Theory*, 前面所引,第 16—18 章。

▶10 时间:最大值原理

10.1 问题的论述

同在不确定性的情形下一样,时间维度上的最优化问题并不需要新的一般化的原理。虽然我们所要选择的变量是关于不同的时期的,但总是可以用一个大的向量 x 把它们归在一起,所以一般的问题仍然是在一个向量不等式 $G(x) \leqslant c$ 的约束下最大化函数 $F(x)$ 的值。当我们在进行决策时,对未来的偏好和技术的了解可能非常不清楚。但是这只要求我们能够在函数 F 和 G 中刻画出不确定性和对风险的态度。随着时间的推进,我们可能有机会重新考虑当前的决定并修改计划。但是,这也仅要求我们在当前的选择中能意识到未来可能有这样的修改。这种考虑会使得我们现在采取更为灵活的决策,从而允许今后有更好的了解时进行选择。但是它同时也意味着我们现在要作出承诺,以排除某些将来的路径,因为未来的偏好会诱使我们选择这些路径,而今天的偏好则不会。此处,今天的决策涉及一场与自己的未来决策之间的博弈。当所有的这些考虑被包含在目标函数和约束函数中时,我们就可以继续应用前面那些章节中的正式的理论。

把对时间维度上的最大化问题的研究作为一个单独的主题,并非因为它要求基本的新的理论。相反,这是因为此类问题通常有一个特别的结构,从而能使我们对它们的解更加了解。这一特别结构中最重要的一点,就是在连续的时间点上各变量之间存在存量—流量关系。其中有一些变量具有存量维度,我在后文将之标以 y,并有适当的时间下标或自变量。其他的变量具有流量维度,标以 z。用数学术语来说,这些存量变量被称为状态变量,

而流量变量则被称为控制变量。应用在通常的生产问题中就是，这一期的经济活动会决定从这一期到下一期的存量的变动。存量的可行的增加同时取决于原来的存量和这一期的流量。因而，生产可能性约束为

$$y_{t+1} - y_t = Q(y_t, z_t, t) \tag{10.1}$$

这里 t 和 $(t+1)$ 是连续的离散时期，y 代表了资本品的存量，z 可以包括劳动供给、消费流量等等，Q 应该被视作生产函数，而时间 t 作为一个自变量明确地出现在这个函数中，以刻画外生的技术变动。从数学上说，控制变量控制了状态变量的变动。为了和通常的允许扔掉物品的想法一致，我应该把(10.1)式写作一个不等式

$$y_t + Q(y_t, z_t, t) \geqslant y_{t+1}$$

但是事实上，这个约束在最优路径上总是紧的，所以我在以后会采用更加简洁的等式形式。

除了决定存量变化的约束条件之外，还有一些在任何一个时点上关于所有变量的约束，例如

$$G(y_t, z_t, t) \leqslant 0 \tag{10.2}$$

其中 G 为向量函数。例如，限制消费不能超过总产出的约束。所有的存量和流量变量为非负的约束也可以被包括在(10.2)式中。

经常出现在时间维度上的最优化问题的另一个特点就是，它的准则函数是加性可分的：它可以被表达成一系列函数的和，而每一个函数依赖于只与一期有关的变量：

$$\sum_{t=0}^{T} F(y_t, z_t, t) \tag{10.3}$$

例如，一个最大化其利润流的折现值的企业自然就会有这样的目标，时间会以折现因子 $(1+r)^{-t}$ 的形式显性地进入这个函数，其中 r 为利率。对于消费者时间维度上的选择（如储蓄决策），假定效用函数在时间维度上是加性可分的，通常处理起来非常方便。这是对偏好的一个限制。粗略地说，它要求“午餐和晚餐之间的边际替代率与早餐的数量无关”。这个简洁的例子来自于 Henry Wan。与不确定性的情形下效用函数的形式一样，出于分析的可行性的考虑，我们通常假定效用函数在时间维度上是加性可分的。

还有一个细节需要具体说明。最优化问题的时间跨度从 $t=0$ 开始。

最初的存量或状态变量 y_0 必定是某些未详细说明的历史的结果；我们仅仅将它们视为给定的。类似地，当最优化在有限的 T 时期中止时，我们必须说明某些终端条件，而最简单的做法就是要求将固定不变的存量向量 y_{T+1} 遗赠给未来。

10.2 最大值原理

现在选择变量为 $y_t(t=1,2,\cdots,T)$ 和 $z_t(t=0,1,\cdots,T)$。它们都必须满足约束(10.1)式和(10.2)式，每一个约束条件对 $t=0,1,\cdots,T$ 都必须成立。目标函数由(10.3)式给出。

和以往一样，我们可以定义影子价格并建立拉格朗日函数。令 λ_t 代表约束(10.2)式的拉格朗日乘子，它们具有通常的关于 t 时刻的约束的影子价格的解释。而(10.1) 式的影子价格则是新的，且更加有趣。如果关于存量增加的约束放松一点，或者说，如果我们被给予存量 y_{t+1} 的一个小的增量，它们就表示，目标函数的一阶增量。因而它们是$(t+1)$时刻存量的影子价格，我把它们记作 π_{t+1}。

记 $\mathcal{L}$ 为整个跨期问题的拉格朗日函数，那么

$$\mathcal{L}=\sum_{t=0}^{T}\{F(y_t,z_t,t)+\pi_{t+1}[y_t+Q(y_t,z_t,t)-y_{t+1}]-\lambda_t G(y_t,z_t,t)\} \quad (10.4)$$

$\mathcal{L}$ 的变量为所有的 y_t、z_t、λ_t 和 π_{t+1}。它们数量太多而没有罗列在等式的左边。本章中我将使用在第 3 章中建立的一阶必要条件，而不再明显地将最优解和其他可行的选择进行比较。因而我不再需要用短横将特定的点从一般的点中区别出来，略去它们会使表示法大大简化。我同时还假定适当的约束规格总是满足的。

关于 $z_t(t=0,1,\cdots,T)$的一阶条件非常简单：

$$\partial L/\partial z_t\equiv F_z(y_t,z_t,t)+\pi_{t+1}Q_z(y_t,z_t,t)-\lambda_t G_z(y_t,z_t,t)=0 \quad (10.5)$$

而那些关于 y_t 的一阶条件则更加复杂，因为每一个 y_t 都出现在和式中的两项中。例如，y_1 出现在 F、Q、G 中，以 $\pi_2 y_1$ 的形式出现在 $t=1$ 的项中，又以 $-\pi_1 y_1$ 的形式出现$t=0$的项中。我们可以重新整理表达式以使得每一个 y_t

只出现在一项中。只取(10.4)式的相关部分：

$$\begin{aligned}
&\sum_{t=0}^{T}\pi_{t+1}(y_t-y_{t+1})\\
&\quad=\pi_1(y_0-y_1)+\pi_2(y_1-y_2)+\cdots+\pi_{T+1}(y_T-y_{T+1})\\
&\quad=y_0\pi_1+y_1(\pi_2-\pi_1)+\cdots+y_T(\pi_{T+1}-\pi_T)-y_{T+1}\pi_{T+1})\\
&\quad=\sum_{t=0}^{T}y_t(\pi_{t+1}-\pi_t)+y_0\pi_1-y_{T+1}\pi_{T+1} \qquad (10.6)
\end{aligned}$$

于是(10.4)式就变成

$$\begin{aligned}
\mathcal{L}=&\sum_{t=0}^{T}\{F(y_t,z_t,t)+\pi_{t+1}Q(y_t,z_t,t)+y_t(\pi_{t+1}-\pi_t)-\lambda_t G(y_t,z_t,t)\}\\
&+F(y_0,z_0,0)+\pi_1 Q(y_0,z_0,0)+y_0\pi_1-y_{T+1}\pi_{T+1} \qquad (10.7)
\end{aligned}$$

最后一行中被放在求和符号外的项只同 y_0 和 y_{T+1} 有关，而这两个不是选择变量。因此关于 $y_t(t=1,2,\cdots,T)$ 的一阶条件为

$$\partial L/\partial y_t\equiv F_y(y_t,z_t,t)+\pi_{t+1}Q_y(y_t,z_t,t)+\pi_{t+1}-\pi_t-\lambda_t G_y(y_t,z_t,t)=0$$

或

$$\pi_{t+1}-\pi_t=-[F_y(y_t,z_t,t)+\pi_{t+1}Q_y(y_t,z_t,t)-\lambda_t G_y(y_t,z_t,t)] \qquad (10.8)$$

这些条件可以写成一个更加紧凑并且在经济学含义上更具启发性的形式。定义一个新的函数 H，称为汉密尔顿函数：

$$H(y,z,\pi,t)=F(y,z,t)+\pi Q(y,z,t) \qquad (10.9)$$

那么(10.5)式表明我们应该在时刻 t 选择控制变量 z_t，使其在约束 $G(y_t,z_t,t)\leqslant 0$ 下最大化 $H(y_t,z_t,\pi_{t+1},t)$。将由此得到的最大值记为 $H^*(y_t,\pi_{t+1},t)$。

定义这个单期最大化问题的拉格朗日函数为 L(以区别于在所有时期上求最大化问题的 $\mathcal{L}$)：

$$L=H(y_t,z_t,\pi_{t+1},t)-\lambda_t G(y_t,z_t,t) \qquad (10.10)$$

于是(10.8)式可以被更简单地写为

$$\pi_{t+1}-\pi_t=-L_y(y_t,z_t,\pi_{t+1},t)$$

在静态的最大化问题中，只有 z_t 是选择变量，y_t 和 π_{t+1} 都是参数。因而适用包络定理，我们有

$$\pi_{t+1}-\pi_t=-H_y^*(y_t,\pi_{t+1},t) \tag{10.11}$$

包络定理还给出了 $H_\pi^*=L_\pi=Q$，并在最优解处取值。因而我们可以把(10.1)式写成和(10.9)式对称的形式

$$y_{t+1}-y_t=H_\pi^*(y_t,\pi_{t+1},t) \tag{10.12}$$

这些结果可以被总结如下：

最大值原理：在满足约束(10.1)式和(10.2)式下最大化(10.3)式的一阶必要条件是：

(i)对每一期 t，z_t 在单期约束 $G(y_t,z_t,t)\leqslant 0$ 下最大化汉密尔顿函数 $H(y_t,z_t,\pi_{t+1},t)$。

