第 6 卷第 3 期 经济学（季刊） Vol. 6, No. 3
2007 年 4 月 China Economic Quarterly April, 2007

择优分配最优化原理及其应用

茅于轼*

摘　要　本文讨论在资源约束条件下如何分配使产出的效果极大。从最简单的两块土地分配化肥使增产的粮食极大化开始，在边际上不断调整，得出一般约束条件下的最优化方法，即拉格朗日乘子法。此优化原理更可以推广到动态过程，借用物理学中的捷降线问题，把落差看成资源，在逐步分配落差中使旅程的时间最短，进而把动态问题一般化，推导出变分法中的欧拉方程。择优分配原理可以将拉氏乘数法和欧拉方程都解释为资源的优化配置问题，从而给出了它们的经济学意义。本方法为一基本的优化原理，有广泛的应用可能。

关键词　资源配置，静态和动态优化原理，拉格朗日乘子法和欧拉方程的经济学解释

一、择优分配原理

在经济工作中经常遇到分配问题，所谓分配问题就是将供应有限的资源分给若干部门，以得到最佳的经济效果。这里的资源，可以是能源、资金、外汇、矿石、木材，甚至可以是人力、时间、土地等等。

为了判断一个分配方案是否优于另一个，必须有评价的标准。如果评价标准还没有建立起来，讨论优劣便没有任何意义。建立评价标准是经济学上的基本问题之一，本文不准备对这个问题加以讨论，而是假定已经具备了这个标准以后，讨论如何制定分配方案的方法。

如果只有一个部门参加分配，最佳分配量是使收益获得最大的量。如果收益对于分配的函数关系为已知，则通过微分方法可以求得切线为水平处的分配量 X^* 为最佳分配量，见图 1。如果资源量少于 X^*，则显然应将全部资源分配完，因为收益随着分配而增长。

从这里所举的例子我们可以看到，“收益”尽管在不同的情况下可以有不同的计量方法，例如铁矿石的分配的收益可以是产出的生铁量，电能的分配的收益可以是由于分配电能而得到的产值，资金的分配的收益可以是利润数等等，但所有这些收益均必须是纯收益，即扣除支出以后的收益。否则，对

* 天则经济研究所。作者通讯地址：北京市三里河南沙沟 2-1-1，100045；电话：(010)68522129；E-mail：maoysh@mail.ied.ac.cn. 本文为作者供职于中国能源研究会和铁道部科学研究院时完成的，最初发表于 1981 年 8 月，为一油印本。原文共 12 节两万余字，现在为了压缩篇幅，减为八节，1.3 万字。压缩掉的四节为：五、择优分配与宏观经济；六、择优分配原理的若干应用；七、可靠性的最优分配；十二、利用 Kukn Tucker 最优性条件求收益函数。现在发表的除了节、公式、图表的编号改变外，其余一字未改。