(ii)差分方程(10.11)式和(10.12)式决定了 y_t 和 π_t 在时间维度上的变动。

这一原理在求解具体应用中的类似问题时非常有用。不过它最大的概念意义上的优点，则是最大化条件(i)的经济学解释。显然，我们不会选择 z_t 以最大化 $F(y_t,z_t,t)$：因为我们知道 z_t 的选择会通过(10.1)式而影响到 y_{t+1}，从而影响到在 $t+1$ 等时期目标函数中的各个项。以生产问题为例来解释，今天大肆挥霍的消费虽然会增加今天的效用，但这将意味着未来只有较少的资本积累，从而在未来只有较少的消费和较低的效用。我们怎样才能用一种简洁的方式刻画所有这些对未来的影响呢？当然是通过利用受影响的存量的影子价格。z_t 对 y_{t+1} 的影响等于它对 $Q(y_t,z_t,t)$ 的影响，由此产生的目标函数值的变动等于这一影响乘以 y_{t+1} 的影子价格 π_{t+1}。而这正是我们加进 F 中以得到汉密尔顿函数的那一项。这样，汉密尔顿函数提供了一种非常简洁的方式，修正了一期的目标函数 $F(y_t,z_t,t)$，使之考虑了在 t 时期控制变量 z_t 的选择对未来的影响。

条件(10.8)式或等价的(10.11)式也有一个有价值的经济学解释。适当注意单期约束的影子成本，y_t 的一个边际单位在时期 t 中产生的边际回报为 $F_y(t)-\lambda_t G_y(t)$，并在下一期还产生额外的 $Q_y(t)$，以 π_{t+1} 计值。（注意，我使用了自变量 t 而不是整个的 (y_t,z_t,t) 以求简洁。）这些都可以被视作一种红利。价格变动 $\pi_{t+1}-\pi_t$ 就像是一种资本利得，只是价格是用现值

的形式来表示的，这样 π_{t+1} 就包含了一个额外的折现因子，它刻画了一期中使用 y_t 的通常的利息或机会成本。当 y_t 为最优解时，全部的边际回报或这些分量的和，应该为零。当写作如下的形式时，这就是(10.8)式表达的含义：

$$[F_y(t)-\lambda_t G_y(t)]+\pi(t)Q_y(t)+[\pi_{t+1}-\pi_t]=0 \qquad (10.13)$$

换句话说，影子价格的取值应该使持有资产不会产生净收益或者超额收益；这是一个跨期的无套利条件。

如果没有施加对存量 y_{T+1} 的终端条件，情况又会如何呢？由于这些存量对于目标函数没有任何贡献，最优的策略应该是让它们的值尽量低，或通常为零。但是在某些情况下，先积累一定的存量以提供产出和效用，然后以保证它们在终端时刻恰好被用完的速度折旧，这样的做法或许是合意的。如果最后还有正的存量余留，那它们必定是无价值的，换句话说，我们应该有

$$y_{T+1}\geqslant 0,\quad \pi_{T+1}\geqslant 0\text{，满足互补松弛条件}$$

更加一般地，如果有一个约束是 $y_{T+1}\geqslant\hat{y}$，我们有

$$y_{T+1}\geqslant\hat{y},\quad \pi_{T+1}\geqslant 0\text{，满足互补松弛条件} \qquad (10.14)$$

这种关于终端存量和它们的影子价格的条件通常被称为横截条件。

10.3　连续时间模型

到目前为止，我一直把时间处理为一个个离散时期的延续。这使得本章前面讨论的理论可以作为标准的拉格朗日一库恩一塔克定理的一个特例进行展开，并得出有关这些条件的容易理解的经济学解释。但是在求解现实中的具体问题时，把时间视为一个连续的变量会方便得多。但不存在理论上的理由让我们偏好离散时间或者连续时间。为了方便记忆，我把离散时间记作一个下标，把连续时间记作函数括号中的变量。

我们可以把连续时间视作离散时期的长度 Δt 趋近于零的极限。这要求对(10.1—10.3 式)略作修改。现在，流量就是每单位时间的比率，所以表示存量一流量关系的(10.1)式的右边必须乘上时期长度 Δt，这个等式就变成

$$y(t+\Delta t)-y(t)=Q(y(t),z(t),t)\Delta t$$

两边同除以 Δt 并令其趋于零，这就给出了等式左边的存量对时间的导数。按惯例，我们可以在这个变量的上方加一个点以表示它对时间的导数，这样，我们就有

$$\dot{y}(t)=Q(y(t),z(t),t) \tag{10.15}$$

只要把(10.2)式中的下标改成自变量就可以得到

$$G(y(t),z(t),t)\leqslant 0 \tag{10.16}$$

(10.3)式中的求和就比较复杂。从 0 到 T 的时间总跨度被分成 $T/\Delta t$ 个小的离散的区间，用 i 来表示这些区间，那么这个和式就可以被写作

$$\sum_{i=0}^{T/\Delta t}F(y(i\,\Delta t),z(i\,\Delta t),i\,\Delta t)\Delta t$$

当 Δt 趋于零时，这个和式的极限就是积分

$$\int_0^T F(y(t),z(t),t)\mathrm{d}t \tag{10.17}$$

与此伴随的优点是 $\pi_{t+\Delta t}$ 会收敛于 π_t，从而在时刻 t 的存量不会和汉密尔顿函数中时刻 $t+1$ 的影子价格混淆在一起。和(10.9)式一样地定义 H，即 z_t 在约束 $G(y(t),z(t),t)\leqslant 0$ 下最大化 $H(y(t),z(t),\pi(t),t)$。记 H^* 为最大值函数，那么 $y(t)$ 和 $\pi(t)$ 满足下面这对微分方程：

$$\dot{y}(t)=H_\pi^*(y(t),\pi(t),t) \tag{10.18}$$

和

$$\dot{y}(t)=-H_y^*(y(t),\pi(t),t) \tag{10.19}$$

通过像(10.4)式那样先定义整个问题的拉格朗日函数 $\mathcal{L}$，我们可以正式地推导这两个条件：

$$\begin{aligned}\mathcal{L}=\int_0^T\{&F(y(t),z(t),t)+\pi(t)[Q(y(t),z(t),t)-\dot{y}(t)]\\&-\lambda(t)G(y(t),z(t),t)\}\mathrm{d}t\end{aligned} \tag{10.20}$$

与(10.6)式中的重新整理类似的是分部积分：

$$-\int_0^T\pi(t)\dot{y}(t)\mathrm{d}t=\int_0^T y(t)\dot{\pi}(t)\mathrm{d}t+y(0)\pi(0)-y(T)\pi(T) \tag{10.21}$$

所以

$$\mathcal{L}=\int_0^T\{F(y(t),z(t),t)+\pi(t)Q(y(t),z(t),t)+y(t)\dot{\pi}(t)$$

$$-\lambda(t)G(y(t),z(t),t)\}\mathrm{d}t+\pi(0)y(0)-\pi(T)y(T) \tag{10.22}$$

现在我们可以把积分看作求和，分别对 $z(t)$ 和 $y(t)$ 求导以得到一阶条件

$$F_z(y(t),z(t),t)+\pi(t)Q_z(y(t),z(t),t)-\lambda(t)G_z(y(t),z(t),t)=0 \tag{10.23}$$

和

$$F_y(y(t),z(t),t)+\pi(t)Q_y(y(t),z(t),t)+\dot{\pi}(t)-\lambda(t)G_y(y(t),z(t),t)=0 \tag{10.24}$$

(10.23)式是 $z(t)$ 最大化汉密尔顿函数的条件，而(10.24)式则类似于(10.13)式，这一跨期无套利方程

$$F_y(t)-\lambda(t)G_y(t)+\pi(t)Q_y(t)+\dot{\pi}(t)=0 \tag{10.25}$$

当然，事实并非这样简单。通常意义上说，积分不能对一瞬间的变量求导。而在连续时间下求最优化的严格的理论则非常复杂。但是像上面那样的捷径确实给我们带来了有用的结论。对绝大多数读者来说这已经足够了。那些想知道更多的读者可以看本章末所引述的参考文献。

最大值原理的实际应用为推导微分方程(10.18—10.19 式)，并在满足 $t=0$、T 的适当条件下求解这两个方程。如果时间没有明确地被引入汉密尔顿函数，那么最大化问题的解可以在(y,π)空间中几何地表示出来。这种图形的表示法被称为相位图。通过一个经济学上有意义的具体问题我们可以最好地展示其用途。我将在例题 10.2 中给出一个应用的例子。

例题

例题 10.1：生命周期储蓄

考虑一个有已知寿命 T 的工人，在其生命期间他将赚取为常数 w 的工资，对积累的储蓄他将获得为常数 r 的利率，或者对积累的债务他将支付同

样的利率。所以,当他累积的资产存量(如果为负就表示债务)等于 k 时,他的流量收入就是$(w+rk)$。记 c 为他的消费流,那么资本积累就由下式决定

$$\dot{k}=w+rk-c$$

注意 k 和 c 都是 t 的函数。一般情况下我假定读者都明白自变量是什么,只有在必要时才把一些特别的值明确地列出来。

用术语说,k 为状态变量,c 为控制变量。假定没有遗产也没有遗赠,那么终端条件为

$$k(0)=k(T)=0 \tag{10.26}$$

假定选择不受其他限制。瞬时效用函数为 $\ln(c)$,效用折现率为 ρ,那么要最大化的目标为

$$\int_0^T \ln(c)e^{-\rho t}\,\mathrm{d}t$$

为了使用最大值原理,定义汉密尔顿函数

$$H=\ln(c)e^{-\rho t}+\pi(w+rk-c) \tag{10.27}$$

使 H 取得最大值的关于 c 的条件为

$$c^{-1}e^{-\rho t}-\pi=0 \tag{10.28}$$

把它代入(10.27)式,最大化的汉密尔顿函数就是

$$H^*=-(\ln(\pi)+\rho t)e^{-\rho t}+\pi(w+rk)-e^{-\rho t}$$

关于 k 和 π 的微分方程为

$$\dot{k}=\partial H^*/\partial\pi=w+rk-\pi^{-1}e^{-\rho t} \tag{10.29}$$

和

$$\dot{\pi}=-\partial H^*/\partial k=-r\pi \tag{10.30}$$

(10.30)式的通解是显然的:

$$\pi=\pi_0 e^{-\rho t} \tag{10.31}$$

其中 π_0 为待定的常数,把它代入(10.29)式,我们有

$$\dot{k}=w+rk-\pi_0^{-1}e\,(r-\rho)t$$

现在

$$d(ke^{-rt})/dt = (\dot{k} - rk)e^{-rt} = we^{-rt} - \pi_0^{-1}e^{-\rho t}$$

积分可得

$$ke^{-rt} - k(0) = w[1 - e^{-rt}]/r - \pi_0^{-1}[1 - e^{-\rho t}]/\rho$$

由于我们已经知道 $k(T)$，所以这个方程就能确定 π_0，从而完成了求解过程。

即使不知道完整的解，我们也可以发现一些经济学上重要的事实，把(10.31)式代入(10.28)式，我们有

$$c = \pi_0^{-1}e^{(r-\rho)t}$$

该式表明，当 $r>\rho$ 时，这个工人的最优消费水平会在他的一生中不断地增长。由于在具有相同的折现值的意义上消费和工资必须在他一生中实现平衡，这意味着在生命的早期有 $c<w$，而在后面的年份中则有 $c>w$。换句话说，消费者在早年储蓄，积累起资产，而到了生命的晚年则用完这些储蓄。如果 $r<\rho$，情况则相反。或许某些制度上的约束会阻止他在生命早期负储蓄从而使他持有负资产。当然，当所有消费者都试图负储蓄时，整个经济就不可能处于均衡状态。不过，这是另外的问题了。

以上只是生命周期储蓄的最简单的例子。该理论可以被扩展以包括更复杂的偏好、劳动供给和退休选择、征税、不确定性、流动性约束，以及很多其他的现实特征。

例题 10.2：最优增长

这也是一个最优储蓄的问题，只是从将整个经济视作一个整体的视角出发。这一视角的转换为这个问题带来了两个新特点。第一，储蓄的回报率不再是一个外生的市场利率，就像在个人储蓄的例子中那样，相反，它必然是内生的资本边际产出。第二，从逻辑上说，这一计划不存在逻辑上的终止时刻。通过正式设定 $T=\infty$，我将展开该理论的讨论，并只在推导过程中提及相伴随的复杂之处。

假定积累的储蓄转变为作为标量的资本存量 k，产出的流量则由生产函数 $F(k)$ 给出。通常假定 F 是一个递增的严格凹函数，且 $F(0)=0$，$F'(0)=\infty$。资本以 δ 的比例折旧。如果消费流是 c，那么资本积累方程为

$$\dot{k} = F(k) - \delta k - c \tag{10.32}$$

初始资本存量 $k(0)$ 给定，且没有其他的约束。

消费流的效用为$U(c)$，递增且严格凹。效用折现率为ρ，那么要最大化的目标为

$$\int_0^T U(c)e^{-\rho t}\,\mathrm{d}t$$

一个明显的潜在困难是要保证这个积分的收敛性。这需要ρ充分大，此处我略过了细节。

为了应用最大值原理，定义汉密尔顿函数：

$$H = U(c)e^{-\rho t} + \pi[F(k) - \delta k - c]$$

c使H取得最大值的条件为

$$U'(c)e^{-\rho t} = \pi \tag{10.33}$$

π满足的微分方程为

$$\dot{\pi} = -\partial H/\partial k = -\pi[F'(k) - \delta] \tag{10.34}$$

我们可以通过(10.33)式求解出c，然后将这结果代入(10.32)式，再结合(10.34)式就能给出关于k和π的一对微分方程。事实上，采取$\pi e^{\rho t} = \varnothing$的形式更加方便，因为关于$k$和$\varnothing$的那对微分方程不会显性地含有时间。不过，我在这里将采用一种不同的方法来处理k和c。

微分(10.33)式得到

$$\dot{\pi} = [U''(c)\dot{c} - \rho U'(c)]e^{-\rho t}$$

使用(10.34)并化简，我们得到

$$\frac{\dot{c}}{c} = \frac{F'(k) - (\rho + \delta)}{\eta(c)} \tag{10.35}$$

其中$\eta(c)$是弹性，它表示当消费增加时消费的边际效用下降的幅度：

$$\eta(c) = -cU''(c)/U'(c)$$

观察可知，例题10.1有一个形式上完全相同的结构，其中，$F'(k)$为常数r，$\delta=0$，$\eta(c)$为常数1。

现在我们把(10.32)式和(10.35)式作为k和c的一对微分方程。时间没有显性地进入方程式，所以我们可以在相位图中画出解来；见图10.1。给定任意一点(k,c)，我们可以从微分方程中找出变动速度$\dot{k}$，$\dot{c}$，这可以用附

在(k,c)上的小的向量箭头来表示。如果我们在各个点上都画出这样的箭头，把这些连续的小箭头连接起来，就可在(k,c)空间中找出整个的运动路径。给定一个初始点，微分方程就决定了沿着经过该点的解的路径随后将发生的变动。两条这样的路径是不可能相交的，因为给定一个初始点，运动的方向是由微分方程惟一决定的。

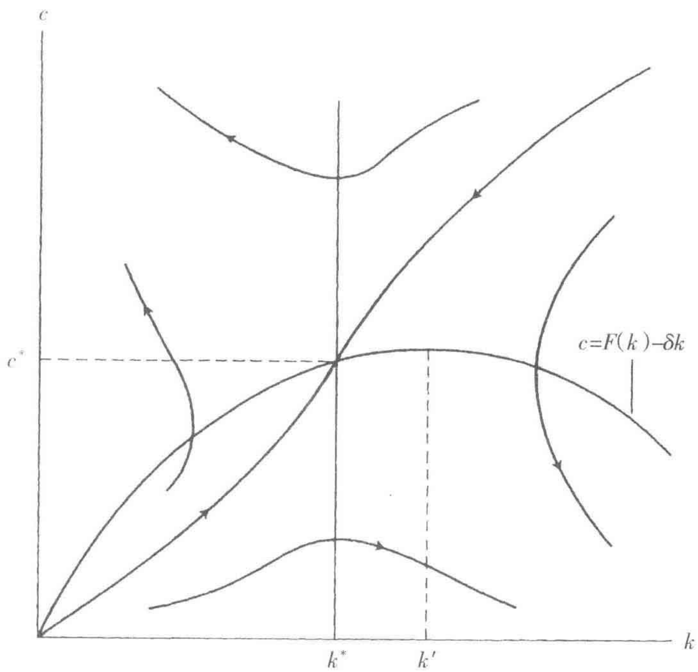

图 10.1　最优增长的相位图

理解相位图的最简单的方法就是认识到每一个 k 和 c 都是可以增加或者减少的，这样整个空间可以分成四个区域，每一个区域分别对应于朝东北、东南等等方向的运动。从(10.32)式中，我们看到，如果 $c<F(k)-\delta k$，那么 k 就增加，这恰是曲线 $c=F(k)-\delta k$ 以下的区域。当 $F(k)-\delta k$ 取到最大值时，即 $k=k'$时，该曲线达到峰值，此时满足 $F'(k')=\delta$。再来看(10.35)式，我们看到当 $F'(k)>\rho+\delta$ 时，c 增加；注意到由于 U 为递增的严格凹函数，则 $\eta(c)>0$。但是 F 也是递增的严格凹函数，因此当 $k<k^*$ 时，c 增加，其中 $F'(k^*)=\rho+\delta$。此外，当 $\rho>0$ 时，这是收敛所需的条件，我们有 $k^*<k'$。综合所有这些信息，我们得到图中所给出的路径情形。

记 $c^*=F'(k^*)-\delta k^*$，我们看到恰有两条路径收敛到(k^*,c^*)，一条从左边另一条从右边收敛到此处。其他所有的路径都是发散的，向坐标轴

中的一条渐近甚至与坐标轴相交。

但由于给定的是 $k(0)$ 而不是 $c(0)$，因此我们必须尝试 $c(0)$ 所有的可能性，来看看它们会导致怎样的结果。如果我们选择的 $c(0)$ 使得从 $(k(0), c(0))$ 出发的路径收敛到 (k^*, c^*)，那么就万事大吉。如果我们选择了任何其他的 $c(0)$，路径最终都会收敛到坐标轴中的一条。沿着 k 轴，消费水平下降到零。沿着 c 轴，消费先是上升，但是一段时间后资本耗尽，那么最终资本、产出从而消费都会跌到零。这两种可能性都不尽如人意。这表明，$c(0)$ 的正确选择是在 $k(0)$ 所对应的稳定路径上。事实上，我将在第 11 章中简要讨论对充分条件的要求，该要求表明，这样的选择正是正确的选择。

本例的解有一个明显的特征：当 k 越大时 c 也越大。但是图形并不能告诉我们 $c/F(k)$ 是否也随着 k 的增加而增加。事实上，对于特定形式的 F 和 U，$c/F(k)$ 是增加还是下降都有可能。所以我们不能下一般的结论说，越是富有的经济就应该把越多比例的收入用于储蓄。

习题

习题 10.1：生命周期储蓄

把例题 10.1 中瞬时效用函数改为如下形式，求解该问题，

$$U(c) = c^{1-\varepsilon}/(1-\varepsilon), \varepsilon > 0$$

边际效用为 $U'(c)=c^{-\varepsilon}$。前面那个例子是 $\varepsilon=1$ 时的特例，可以通过洛必达法则取极限而加以验证。

接下来假定消费者继承了资产 k_0 作为遗产，并且计划留下 k_1 作为遗赠。k_1 最多可以有多大，以使得这个问题还能有可行的解？

习题 10.2：最优增长

解释定义在例题 10.2 中的变量 $\varnothing$，证明 k 和 $\varnothing$ 满足下面一对微分方程

$$\dot{k} = F(k) - \delta k - G(\varnothing)$$

和

$$\varnothing = -\varnothing[F'(k) - \rho - \delta]$$

其中 G 是 U' 的反函数。画出相位图，它看上去应该和图(10.1)相类似，但是上下倒置。完成这个问题的求解。

习题 10.3：进入阻挠

一个行业在时刻 t 的需求函数由下式给出

$$q(t) = a - b\,p(t)$$

其中 a 和 b 是正的常数，$p(t)$ 和 $q(t)$ 分别代表价格和产量。这个行业中有一个大的企业制定价格，许多小的企业接受这个价格并出售它们全部的产出。一旦大企业制定的价格高于 p^*，新的小企业就会进入。把 $x(t)$ 记作这些小企业的产出。初始的 $x(0)$ 是给定的。$x(t)$ 满足如下的微分方程

$$\dot{x} = k[p(t) - p^*]$$

大企业的销售量是 $q(t) - x(t)$，它的平均成本为常数 c。因而它的利润折现值为

$$\int_0^{\infty} [p(t) - c][a - x(t) - bp(t)]e^{-\rho t}\,\mathrm{d}t$$

其中 ρ 为利率，假定 $p^* > c$。

将最大值原理应用于这一问题，把 x 视为状态变量，p 为控制变量。在 (x, p) 空间中画出相位图。找出大企业最优定价策略的数量特征。并找出在怎样的参数条件下，当 t 趋于∞时，竞争性企业依旧保持正的销售量。

▶ 进一步阅读

最大值原理及其应用的一个更严格的分析见 Intriligator(1971，参见第 2 章的进一步阅读)第 14 章和第 16 章。该书第 11 章和第 12 章还给出了一些相关的思想和技巧。

最优储蓄理论的相对简单的阐述见 Robert M. Solow, *Growth Theory: An Exposition*, Oxford University Press, 1970，第 5 章；和 Avinash K.