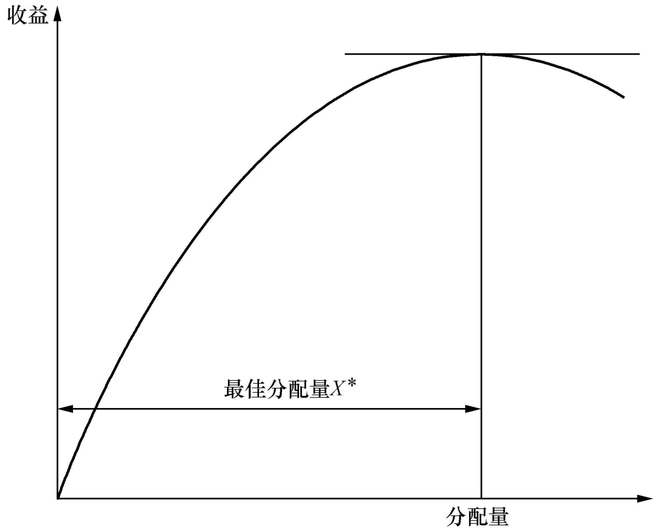

图 1 投入和产出——分配和收益

于亏损企业，分配量越多亏损越大，尽管毛收益也随着分配而增加。

当有几个部门参加分配时，显然不能用简单的微分方法求得最优解。考虑只有两个部门参加分配的情况。我们先定义收益 g 对于分配 x 的导数$\frac{\mathrm{d}g}{\mathrm{d}x}$，称为边际收益率，或边际产出率。对于这两个部门各有其收益函数 $g_1(x_1)$及 $g_2(x_2)$。对应于任何一个分配方案，两个部门各有其边际收益率$\frac{\mathrm{d}g_1}{\mathrm{d}x_1}$及$\frac{\mathrm{d}g_2}{\mathrm{d}x_2}$，见图 2。如果这两个边际产出率不相等，这个分配方案必有改善的可能。设想对边际收益率较大的第一部门增加微量分配 $\mathrm{d}x$，同时对边际收益率较小的第二部门减少分配 $\mathrm{d}x$，则在总分配额 $x_1+x_2=X$ 不变的条件下，两部门的总收益可以获得增量$\frac{\mathrm{d}g_1}{\mathrm{d}x_1}\mathrm{d}x-\frac{\mathrm{d}g_2}{\mathrm{d}x_2}\mathrm{d}x=\left(\frac{\mathrm{d}g_1}{\mathrm{d}x_1}-\frac{\mathrm{d}g_2}{\mathrm{d}x_2}\right)\mathrm{d}x$。可见当$\frac{\mathrm{d}g_1}{\mathrm{d}x_1}\neq\frac{\mathrm{d}g_2}{\mathrm{d}x_2}$时，分配方案可以继续改进，因此不是最优分配方案。仅当两部门具有同样的边际收益率时，分配方案才可能是最优的。

当有多于两个部门参加分配时，上面的结论同样成立。因为如果有一个最优分配方案，则任意两个部门之间均应具有同样的边际收益率。否则仅对这两个部门的分配加以调整就可以改善总收益量。任意两部门都有相同的边际收益率，只有当所有各部门的边际收益率相同时才有可能。因此，边际收益率相等是最优分配方案的必要条件。

现在考虑收益函数是凹函数时，通过逐步的微量调整以改善分配方案，最终会得到什么结果。图 1、图 2 中所绘的均系凹收益函数的情况，即对于同样的分配量 $\mathrm{d}x$，当分配数量 x 增多时，如由 $x=a$ 增加到 $x=b$，$a<b$ 收益的