Dixit, *The Theory of Equilibrium Growth*, Oxford University Press, 1976,第5章和第7章。

进入阻挠问题(习题10.3)的细节见 Darius W. Gaskins,"Dynamic Limit Pricing: Optimal Pricing under Threat of Entry", *Journal of Economic Theory*, 3(3), 1971, pp. 306—22。

虽然在应用最大值原理时,时间是一个自然的独立变量,但其中的数学应用并不依赖于这一解释。事实上,城市中的最优道路设计已经使用相同的技术来处理了,此时的 t 就可以被解释为距离。一个例子可见 Avinash K. Dixit,"The Optimum Factory Town", *Bell Journal of Economics and Management Science*, Autumn 1973, pp. 637—51。

11

动态规划

11.1 贝尔曼方程

在第 10 章中我们研究了时间维度上的最优化问题。表示存量或状态变量增量的约束(10.1)式和目标函数加性可分的形式(10.3)式这两大特征,使我们能将一阶条件表达成一种有用的特殊形式,即最大值原理。动态规划是解决相同问题的另一种方法。当时间和不确定性同时出现时,现实中也往往确实如此,动态规划就显得特别有用。为保证阐述的简单性让我们先从时间开始,并在后面引入不确定性。

给定初始存量向量 y_0 和末期存量向量 y_{T+1},我们最大化

$$\sum_{t=0}^{T} F(y_t, z_t, t) \tag{10.3}$$

须满足约束

$$y_{t+1} - y_t = Q(y_t, z_t, t) \tag{10.1}$$

和

$$G(y_t, z_t, t) \leqslant 0 \tag{10.2}$$

其中,$t=0,1,\cdots,T$。现在保持末期存量的规定不变。同在第 5 章中一样,我们可以将由此产生的最大值定义为初始存量的一个函数,即 $V(y_0)$。导数向量 $V_y(y_0)$就是这些初始存量的影子价格向量。

目标函数和约束条件的可分性使得我们可以将此大大地一般化。我们不是从时点 0 开始,而是考虑另一个特别的时间,比如 $t=\tau$。对始点为

τ的决策，关于过去惟一重要的事就是以往的决策产生的存量向量 y_τ。我们可以将它看作一个参数，并将整个问题在 τ 处重新开始。换言之，我们最大化像(10.3)式那样的一个和，但只从 τ 到 T 展开，也必须满足像(10.1)式和(10.2)式一样的约束，但只对 $\tau,\tau+1,\cdots,T$ 成立。令 $V(y_\tau,\tau)$ 为这个问题的最大值函数，显性变量 τ 是必要的，因为求和的极限取决于它。当我们从 τ 处给予初始存量 y_τ 一个小的增量时，导数向量 $V_y(y_\tau,\tau)$ 即为最大化的和的边际增量，即从 τ 开始的最优化问题中初始存量的影子价格向量。

当我们将始点为 τ 的子问题加入到始点为 0 的整个问题中时会发生什么呢？在第 10 章中我们可以将 τ 时的约束(10.1)式的拉格朗日乘子解释为在 $\tau+1$ 时的存量的影子价格向量。这个约束的微小放松意味着 $y_{\tau+1}$ 的一个外生增量，而乘子告诉我们目标函数(10.3)式会由此而增加多少。初看起来这不同于始点为 $\tau+1$ 的子问题的导数向量 $V_y(y_{\tau+1},\tau+1)$。在整个问题中，我们在时点 0 知道，在 $\tau+1$ 时的存量将有一个小的增加。因而我们可以预先计划，并在早先几期中改变控制变量。例如，如果在时点 0 我们知道，在 $\tau+1$ 时将有一笔飞来横财，我们会在这笔横财到来之前和之后消费得更多。但包络定理施以援手。当我们在研究一阶导数所涉及的微小变动时，其对目标函数的直接影响即是全部的影响，由选择变量的最优重新调整而导致的间接影响可以被忽略不计。因而我们确实可以始终将导数 V_y 等同于影子价格 π。

现选择任意的 t，考虑在那个时候关于控制变量 z_t 的决策，以及由于选择任意特定的 z_t 而带来的结果。根据(10.1)式它将产生下一期的存量 y_{t+1}。此后我们仍须求解始点为 $t+1$ 的子问题，并得到最大值$V(y_{t+1},t+1)$。因而在 t 时刻以 y_t 开始的总值可以分解成两项：即刻得到的$F(y_t,z_t,t)$和稍后得到的 $V(y_{t+1},t+1)$。z_t 的选择应使这两项之和最大。换言之，

$$V(y_t,t)=\max_{z_t}\{F(y_t,z_t,t)+V(y_{t+1},t+1)\} \tag{11.1}$$

只要满足这一时刻 t 的约束(10.1)式和(10.2)式。

这个等式给我们提供了一条求解原来那个最优化问题的强有力的途径。其思路是：从末期开始并递归地向前面时点进行。时点 T 没有将来，只有固定的末期存量 y_{T+1}。因此，

$$V(y_T,T)=\max_{z_T}F(y_T,z_T,T)$$

满足

$$y_T+Q(y_T,z_T,T)=y_{T+1},G(y_T,z_T,T)\leqslant 0$$

原则上这是一个简单的静态最优化问题，并可以得到最大值函数$V(y_T,T)$，因而可以被用在 $t=T-1$ 时的(11.1)式的右边。同时，这也是另一个静态问题，并得到最大值函数 $V(y_{T-1},T-1)$。如此一直下去直到回到 0 时刻。在现实操作中这种方法只对最简单的问题有用。只有当函数 F、G 和 Q 具有非常简单的形式时这类问题的解析解才有可能存在。对更困难的问题可以算出数值解来，但如果状态变量是二维以上的向量，则数值解甚至也立即变得不可得。幸运的是，这个强有力的方法只是一个"后台屏障"，在许多经济学的应用中，有更好的方法找到解，或者至少可以得到一些关于解的有益的洞见。

作为静态规划问题的后续，这个在时间维度上的最优化方法由理查德·贝尔曼首先提出来，并被命名为动态规划。不管 t 时刻的决策是什么，随后的决策对从$(t+1)$开始的子问题而言应该是最优的，这一思想就是著名的贝尔曼最优化原理。最大值函数 $V(y_t,t)$被称为贝尔曼值函数，而且方程(11.1)式被称为贝尔曼方程。

让我们来看贝尔曼方程右边的最大化问题。根据(10.1)式替代掉 y_{t+1}，我们将选择 z_t 以最大化

$$F(y_t,z_t,t)+V(y_t+Q(y_t,z_t,t),t+1)$$

满足

$$G(y_t,z_t,t)\leqslant 0$$

令 λ_t 表示关于约束的乘子的行向量，一阶条件为

$$F_z(y_t,z_t,t)+V_y(y_{t+1},t+1)Q_z(y_t,z_t,t)-\lambda_t G_z(y_t,z_t,t)=0$$

将导数 V_y 看作影子价格π，上式就变成

$$F_z(y_t,z_t,t)+\pi_{t+1}Q_z(y_t,z_t,t)-\lambda_t G_z(y_t,z_t,t)=0$$

这个式子就是使第 10 章中定义的汉密尔顿函数 $H(y_t,z_t,\pi_{t+1},t)$在(10.2)式的单期约束下最大化的一阶条件。因而在确定选择变量时，动态规划得

到了与最大值原理相同的规则。

事实上在解决时间维度上的最优化问题时，最大值原理和动态规划是两种完全等价的方法。在解决特定的问题时，哪一种更简单就用哪一种。一般而言，当时间是连续的且没有不确定性时最大值原理更好；在相反的情形中，即存在离散时间和不确定性时，动态规划则更好。但这并非是一成不变的原则。

在本章后面我将通过一些经济学应用的例子来说明动态规划的应用。而在这一部分的最后，我将用它以不同的方式推导出跨期的套利方程(10.13)式。当 z_t 是最优选择时，(11.1) 式以等式成立，即

$$V(y_t, t) = F(y_t, z_t, t) + V(y_{t+1}, t+1)$$

对 y_t 进行微分，注意 y_{t+1} 依赖于 y_t，并在右边应用包络定理。因此，

$$V_t(y_t, t) = F_t(y_t, z_t, t) + V_y(y_{t+1}, t+1)(1 + Q_y(y_t, z_t, t)) - \lambda_t G_y(y_t, z_t, t)$$

利用影子价格 π，这个等式就变为(10.13)式了。

11.2 不确定性

动态规划特别适宜兼有时间和不确定性的最优化问题。假定存量 y_t 沿时间路径演化的过程有一个随机的成分。给定初始时刻 t 的存量 y_t 和这个时期中的控制变量 z_t，我们只知道下一时期存量 y_{t+1} 的概率密度函数，将其写成 $\emptyset(y_{t+1}, y_t, z_t)$。为了表示上的清晰，变量们被分离出来。第一个变量 y_{t+1} 就是实际的随机变量向量，它的概率密度函数即如上所示；其他的变量类似一些能改变该分布的函数形式的参数。举一个简单的例子，y_{t+1} 可以是一个向量正态分布，其均值向量 μ 和方差—协方差矩阵 Σ 都依赖于 (y_t, z_t)。一个更特殊的情形是，μ 可以等于 $y_t + Q_y(y_t, z_t, t)$，就是在前面讨论过的没有不确定性的 y_{t+1} 的值。

现在的问题是使(10.3)式的数学期望最大，对所有的 t 都要满足(10.2)式，以及由函数 $\emptyset$ 所刻画的 y_{t+1} 的随机运动律来替代的(10.1)式。将从 t 开始的子问题的最大值函数写成 $V(y_t, t)$。暂且固定 z_t 的选择。考虑我们在时期 $(t+1)$ 之初得知 y_{t+1} 的实际值后会发生什么。余下的决策将会最优地进行，并且得到 $V(y_{t+1}, t+1)$。如果我们从时期 t 的角度来看，这

仍然是一个随机变量,而我们关心的是它的数学期望

$$E[V(y_{t+1},t+1)]=\int V(y_{t+1},t+1)\,\emptyset(y_{t+1};y_t,z_t)\mathrm{d}y_{t+1} \quad (11.2)$$

其中积分在 y_{t+1} 分布的整个区域中进行。因而最优原理变为

$$V(y_t,t)=\max_{z_t}\{F(y_t,z_t,t)+E[V(y_{t+1},t+1)]\} \quad (11.3)$$

(11.3)式右边的最大化比相应的确定性的情形(11.1)式更有难度。关于 z_t 的一阶条件要求在积分中 $\emptyset$ 对 z_t 的微分,其结果可能难以刻画和理解。但原则上(11.3)式允许我们从 T 开始并且像以前一样递归地求解问题一直向后到 0。在简单有效的模型中,我们可以得到完整的解析解。在本章末,我会给出两个将不确定性下的动态规划应用到经济学问题中的例子,并给出一些参考文献,使读者可以从中寻求对该论题的进一步研究。

11.3 连续时间

最大值原理可以被用来阐述离散或连续时间的问题,动态规划也可以。回想在第 10 章中的那个问题,它可以被形式化为最大化

$$\int_0^T F(y(t),z(t),t)\mathrm{d}t \quad (10.17)$$

满足运动律

$$\dot{y}(t)=Q(y(t),z(t),t) \quad (10.15)$$

及瞬时约束

$$G(y(t),z(t),t)\leqslant 0 \quad (10.16)$$

将 $V(y(t),t)$定义为从 t 开始的子问题的最大值函数。在接下来的那个小的时间区间 $\mathrm{d}t$ 上,假定控制变量取值为 $z(t)$,则在这个小区间上它对目标函数的贡献将是 $F(y(t),z(t),t)\mathrm{d}t$。在$(t+\mathrm{d}t)$时存量的增量为

$$y(t+\mathrm{d}t)-y(t)=Q(y(t),z(t),t)\mathrm{d}t$$

且此后的最优政策将得到$V(y(t+\mathrm{d}t),t+\mathrm{d}t)$。贝尔曼最优原理给出

$$V(y(t),t)=\max_{z(t)}\{F(y(t),z(t),t)\mathrm{d}t+V(y(t+\mathrm{d}t),t+\mathrm{d}t)\}\quad(11.4)$$

且满足(10.15)式和(10.16)式。将右边用泰勒级数展开：

$$\begin{aligned}V(y(t+\mathrm{d}t),t+\mathrm{d}t)&=V(y(t),t)+V_y(y(t),t)[y(t+\mathrm{d}t)-y(t)]\\&\quad+V_t(y(t),t)\mathrm{d}t\\&=V(y(t),t)+V_y(y(t),t)Q(y(t),z(t),t)\mathrm{d}t\\&\quad+V_t(y(t),t)\mathrm{d}t\end{aligned}$$

代入(11.4)式，我们看到$V(y(t),t)$从两边消去，然后方程可以除以$\mathrm{d}t$。（更加严格的证明应该用有限增量Δt，然后取极限。）这样就得到

$$0=\max_{z(t)}\{F(y(t),z(t),t)+V_y(y(t),t)Q(y(t),z(t),t)\}+V_t(y(t),t)\quad(11.5)$$

满足瞬时约束(10.16)式。

由于$V_y(y(t),t)$就是影子价格$\pi(t)$，最大化的部分就是第10章中的汉密尔顿函数$H(y(t),z(t),\pi(t),t)$，且结果是取得最大值的汉密尔顿函数$H^*(y(t),\pi(t),t)$。因此(11.5)式可以被写成

$$V_t(y(t),t)+H^*(y(t),V_y(y(t),t),t)=0\quad(11.6)$$

由于H^*是已知函数，所以这是关于贝尔曼值函数V的偏微分方程。利用具体问题中具有适当定义的边界条件，它可以被求解出来。同样的道理，解析解只有在非常简单及特殊的情形中才能被解出来。但是随着计算技术的提高，数值解变得越来越可行了。

11.4 横截条件

到目前为止我们一直保持终端时间T和与此相联系的目标存量y_{T+1}固定，同时为建立贝尔曼值函数而允许初始时间和存量变动。但是反过来的途径也会得到一些有价值的洞见。我将在连续时间情形中尝试这个方法，而将相应的离散时间的情形留给读者去推导。将F在$[0,t]$上的积分的最大值写为$W(y,t)$，初始存量向量固定为y_0，在t时刻要求有可变的y。期末必须留下来的存量越大，取得最大值的积分就越小，且影子

价格的解释变为 $\pi(t)=-W_y(y(t),t)$。将此问题分成一个初始区间 $[0,t-dt]$ 和一个小的最终区间 $[t-dt,t]$，像上面一样我们可以应用贝尔曼最优原理，得

$$W_t(y(t),t)-H^*(y(t),-W_y(t),t),t)=0 \tag{11.7}$$

这个可供选择的方法使我们可以将该理论扩展到更一般的端点条件。接受一个固定的初始时间和在那个时刻由历史给定的初始存量是自然的，但不同于固定的日期和存量的终端条件也是容易想象的。例如，我们可能希望在最小时间内获得一个给定的目标存量，或者面对在终端时间和存量之间的更一般的权衡。假定目标是最大化积分(10.17)式，在满足一个一般约束

$$J(y(T),T)\leqslant 0 \tag{11.8}$$

的情况下同时选择 T 和 $y(T)$。这可以分两步来处理。首先我们将 T 和 $y(T)$ 看作是固定的，求解标准的动态规划问题，并找到上面定义的值函数 $W(y(T),T)$。然后我们选择 T 和 $y(T)$ 使得这个函数在满足约束(11.8)式下最大化。这是一个标准的静态最优化问题，有一阶条件

$$W_y(y(T),T)=\xi J_y(y(T),T),W_t(y(T),T)=\xi J_t(y(T),T)$$

其中 ξ 为拉格朗日乘子。用影子价格和汉密尔顿函数的表示法，这些条件变为

$$\pi(T)=-\xi J_y(y(T),T),H^*(y(T),\pi(T),T)=\xi J_t(y(T),T) \tag{11.9}$$

这表明，当两者都在最优终端取值时，向量 $(\pi,-H^*)$ 应该与向量 (J_y,J_t) 平行。由于后一向量垂直于约束平面 $J(y,t)=0$，这些条件表明前一向量也垂直于同一平面。因而条件(11.9)式被称为横截条件。

举个例子，假定我们希望在最少的时间内达到一个给定的目标，比如 y^*。那么在满足 $y(T)=y^*$ 的条件下 $W(T)=-T$ 要实现最大。现 J_t 等于零，(11.9)式表明在最优选择的端点，汉密尔顿函数 H^* 应该也为零。类似地，如果 T 固定但 $y(T)$ 不受限制，则 J_y 等于零，并且横截条件为 $\pi(T)=0$。这些都作为边界条件，帮助我们确定动态规划问题的解。

11.5 无限期界

跨期最优化问题的另一个扩展在许多经济学问题中也很重要。通常不存在自然的方法确定决策最优化的终端日期。事实上，我们几乎无法提前固定一个日期，并宣称可以完全忽略在此日期以后的考虑。对个体而言这或许并不是一个重要的问题，但当我们考虑越来越广泛的决策背景时，如一个扩大的家庭、一个企业和整个经济，它就变得越来越重要了。保持时间期界有限，我们能认识到终端存量将提供超出时限的效用流，这样就间接地将未来考虑进来了。但这是一个不完美的解。如果对于时限之外的未来没有给予明确的关注，我们就不能确定正确的终端存量目标，或像上面的 J 一样的评价函数。但是那无非意味着求解一个与原来的问题完全一样的但有更长的时限的问题。当然，这个论证不存在逻辑上的终止点，从而迫使我们来考虑无限期界。

当我们考虑在一个无限时间期界上的决策时，就陷入了一些技术困境。最重要的问题就是积分(或在离散时间情形下的求和)可能不收敛。例题10.2的最优增长问题就是这样的一个典型例子。假定生产函数和效用函数都是线性的，

$$F(k)=\beta k,\ U(c)=c$$

并且资本的边际产品 β 超过了效用折现率 ρ。不考虑折旧，即设定 $\delta=0$。现在考虑将一单位产出从消费转为储蓄，并且让其累计到时间 T，到那时再消费这增加的产出。在 T 时这额外的产出变为 $e^{\beta T}$，它的消费现值为 $e^{(\beta-\rho)T}$，超过放弃当前消费的机会成本1。因而推迟消费总是值得的，且越来越长的推迟可以使效用积分越来越大。但这种策略的极限意味着没有一点儿消费，而这正是所有策略中最坏的。

消除这种反常情形的充分并且通常也是必要的条件，要求终端存量的影子价值应该趋向于零：

$$\lim_{T\to\infty}\pi(T)y(T)=0 \tag{11.10}$$

对这一命题的严格证明超出了我们这里讨论的范围。但注意，它是由非负终端存量规定的有限期界问题的横截条件的一个自然扩展(恰好就是互补

松弛条件)。

除此新条件,无限期界最优化没有什么不同于有限期界的情形。贝尔曼最优原理马上告诉我们为什么。考虑任何被更大的问题固定初始和终端条件的有限期界的子问题。对此子问题最大值原理或动态规划条件都适用。但子问题的初始和终端时间可以是任意的,因而这些条件事实上必须在$(0,\infty)$的整个期界上都成立。

无限期界横截条件的一个应用是例题 10.2 的最优增长问题。在相位图中收敛到稳定状态(k^*,c^*)的路径满足这一条件;而发散的路径一般不满足。这使得我们可以拒绝发散的路径,并对给定的初始资本$k(0)$,选择初始消费水平$c(0)$使之位于一个收敛的路径上。

例题

例题 11.1:搜寻

这是一个简化的工作搜寻模型。它并不旨在符合现实;而只是向你介绍动态规划方法,并让你为本章末所引用的更加丰富的模型做好准备。

在这个经济中存在一个支付不同工资的各种工作的完整范围。累积分布函数——一项随机选择的工作支付w或者更少的概率——为$\Phi(w)$。相应的密度函数是$\emptyset(w)=\Phi'(w)$。个人知道这些概率,但不知道任何特定的工作可以在哪里找到或者它支付工资的细节;他必须投入到那些信息的搜寻中去。为了获知一项工作的情况,他必须以零收入保持失业状态并且搜寻工作一个时期。在下一期初,他可以接受刚找到的工作或者继续搜寻工作。如果他拒绝此项工作,他以后不能回过来再接受它。他的目标是使他工资的现值的数学期望最大。如果利率是r,定义折现因子$\delta=1/(1+r)$,因而在t年后的 1 美元现在值δ^t美元。