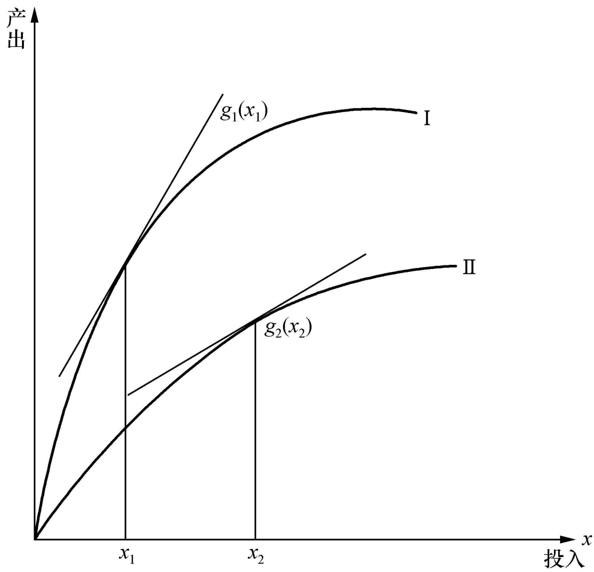

图 2　边际产出率

增量$\left.\frac{\mathrm{d}g}{\mathrm{d}x}\right|_{x=a}>\left.\frac{\mathrm{d}g}{\mathrm{d}x}\right|_{x=b}$，在改善分配效果时，是将$\frac{\mathrm{d}g}{\mathrm{d}x}$较大的部门增加分配量，$x$ 增加以后该部门的$\frac{\mathrm{d}g}{\mathrm{d}x}$必减小；而$\frac{\mathrm{d}g}{\mathrm{d}x}$较小的部门减少分配量后，$\frac{\mathrm{d}g}{\mathrm{d}x}$增大，这是凹函数的条件，改善分配的根据是$\frac{\mathrm{d}g_1}{\mathrm{d}x_1}\neq\frac{\mathrm{d}g_2}{\mathrm{d}x_2}$，经过调整改善之后各部门的$\frac{\mathrm{d}g}{\mathrm{d}x}$大的变小，小的变大。因此逐步调整的结果，最终将使所有部门的边际收益率均收敛于同一数值 λ。即$\frac{\mathrm{d}g_i}{\mathrm{d}x_i}=\lambda(i=1,2,\cdots,N)$，$N$ 为参加分配的部门数。

如果我们将总资源量 X 分成若干小份 $\mathrm{d}x$，然后一份一份选择效果最佳的部门逐步分配下去，这就是一个择优分配的过程。此时最先得到分配的部门必是在原点附近收益函数最陡的部门。以后的分配必定是使$\frac{\mathrm{d}g}{\mathrm{d}x}$最大的部门优先照顾，但因 $\mathrm{d}x$ 是微增量，因此结果一定是每一个分配方案都使各部门具有同样的边际收益率。X 从零开始一直到全部分配完毕，历经了无数的分配方案，每个方案都有$\frac{\mathrm{d}g_i}{\mathrm{d}x_i}=\lambda$ 的特点，而且它们都是最优分配方案，因此，等边际产出率同时又是最优分配方案的充分条件。因此可以得出结论：对凹的收益函数，各分配部门具有相等的边际产出率，是最优分配的充分必要条件。此时我们还假定了收益曲线是通过原点的，即分配为零时收益亦为零。

根据这个结论，我们很容易得到求各部门最优分配量的方法：任意确定

一个切线的斜率即$\frac{\mathrm{d}g}{\mathrm{d}x}$，对各部门的收益函数作平行切线，各切点所对应的横坐标 x_i 即为第 i 部门应该分得的资源量。如果 $\sum x_i \neq X$ ，可以改变$\frac{\mathrm{d}g}{\mathrm{d}x}$的值，当切线斜度变陡，由于函数的凹性，各个 x_i 均变小，反之 λ 变小时各个 x_i 均变大。最后总可试凑使 $\sum x_i = X$ 。这个试凑过程在电子计算机上很容易实现。另一个求得最优分配方案的方法是：由任意一个满足 $\sum x_i = X$ 的可行方案出发，不断优化这个方案，即减少$\frac{\mathrm{d}g}{\mathrm{d}x}$小的部门的分配量，改配给$\frac{\mathrm{d}g}{\mathrm{d}x}$大的部门。最后各部门的$\frac{\mathrm{d}g}{\mathrm{d}x}$均相等时即为最优分配方案。

如果收益函数是凸的，逐步改善分配的结果不会收敛于等边际收益率。因为对于凸函数，$\frac{\mathrm{d}g}{\mathrm{d}x}$大的部门增加分配以后，$\frac{\mathrm{d}g}{\mathrm{d}x}$变得更大，而小的变得更小。结果全部资源都将分配给一个部门。而且分配的结果与最初方案的选定有关，因为全部资源最终都将分配给最初方案中$\frac{\mathrm{d}g}{\mathrm{d}x}$最大的那个部门，而不同的初始分配方案下，各部门相对的$\frac{\mathrm{d}g}{\mathrm{d}x}$的大小次序不同。然而凸函数求极小值时，完全可以应用方才讨论的凹函数求极大值的理论。在经济问题中成本函数对于生产的各种要素（投入的资本、劳力、应用的物料、生产进度安排等）往往具有凸特性。使成本极小化就可应用成本对生产要素分配量的边际变化率相等的最优化条件，来确定生产要素的最佳分配。