考虑一个开始其最后一期工作生涯的人。假定在他前一期的搜寻中他观察到的工作支付的工资$x>0$。接受此工作并工作一个时期要明显胜过仍然处于失业状态。这样他将接受任何正的x,他的值函数即为$V_1(x)=x$,其中下标代表剩下的工作生涯的时期数。

接下来假定同一个搜寻者还有两期工作生涯,他刚观察到的工资为x。

如果他接受此工作，他这一期和下一期得到 x，即 $x+\delta x$ 的折现值。如果他搜寻，这一期他没有收入，但获知了另一工作前景和其支付的工资 y。由于下一期是他最后一期工作生涯，他必将接受它。因而下一期的期望收入是数学期望

$$E[y] = \int_0^\infty y\varnothing(y)\mathrm{d}y$$

这必须乘以 δ 以得到在这一期的现值。

在这两者中，搜寻者将选择较好者，因而

$$V_2(x) = \max((1+\delta)x, \delta E[y]) \tag{11.11}$$

存在临界值

$$x_2^* = \delta E[y]/(1+\delta) \tag{11.12}$$

使还有两期工作生涯的搜寻者将接受最近的工作机会 x，当且仅当它大于 x_2^*。因此 x_2^* 被称为这个搜寻者的保留工资。用相同的方式我们可以为搜寻者在他的最后一期定义 x_1^*，如我们上面所看到的，$x_1^*=0$。

更一般地，还有 n 期，搜寻者接受最近的工资为 x 的工作得到

$$(1+\delta+\delta^2+\cdots\delta^{n-1})x = x(1-\delta^n)/(1-\delta)$$

继续搜寻会观察到一个新的 y 和开始于下一期(因而以 δ 折现)并且还剩 $(n-1)$ 期的一个最优决策。贝尔曼方程变为

$$V_n(x) = \max(x(1-\delta^n)/(1-\delta),\ \delta E[V_{n-1}(y)]) \tag{11.13}$$

这产生了保留工资 x_n^*。一个归纳法的论证立即表明保留工资的序列 x_n^* 随 n 增加，即面临更长工作生涯的搜寻者对他们所接受的工作更挑剔。利用(11.13)式我们也可以递归地求解出函数 $V_n(x)$。

如果我们令 n 趋向无穷大，$V_n(x)$ 和 $V_{n-1}(x)$ 同样地收敛到一个极限函数 $V(x)$，它满足

$$V(x) = \max(x/(1-\delta), \delta E[V(y)]) \tag{11.14}$$

由于同一个函数 V 出现在两边，(11.14)式的解包含循环(或者更恰当地，不动点)逻辑。把上式的右边看作一个算子，或一个“函数的函数”，即以函数 V 开始并产生一个新的函数。因而我们寻找以这种方式得到自己的一个特

定的 V。此见解也提供了一个数值计算的方法。从任何的 V 开始，应用此算子得到一个新的 V，然后对其应用同一个算子，一直下去。只要解存在，这个过程就会收敛到这个解，而在这个例题中，只要 $\delta<1$，解就会存在。

从直觉上说，这个解应该与长期但有限的生命跨度情形下的解具有相同的定性特征，即应该存在一个保留工资，或临界值 x^*，提供的工资在此之上的工作机会被接受，而提供比这个低的工资的工作机会就会促使搜寻者继续搜寻。让我们沿着这个假设继续前进，看它能带来什么结果。对 $x>x^*$，(11.14)式变为

$$V(x)=x/(1-\delta)$$

对较小的 x，我们有

$$V(x)=\delta E[V(y)]$$

它与 x 无关。当 x 从右边和左边趋向于 x^* 时，这两个式子的极限相等，

$$V(x^*)=x^*/(1-\delta)=\delta E[V(y)] \tag{11.15}$$

现在

$$\begin{aligned}
E[V(y)]&=\int_0^{\infty}V(y)\varnothing(y)\mathrm{d}y\\
&=\int_0^{x^*}V(x^*)\varnothing(y)\mathrm{d}y+\int_{x^*}^{\infty}[y/(1-\delta)]\varnothing(y)\mathrm{d}y\\
&=V(x^*)\Phi(x^*)+[1/(1-\delta)]\int_{x^*}^{\infty}y\varnothing(y)\mathrm{d}y
\end{aligned}$$

因此

$$V(x^*)[1-\delta\Phi(x^*)]=[\delta/(1-\delta)]\int_{x^*}^{\infty}y\varnothing(y)\mathrm{d}y$$

或

$$x^*[1-\delta\Phi(x^*)]=\delta\int_{x^*}^{\infty}y\varnothing(y)\mathrm{d}y \tag{11.16}$$

由于我们知道函数 Φ 和 $\varnothing$，这个方程可以求解出 x^*，于是函数 V 一直为常数直到 x^*，之后等于 $x/(1-\delta)$。

例题 11.2：不确定性下的储蓄

这又是一个简化的"初学者"的例子。考虑一个有无限生命、财富为 W 的消费者，他每期赚得随机的总收益（本金加利息）r，且没有其他收入。像在习题 10.1 中一样，在任何一期的消费 C 给他带来效用

$$U(C) = C^{1-\varepsilon}/(1-\varepsilon) \tag{11.17}$$

其中，$\varepsilon > 0$。他的效用折现因子是 δ，整个时期的目标是使效用折现值的数学期望最大。

从财富为 W 的一个时期开始，如果他消费 C 和储蓄 $(W-C)$，下一期初他的随机财富将为 $r(W-C)$。将他的贝尔曼值函数写为 $V(W)$，贝尔曼方程为

$$V(W) = \max_{C}(C^{1-\varepsilon}/(1-\varepsilon) + \delta E[V(r(W-C))]) \tag{11.18}$$

求解此问题的捷径是猜测一个有具体形式的解然后证实它。由于财富是分到许多期上消费的且每期带来形式为(11.17)式的效用，我们很自然地会想到尝试以下的形式

$$V(W) = AW^{1-\varepsilon}/(1-\varepsilon) \tag{11.19}$$

其中 A 为将要决定的常数。在(11.18)式中代入 $V(W)$ 的这个形式，我们有

$$\frac{AW^{1-\varepsilon}}{1-\varepsilon} = \max_{C}\left\{\frac{C^{1-\varepsilon}}{1-\varepsilon} + \frac{\delta A}{1-\varepsilon}E[r^{1-\varepsilon}](W-C)^{1-\varepsilon}\right\} \tag{11.20}$$

一阶条件为

$$C^{-\varepsilon} - \delta AE[r^{1-\varepsilon}](W-C)^{-\varepsilon} = 0$$

这化简为

$$C = W/(1+B) \tag{11.21}$$

其中我用了缩写

$$B = (\delta AE[r^{1-\varepsilon}])^{1/\varepsilon}$$

代回到(11.20)式并化简，我们得到

$$A^{1/\varepsilon}[1 - \varepsilon^{1/\varepsilon}E[r^{1-\varepsilon}]^{1/\varepsilon}] = 1$$

这决定了 A，如果

$$\delta E[r^{1-\varepsilon}] < 1$$

从而成为此问题的解存在的条件。在这个例题中它可以被证明就是保证无限效用之和收敛所需的条件。代回到(11.21)式,我们得

$$C/W = 1 - (\delta E[r^{1-\varepsilon}])^{1/\varepsilon} \tag{11.22}$$

消费占财富的比例的最优规则(11.22)式是一个相对简单的比例问题,但这个比例依赖于效用函数的参数 ε 和 δ,以及随机变量 r 的分布。但有一个特殊的情形会产生明确的解。假定 r 是对数正态分布的,即 $\ln(r)$是标准差为 σ 的正态分布,那么对数正态分布的标准公式给出

$$E[r^{1-\varepsilon}] = (E[r])^{1-\varepsilon}\exp(-\varepsilon(1-\varepsilon)\sigma^2/2)$$

为得到结果,考虑 $\varepsilon<1$ 的情形。现在保持 σ 不变时,$E[r]$的增加会降低消费—财富的比例,而 $E[r]$不变时,σ 增加会增加这个比例。如果 $\varepsilon>1$,相反的结论就会成立。

例题 11.3：最短距离

这个例题没有经济学上的内容,但它的巨大优点是这个问题在一开始就知道答案,从而使我们能更加好地集中在求解的技术上。而且,它证明了这样一个观点:尽管在理论上独立变量 t 有作为时间的自然的解释,但任何其他变量,如空间,可以起到同样的正式的作用,因此动态规划的理论也可以继续应用。

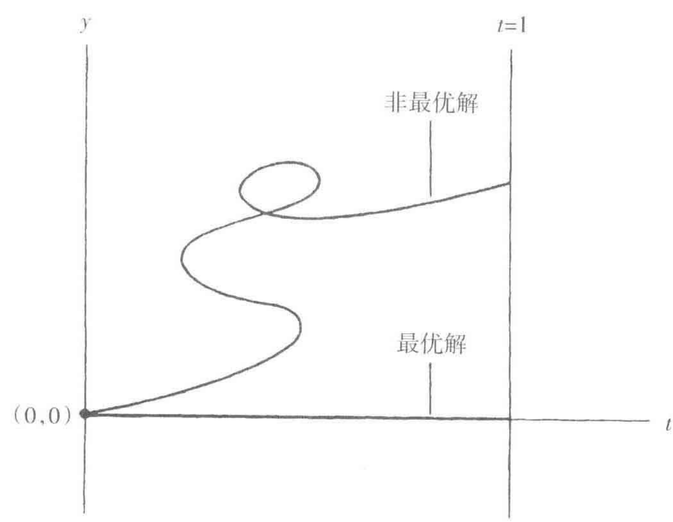

图 11.1　最短距离

考虑平面上点(0,0)和(1,0)之间最短的路径;看图11.1。称横坐标为t,纵坐标为y。显然,任何成环或弯曲的路径不可能有最短长度,因为我们可以省略掉环或S形以得到更短的路径。因而我们可以将讨论限制在y为t的单值函数的情形。相邻点(t,y)和$(t+\mathrm{d}t,y+\mathrm{d}y)$之间的距离为$[(\mathrm{d}t)^2+(\mathrm{d}y)^2]^{1/2}$。令$\mathrm{d}y/\mathrm{d}t=z$,为控制变量。因而我们将最大化

$$-\int_0^1 (1+z(t)^2)^{1/2}\mathrm{d}t \tag{11.23}$$

须满足

$$\dot{y}(t)=z(t) \tag{11.24}$$

和$y(0)=y(1)=0$。

为运用最大值原理,定义汉密尔顿函数

$$H=-(1+z^2)^{1/2}+\pi z \tag{11.25}$$

为使其最大,对z的一阶条件为

$$-(1+z^2)^{-1/2}z+\pi=0$$

化简得

$$z=\pi/(1-\pi^2)^{1/2} \tag{11.26}$$

因此最大化的汉密尔顿函数为

$$H^*=-(1-\pi^2)^{1/2} \tag{11.27}$$

现在两个微分方程为

$$\dot{y}=\partial H^*/\partial\pi=\pi/(1-\pi^2)^{1/2} \tag{11.28}$$

和

$$\dot{\pi}=-\partial H^*/\partial y=0 \tag{11.29}$$

这样,在最优路径上π为常数,于是z也为常数。但是惟有当常数z为零时,$y(1)=y(0)$。因此最短路径为连接(0,0)和(1,0)的水平直线。

我们可以应用动态规划,考虑一个更一般的问题,即找到连接(0,0)到直线$t=1$的最短的直线。首先我们考虑连接(0,0)到一般的点(T,y)的子问题,其中$T>0$但y是无限制的。继续应用上面的理论,我们看到满足要求的为直线$y(t)=ty/T$。现在剩下的问题是最优地选择y。它的横截条件为