二、择优分配的对偶原理

在各个部门的收益函数已经完全确定的情况下，一定数量的资源量 x 对应着一定的边际收益率λ，并且对应着一定的总收益 $\sum g_i(x_i^*) = G$ 。当 x 变化时，λ 和 G 也跟着变化。如果我们分别用 X、Λ、G 表示资源量、统一边际收益率和总收益量的实数集合，则三个集合之间存在着一一对应的映射关系，如图 3。X 到 Λ 的映射由 φ 表示，Λ 到 G 的映射由 ϕ 表示。一一对应的关系由收益函数的凹性所决定。举例说，如果资源量为 100 单位，对几个部门分配得到彼此统一的边际收益率为 0.3，此时总收益量为 70 个单位，如果资源增为 101 单位，由于函数的凹性，边际收益率必将减少，设减少为 0.29，总收益量将因投入的资源总量有所增加而增加，设增加为 71。这个例子说明了 X 中的任一个数都与 Λ 及 G 中唯一的某一个数相对应，反之 Λ（或 G）中的任一个数也与 X 和 G（或 X 和 Λ）中唯一的数相对应。

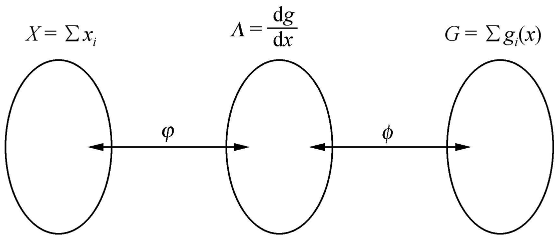

图 3　X、Λ、G 集合间的关系

这三个集合之间的对应关系，启示了择优分配的对偶原理，即某一个资源总量通过求出的统一边际收益率对应着一个最大的收益总量，同时某一收益总量通过统一的边际收益率对应着一个最小的资源总量。同一个资源总量由于分配方案的不同，可以有不同的收益总量，但最大的收益总量只有一个，即对应在 G 集合中的数，那是通过统一边际收益率求出的；同样，为达到一定的收益总量，可以耗用不同数量的资源，但最少的资源量只有一个，即对应在 X 集合中的数，那也是通过统一边际收益率求出的。

为什么 X 对应着 G 集合所需的最少资源需要量呢？这可用上面所举的例子来加以证明。为了获得 71 个单位的总收益，最少的资源需要量不可能大于 101 个单位，因为从 φ 和 ϕ 的对应我们确实只用了 101 个总资源获得了 71 个总收益。同时最少的资源需要量又不可能小于 101 个单位，因为如果只用了少于 101 个资源也能获得 71 个收益，那就违背了上升段凹函数的定义，即仅当 x 增加时方始有 λ 的减小，在上升段 $\lambda\left(=\frac{\mathrm{d}g}{\mathrm{d}x}\right)$无论如何减小还是大于零的，即 g 的变化方向与 x 相同，它随 x 的增加而增加，随 x 的减小而减小，现在 x 既然从 101 减小，g 也必然从 71 减小。可见最少的资源需用量既不可能大于 101 也不可能小于 101，于是 101 就是为得到 71 个总收益所必需的资源量。因此证明了 x 集合中的每一个数对应着 G 集合中每一个数所需的最少资源量。但这个结论仅当资源分配量对应着收益函数上升段的情况是对的，如分配量超过了图 1 中的 X^*，收益函数转入下降段，结论就不再适用。但实际经济工作中发生的问题中的收益函数都处于上升段。

三、统一边际收益率的含义

对于一定的资源量，对于一定的收益量，都有唯一的统一边际收益率与其对应。边际收益率的单位在能源分配问题中可能是每吨标准煤的工业产值；在铁矿石的分配问题中可能是每吨矿石产出的生铁量；在资金分配的问题中就是资金利润率，此时 λ 的含义是边际资本产出率，当它以年率的形式表示时，其倒数就是投资还本期。

由于 X、Λ、G 三个集合的对应关系，一定的边际资本产出率和一定的资本总额互相对应着。当资本充裕时对应的边际资本产出率较低，也就是投资还本期可以放宽。在市场调节的情况下，资本供应充足将使银行利率降低。因此银行利率实际上和边际资本产出率相联系。一个企业如果从银行贷款从事经营，则贷款利率不得高于企业的资本产出率，否则不但本金永远无法偿还，连利息也付不起。