$\pi(T)=0$。由于在最优路径上 π 为常数，它必须处处为零。于是 z 必须为零，因而 $y(T)=0$ 是最优的。换言之，从一点到一条直线的最短路径为从该点出发并垂直于给定直线的直线段。

习题

习题 11.1：搜寻

考虑例题 11.1 搜寻问题的一个变化。假定搜寻者可以回头再接受他曾经拒绝的以前的工作。因而在任何时间，状态变量 x 为到此为止他观察到的最好的工作机会。如果他拒绝了这个工作，继续搜寻并观察到 y，那么下期他将从 x 和 y 中的较优者开始。因此，代之以(11.14)式，我们有

$$V_n(x)=\max(x(1-\delta^n)/(1-\delta),\delta E[V_{n-1}(\max(x,y))])$$

证明存在一个保留工资 x_n^*，并且保留工资的序列随 n 递增。当 n 趋向无穷大时，写出贝尔曼方程，并给出 x^* 的特征。注意

$$\begin{aligned}E[V(\max(x,y))]&=\int_0^x V(x)\varnothing(y)\mathrm{d}y+\int_x^\infty V(y)\varnothing(y)\mathrm{d}y\\&=V(x)\Phi(x)+\int_x^\infty V(y)\varnothing(y)\mathrm{d}y\end{aligned}$$

习题 11.2：研究努力的强度

一种新产品需要的研究与开发过程包括许多阶段。将此想成是一个长度为 L 的连续统。如果 $x(t)$ 阶段在时点 t 完成，此时在研究与开发(R&D)项目上引起的成本流为 $c(t)$，那么完成全部阶段的过程以微分方程

$$\dot{x}=f(c)$$

展开。当所有 L 阶段都完成了，得到回报 R。如果这发生在时点 T，$r>0$ 为利率，这个项目的净现值为

$$Re^{-rT}-\int_0^T c(t)e^{-rt}\mathrm{d}t$$

我们的目标是选择完成的时间 T 和 $[0,T]$ 上的成本流 $c(t)$ 以使上述这个值最大。

将$V(x)$定义为从已完成的x阶段开始的相应问题的值函数。在一个初始的小的时间区间dt上应用贝尔曼原理,我们有

$$V(x)=\max[-c\mathrm{d}t+V(x+f(c)\mathrm{d}t)e^{-r\mathrm{d}t}]$$

像在(11.5)式的推导中一样将右边用泰勒级数展开,把这个式子转换成

$$rV(x)=\max_{c}[V'(x)f(c)-c] \tag{11.30}$$

对于情形$f(c)=c^{\alpha}$,$0<\alpha<1$,(11.30)式中的最大化可以被明确地实现。证明它,得出

$$V'(x)=\beta V(x)^{1-\alpha}$$

其中β为依赖于r和α的已知常数。求解这个方程以证明

$$V(x)^{\alpha}=R^{\alpha}+\alpha\beta(x-L) \tag{11.31}$$

推导V是凸函数,因而研究努力的最优强度开始处于一个低的水平,并随着更多阶段的完成而增加。(试着将此作为你拖延工作时给你的教师、老板等的解释。)

如果L足够大,在此项目上一点也不进行研究或许是最优的。找到L的这个极限。用最大值原理求解整个问题。

可以怎样修改问题以允许研究过程中的不确定性?

▶ 进一步阅读

动态规划的一个严格而细致的阐述在Intriligator(1971,参见第2章的进一步阅读)第13章。将其应用到如搜寻等问题,可参见Morris H. Degroot, *Optimal Statistical Decisions*, New York: McGraw-Hill, 1970。

这里的最优储蓄的处理参考了David Levhari and T. N. Srinivasan, "Optimal Saving under Uncertainty", *Review of Economic Studies*, 36, 1969, pp. 153—63。

对动态规划在投资和增长理论中的重要应用,参见Nancy L. Stokey and Robert E. Lucas, Jr., with Edward C. Prescott, *Recursive Methods in Economic Dynamics*, Cambridge, MA: Harvard University Press, 1989。

研究与开发的练习是基于 Gene M. Grossman and Carl Shapiro，"Optimal Dynamic R&D Programs"，*Rand Journal of Economics*，Winter 1986，pp. 581—93.

附录：

库恩—塔克定理

本附录的目的是提供库恩—塔克定理的严格证明的一个概略，这个定理是关于非负和不等式约束下最大化的一阶必要条件的最一般的定理。注意“概略”这个词：这不是一个详细的证明，但在这个概略中每一个富有启发性的论证可以很容易地转换成一个严格的论证。对大多数读者而言，这应该足够了。我从对第3章中这个定理的扼要重述开始：

库恩—塔克定理：假定 x 为 n 维向量，c 为 m 维向量，F 为取标量值的函数。G 为取 m 维向量值的函数。定义

$$L(x,\lambda)=F(x)+\lambda[c-G(x)] \tag{3.4}$$

其中 λ 为 m 维行向量。假定满足 $G(x)\leqslant c$ 和 $x\geqslant 0$ 时，$\bar{x}$ 使 $F(x)$ 最大，同时约束规格成立，即由那些满足 $G^i(\bar{x})=c_i$ 的 i 行组成的 $G_x(\bar{x})$ 的子矩阵有最大可能的秩。因而存在 λ 的一个值，使得

$$L_x(\bar{x},\ \lambda)\leqslant 0,\ \bar{x}\geqslant 0,\text{满足互补松弛条件} \tag{3.7}$$

和

$$L_\lambda(\bar{x},\ \lambda)\geqslant 0,\ \lambda\geqslant 0,\text{满足互补松弛条件} \tag{3.10}$$

在我们开始证明之前，我们可以简化表示法。非负约束可以表示成 $-x_j\leqslant 0,j=1,2,\cdots,n$。这样我们可以将它们包含在一般的不等式约束中，因此不等式的数量 m 将增加。我们也可以在分量函数 G^i 上选择标记使得约束 $1,2,\cdots,k$ 是紧的，约束 $k+1,\cdots,m$ 是松的，也就是

$$G^i(\bar{x})\begin{cases}=c_i,\ j=1,2,\cdots,k\\<c_i,\ i=k+1,\cdots,m\end{cases} \tag{A.1}$$

由于 $\bar{x}$ 是 $F(x)$ 满足 $G(x) \leqslant c$ 时的(局部或全局)最大值解(maximizer),因此不存在相邻的 x 使得

$$G^i(x) \leqslant G^i(\bar{x}) = c_i,\ i = 1,2,\cdots,k \tag{A. 2}$$

和

$$F(x) > F(\bar{x}) \tag{A. 3}$$

注意我们不需要由约束 $k+1,\cdots,m$ 引起的任何限制;由于在 $\bar{x}$ 处,这些约束以严格的不等式的形式成立,根据连续性,对足够邻近于 $\bar{x}$ 的 x 它们将继续成立。

现在记作 $x=\bar{x}+\mathrm{d}x$,其中 $\mathrm{d}x$ 为无穷小量。因而我们想通过一阶泰勒近似取代(A. 2)式和(A. 3)式:

$$G_x^i(\bar{x})\mathrm{d}x \leqslant 0,\ i = 1,2,\cdots,k \tag{A. 4}$$

和

$$F_x(\bar{x})\mathrm{d}x > 0 \tag{A. 5}$$

如果 $G_x^i(\bar{x})$ 中的 k 行排放在一起形成的 $k \times n$ 矩阵 $G_x(\bar{x})$ 有秩 k,这些约束就是有效的。如果它的秩小于 k,它就有一个太大的零空间,即,当乘以此矩阵时太多的向量 $\mathrm{d}x$ 得到零。因而更多的 $\mathrm{d}x$ 满足线性近似(A. 4)式超过了 $\bar{x}+\mathrm{d}x$ 满足真正的约束(A. 2)式。结果,尽管 $\bar{x}$ 是最优解,一阶条件也可能失效。

我将用两个变量两个约束的例子证明这一点。假定目标函数为

$$F(x_1,\ x_2) = x_1$$

约束为

$$x_2 \geqslant x_1^3,\ 或\ G^1(x_1,\ x_2) \equiv x_1^3 - x_2 \leqslant 0$$

$$x_2 \leqslant -x_1^3,\ 或\ G^2(x_1,\ x_2) \equiv x_2 + x_2^3 \leqslant 0$$

图 A. 1 中阴影区域表示可行集,因而(0,0)为约束最大化问题的解是显然的。

在两个约束同时为紧的($k=2$)点处,约束的导数向量为

$$G_x^1(0,0) = (0,\ -1)\ 和\ G_x^2(0,0) = (0,1)$$

这两个行向量是线性相关的,导数矩阵的秩为 1,小于 k。线性近似(A. 4)式

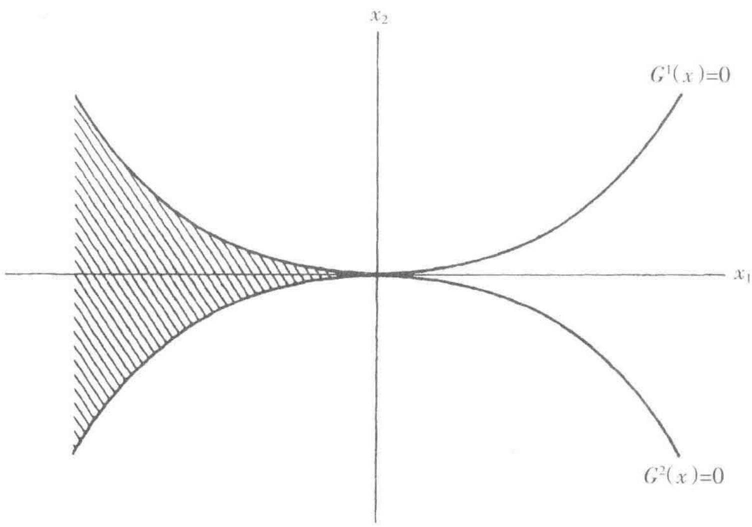

图 A.1 约束规格的失效

变为

$$-\mathrm{d}x_2 \leqslant 0 \text{ 和 } \mathrm{d}x_2 \leqslant 0, \text{即 } \mathrm{d}x_2 = 0$$

为整个横轴。事实上原点右边的点不是可行集的线性近似，但它们看起来是这些函数的线性近似中可行的偏离。

现在我们也可以看到先前在线性规划(例题 7.1)的背景下作出的评论的意义：在那里约束规格是不需要的。当约束已经是线性的时候，不需要找出对它们的线性近似，因而问题就不会出现了。

让我们继续讨论 $G_x(\bar{x})$的秩等于 k 的假设(约束规格)。图 A.2 表示了这种情形。它这样画是为了让所有的函数都随 x 递增，但那只是为了符合经济学的直觉；其他的形状没有区别。可行集和 $F(x)$的最优值的上等值集都以在它们的边界上标以影线的方式表示出来。每一个函数的导数向量也都表示出来了；且每一个都正交(垂直)于相应函数的轮廓线(contour curve)。

当 $\bar{x}$ 为最优解时，可行区域和目标函数的上等值集不应该相交。假定约束规格是有效的，(A.4)式和(A.5)式中的偏离所定义的它们的线性近似也不应该相交。为使此成立，导数向量 $F_x(\bar{x})$应该位于由约束函数导数的向量 $G_x^i(\bar{x})$形成的锥面上。从图中相反的方向看，如下的等价说法就更加容易理解：如果 $F_x(\bar{x})$在锥面外，经过 $\bar{x}$ 的 F 的轮廓线就会切入可行区域，因而一个有更高值的邻近点可以被找到，从而使 $\bar{x}$ 不可能为最优解。

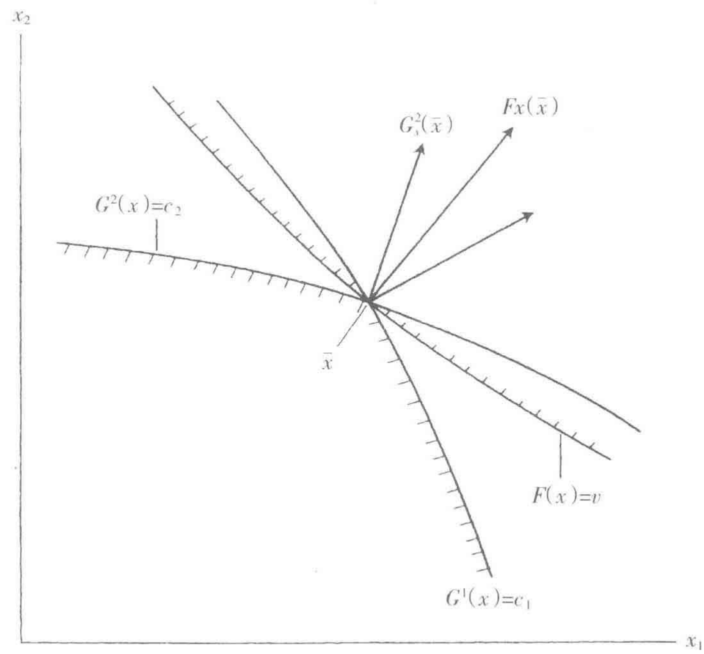

图 A.2 库恩—塔克定理

一个有限维锥面上的点是定义这个锥面的向量的非负线性组合。因而存在非负的数 $\lambda_1, \lambda_2, \cdots, \lambda_k$，使得

$$F_x(\bar{x}) = \sum_{i=1}^{k} \lambda_i G_x^i(\bar{x}) \tag{A.6}$$

通过定义 $\lambda_{k+1}, \cdots, \lambda_m$ 都为零将求和的范围扩大到 m 并无影响。扩大后的(A.6)式就是没有非负约束时由(3.7)式缩减成的条件 $L_x(\bar{x}, \lambda)=0$(记住非负约束被包括在不等式约束中，因而不起单独的作用)。鉴于 λ_i 被定义在松弛约束上的方式，互补松弛条件(3.10)式是满足的。这完成了证明的概略。

▶ 进一步阅读

一个更详细的证明可以在 E. Malinvaud, *Lectures on Microeconomic Theory*, Amsterdam: North-Holland, 2nd edn., 1985, J.-C. Milleron 作的附录中找到。Intriligator (1971，参见第 2 章的进一步阅读)第 4 章更加详尽地分析了凹的情形。

英汉名词对照表

Adverse selection　逆向选择
Arbitrage　套利
Bellman Equation　贝尔曼方程
Bellman's Principle of Optimality　贝尔曼最优化原理
Boundary point　边界点
Budget line　预算(约束)线
Capital　资本
Cobb-Douglas function　柯布一道格拉斯函数
CES function　不变替代弹性函数
Comparative statics　比较静态
Complementary slackness　互补松弛
Concave function　凹函数
Concave programming　凹规划
Constraint qualification　约束规格
Consumer behavior　消费者行为
Control variables　控制变量
Convex function　凸函数
Convex set　凸集
Corner solution　角点解
Cost curves　成本曲线
Cost function　成本函数
Decentralization　分散化
Demand functions　需求函数
　Compensated　补偿需求函数
　Derived　派生需求函数
Dynamic programming　动态规划
Entry-deterrence　进入阻挠
Envelope Theorem　包络定理
　Second-order　二阶(包络定理)
Expected utility　期望效用
Expenditure function　支出函数
First-order conditions　一阶条件
　Necessary　一阶必要条件
　Sufficient　一阶充分条件
Global maxima　全局最优
Hamiltonian　汉密尔顿函数
Indifference curve　无差异曲线
Indirect utility function　间接效用函数
Inequality constraints　不等式约束
Integration by parts　分部积分
Interior point　内点
Invisible hand　看不见的手
Kuhn-Tucker Theorem　库恩一塔克定理
Lagrange multipliers　拉格朗日乘子
Lagrange's method　拉格朗日方法
Langrangian　拉格朗日函数
Le Chatelier-Samuelson Principle　李・查特里一萨缪尔森原理

Linear programming　线性规划
Marginal cost　边际成本
Marginal rate of substation of transformation　边际转换替代率
Marginal utility of income　收入的边际效用
Maximum principle　最大化原理
Maximum Value Function　最大值函数
Mean-variance frontier　均值一方差前沿
Moral hazard　道德风险
Necessary conditions　必要条件
No-arbitrage condition　无套利条件
Non-negativity constraints　非负约束
Numeraire　计价物
Objective function　目标函数
Optimum growth　最优增长
Phase diagram　相位图
Portfolio choice　投资组合
Producer behavior　生产者行为
Production possibility frontier　生产可能性边界
Profit function　利润函数
Quadratic form　二次型
Quasi-concave function　拟凹函数
Quasi-concave programming　拟凹规划
Quasi-convex function　拟凸函数
Rationing　配给
Regular maxima　正则最大
Revealed preference　显示偏好
Risk-aversion　风险厌恶
Saving behavior　储蓄行为
Second-order conditions　二阶条件
　Necessary　二阶必要条件
　Sufficient　二阶充分条件
Search behavior　搜寻行为
Separation Theorem　分离定理
Shadow prices　影子价格
Slater condition　斯莱特条件
Slutsky-Hicks equation　斯拉茨基一希克斯等式
State variables　状态变量
Stationary point　稳定点
Substitution effect　替代效应
Transversality conditions　横截条件
Uncertainty　不确定性
Uniqueness　惟一性
Utility function　效用函数

译者后记

在这本广为流传的教科书和令人敬仰的作者面前，富有敬意的做法是保持沉默，以符合早先向作者所作的承诺：最大限度地保持原貌；并且向读者传递我们在与作者有限的交往过程中所感受到的一个经济学家的谦逊和正直。尽管如此，译者仍想对即将迈入经济学神圣殿堂的有志者作一点提醒：尽量读原著，读译本只是初学阶段和条件限制下的选择。

本书是译者为复旦大学经济学系99级本科生讲授数理经济学的主要参考教材。此译本的顺利出版得到了作者迪克西特教授、复旦大学中国经济研究中心张军先生和上海世纪出版集团何元龙先生的帮助，在此一并致谢。

中英文语言结构和表达习惯上的差异可能使得书中的一些表述尚未达到最理想的效果，译者水平所限，敬请读者指正。我们将尽量做到：第一，书中的专业术语与目前已经约定俗成的译法相统一；第二，不存在对内容的理解上的错误。

最后，我们希望在适当的时候及在获得迪克西特教授许可的情况下，可以向读者提供书中所有练习的参考答案。因为我们相信，书中富有启发性和足以举一反三的练习会让读者和我们一样体会到“干中学”的乐趣。

（第二版）

经济理论中的最优化方法

迪克西特（Avinash K. Dixit）教授是美国当代最负盛名的经济学家之一，现任普林斯顿大学经济学讲座教授。他1968年获麻省理工学院经济学博士学位，1977年当选计量经济学会（*Econometric Society*）院士，1992年荣膺美国文理学院（*American Arts and Sciences Academy*）院士，2002年任美国经济学会副会长。曾在加州伯克利大学、牛津大学任教。研究领域广泛，在博弈论、国际贸易理论、机制设计、产业组织、公共经济学与经济发展等多个领域有重要建树，近年来研究政策制订中的政治经济学。著作甚丰，除在《美国经济评论》(*American Economic Review*)、《经济学季刊》(*Quarterly Journal of Economics*)、《政治经济学杂志》(*Journal of Political Economy*)、《经济研究评论》(*Review of Economic Studies*)等顶级专业期刊发表多篇学术论文外，有多本著作被译成多国文字，影响广泛。本书《经济理论中的最优化方法》就是其中之一。自1976年初版以来，此书一直是微观经济学、宏观经济学的主流研究生教科书必引的参考书。