基本建设投资的最优分配，就应该用边际资本产出率或投资还本期来衡量。对应于一定的资金总额有一定的投资还本期，基建项目中投资还本期超过这个统一投资还本期的都属缓建项目。短于统一还本期的都属应建项目。按照这样一个极简单的原则分配投资，必定得出投资的最优分配方案。当然这只是从经济效果来衡量。重大投资项目的决定还要考虑政治、军事等其他因素。

上面所说的统一的投资还本期应按照资金多寡的情况每年调整。资金紧张时统一投资还本期短，即仅对获利最丰的项目投资。资金充裕时还可对获利次佳的部门投资。

正因为边际资本产出率对应着资金总额，所以一个国家从国外取得贷款的数量应与贷款的利息相关。利息率越高贷款额越少。这两者的数量应通过专门研究客观地确定，并作为国际间金融往来的重要参考资料。显然，贷款额不仅与利率有关，而且与参加分配的部门数 N 有关。N 越大，即兴建的项目越多，在同样利率条件下贷款额可以增大，同时收益总量 G 也增大。开发生产门路、发展经济优势，就是增加 N 的具体措施。

收益总量对应着资源总量。在资金分配的情况下，总的利润额决定了应该准备的资金总额。如果资金不足此数，就不可能达到总的利润额；如果资金超过此数，利润虽能保证，但降低了资金利润率，保证不了投资效果。根据预定的目标估计需要的资源最低数量，是经济计划中的一项重要的任务。

统一边际收益率 λ 又标志着资源的匮缺程度。λ 越大说明这种资源短缺得越严重，因为 λ 大意味着投入资源可以得到的收益很多，或者意味着由于缺少资源损失的收益很多。如果用每吨煤影响多少产值来表示煤炭的短缺程度，这个数字应该是选择分配最后得到的统一边际收益率。如果煤矿附近发生着大量浪费，而在远离煤矿的工业区又严重地短缺，这两处的边际收益率相差很大，没有做到择优分配，就不存在统一的边际收益率，也就无从衡量煤的短缺程度。只有进行了择优分配，将浪费的煤改配给缺煤地区，才能最后估计出增加煤的产量是否还能增加产值，以及每吨煤能增加的产值是多少。

统一的边际收益率可用来衡量稀缺物资的经济价值。对于不受资源限制的产品，只要投入资本和劳力就可制造出来，但诸如石油，某些有色金属、热带作物等，由于受到资源及其他限制，它们的生产带有垄断性质。因此当对它们的需求量超过资源条件所能提供的数量时，它们的价格往往超过一般

条件下确定的数值。在资本主义生产条件下这就是广义的级差地租发生的原因。在社会主义条件下如何正确度量这类稀缺物资的价值，并据此在技术经济比较中作出判断，合理地分配和使用，必须借助于这类物资对各有关部门分配最终得到的统一边际收益率来计算。这个边际收益率就是该项物资的影子价格，以后我们将讨论它的数学定义和含义。

社会主义企业由于自然条件不同，在同样经营管理水平及同样努力程度下，产品成本有颇大的区别，这种苦乐不均的现象必须用征收资源税的方法来纠正。一般地说，统一边际收益率可用以确定在正确制定可比成本时所要考虑的资源税的税率。从这里我们也可以看到，在技术经济比较中对稀缺物资（如石油制品）的价值强调用生产成本来计算是不对的，应该用成本加上利润和资源税，或者直接按择优分配最终达到的统一边际收益率再加适当调整来计算。

边际收益率是两个微增量的比，用它来说明资源稀缺程度时，必须注意随着资源供应的增加，边际收益率将逐渐降低。说明能源匮缺就不能用每亿吨能源增产若干亿元，因为这样表示远远不是微增量之比。

四、经济活动中收益函数的凹性

择优分配结果得到统一边际收益率，仅当收益函数为凹函数时才有这个结论。事实上经济活动中的收益函数是否具有凹性？其原因是什么？有无例外情况？这是有待进一步讨论的。

首先，我们相信，在一般情况下收益函数具有凹性，即收益的增长比投入的增长慢。其根本的原因是收益取决于一系列的因素，其中的任一个因素受到某种限制不能跟着变化时，投资资源的效果将受它的限制，收益就逐渐减小。农作物的收成可随施肥量而增加，但当土地面积一定时，收成因受土地的限制而不可能与施肥量成正比增加。工厂的产量与能源供应量有关，但由于受工厂设备、劳力、空间等的限制，产量也不可能与能源供应量成正比地增加。空间、时间、人力、资源等的限制都是客观存在的，它们最终都会在经济活动中以某种形式表现出来。

其次，经济关系中的逐级分配现象也使收益函数出现凹性。如果 x_1、x_2、x_3 分别表示农、轻、重的投资分配，则对农业来说，x_1 又可分配于粮食、林业、科研、教育等。$g_1(x_1)$所表示的农业投资的收益将是由上述各属于农业的各部门的收益所构成。于是将 x_2 再对粮食、林业、科研等农业内的各部门进行择优分配，先分配给收益最显著的部门，其次分配给收益稍次的部门。可见 x_1 增大时，由于最后分配的部门必是收益最差的部门，故 $g_1(x_1)$的增加在 x_1 越大时就越小，说明了 $g_1(x_1)$必然是凹的。不仅农业内部有逐级分配，工业、交通业亦均如此。甚至小到一个具体产品也有投资的逐级分配问题，

例如花费一笔投资用于提高电视机某一零件的可靠性，这笔投资一定首先用于可靠性最敏感的某道工序。投资数很大时开始照顾到提高可靠性比较困难的工序。这个例子说明了逐级分配现象带有普遍性。每一级的择优分配都促使上一级的收益函数呈凹性。

但是收益函数的凹性却不是没有例外的。例如由于协作的关系两个人完成的工作量大于一个人工作量的两倍。规模很大时，由于社会分工的发展，产出量的增加亦可能大于投入量的增加。由于经营管理方法的改进（例如采用了择优分配的方法等），或者由于工艺技术的改进，甚至可以在投入量不变的条件下增加收益。这些都会导致收益函数呈凸性。从人类生产力发展的总过程来看，改进生产组织和技术进步始终是最活跃的因素，但是在某一个特定时期生产方式相对地比较定型的时候，收益函数就表现出凹性。

在资本主义社会，由于企业和个人追逐利润，并由于市场调节的作用，分配过程会不断地优化改进。在资金分配的情况下，表现为资金利润率趋于统一的自发过程。这种现象说明了资金分配的收益函数确实是凹的，因为对于凸的收益函数，分配优化的结果不会收敛于统一边际收益率，这在前面已经证明过了。然而由于资本主义生产缺乏统一的计划调度，它的优化过程是在利益互相对立的条件下进行的。技术垄断和情报保密妨碍了最优条件的达到，实际的分配结果是在最优分配方案的周围振荡。由于实力巨大的垄断组织的出现，或者单纯由于不利条件的随机组合，这种振荡的振幅有时可能是很大的，因此可能造成严重的社会后果。在社会主义的生产条件下，从理论上讲我们可以不必依赖于市场和竞争的作用，只要通过计划的最优化就可实现最佳的经济效果。但事实上由于原始资料不可能十分详尽周全，数据的传递不可能毫无延误，最优化问题的数学计划也远未臻完善，我们还不得不依靠市场调节的反馈作用来补充计划的不足。计划最优化的工作做得越差，我们将越多地依赖于市场调节，计划最优化做得越是详尽严密，市场调节的作用越小。最优化的计划调节在整个经济调节中所占的比例越高，说明经济活动中的随机现象越少。如果可以用熵来度量经济活动中的随机化程度，那么不断完善和加强计划调节的作用，使经济活动目标更明确，配合更协调，效果更完美，实际上就是一个不断减熵的过程。因此社会主义生产的高度发展是组织最优化作用逐渐加强的过程。相反，如果既不依靠计划的最优化，又不利用市场调节的反馈机制，经济工作必然发生严重的混乱，其后果将是灾难性的。回顾世界经济日益商品化依赖的历史，许多重大的经济危机就是由于上述两种现象同时存在而引起的，尽管发生这两个现象的原因可能是经济本身的，也可能是政治性或军事性的重大动荡。

五、择优分配原理和 Lagrange 乘数法

前面讨论的资源分配问题和可靠性分配问题等，都是在约束条件下求极值的问题。这类问题普遍地可以用 Lagrange 乘数法求得解决。但前文我们并未依靠 Lagrange 乘数法，同样求得了解决方法，而且这种方法的经济意义十分明显。本节中我们将进一步证明，用择优分配原理可以导得 Lagrange 乘数法的一般结果。

在资源分配中 Lagrange 乘数法的提法是：

求能使目标函数：$G(x_1, x_2, \cdots, x_n) = \sum_{i=1}^{N} g_i(x_i)$ 取极大值的各个 $x_i (i = 1, 2, \cdots, N)$，同时满足约束条件 $\sum_{i=1}^{N} x_i = X$，用 Lagrange 乘数法求解，作函数 $L(x_1, x_2, \cdots, x_N) = \sum_{i=1}^{N} g_i(x_i) + \lambda(X - \sum_{i=1}^{N} x_i)$，令 L 函数取极值，其必要条件是 $\frac{\partial L}{\partial x_i} = 0$，即 $\frac{\partial g_i}{\partial x_i} = \lambda (i = 1, 2, \cdots, N)$。

前面我们对择优分配作经济意义讨论的时候，中心的思想是：资源在各部门间互相作微量调整时（但这种调整不能破坏规定的约束条件，即 $\sum_{i=1}^{N} x_i = X$），目标函数可以不断地改善。但在最优分配时此种调整已不能使目标函数得到改善。这是最优分配所应满足的必要条件。但资源分配量的微量调整的方向可正可负，即既可能从甲部门给乙部门 dx，也可以从乙部门给甲部门 dx。两种调整方法都不能使目标函数增大，故这两种调整的结果只能是使目标函数的改变为零。换言之，从甲部门减少 dx 的分配量如果使目标函数改变 $-dG$，则对乙部门增加 dx 的分配量必定使目标函数改变 $+dG$，二者相加为零，也就是目标函数的边际收益率大小相等且符号相反。这个条件对任意两个分配部门均应成立。当约束条件不是简单的 $\sum x_i = X$ 时，上述推论依旧成立，所不同的只是某一 x_i 减少 dx 时，另一 x_k 不是增加 dx，而是要根据约束条件的具体要求增加。

根据以上论证，我们得到极其重要的择优分配的结果：边际收益率相等为最优化的必要条件。具体表述如下：一个约束条件下求最优化的问题，可以视为对各个变量的分配问题。在达到最优分配时，任意两个变量在遵循约束条件的情况下作相应的微量调整时，对目标函数的边际收益率应大小相等，符号相反。下面我们用这个最优化条件来证明 Lagrange 乘数法。

目标函数 $G(x_1, x_2, \cdots, x_N)$ 取极大，　　(1)

约束条件 $R(x_1, x_2, \cdots, x_N) = 0.$　　(2)

令任一个 x_j 改变 $\mathrm{d}x_j$ 另一个 x_k 改变 $\mathrm{d}x_k$，其余的 x_i 均不变，约束函数 R 的改变量在忽略 $\mathrm{d}x$ 高阶项时可表示为：

$$\frac{\partial R}{\partial x_j}\mathrm{d}x_j + \frac{\partial R}{\partial x_k}\mathrm{d}x_k = 0, \tag{3}$$

（3）式必须等于零，否则约束条件被破坏。

从（3）式得

$$\mathrm{d}x_j = -\frac{\dfrac{\partial R}{\partial x_k}}{\dfrac{\partial R}{\partial x_j}}\mathrm{d}x_k, \tag{4}$$

（4）式说明，为了遵循约束条件，x_k 改变 $\mathrm{d}x_k$ 时 x_j 应该变多少。

x_j、x_k 作上述调整时，它们对于目标函数的边际收益率为 $\frac{\partial G}{\partial x_j}\mathrm{d}x_j$ 及 $\frac{\partial G}{\partial x_k}\mathrm{d}x_k$，二者大小相等，符号相反，故

$$\frac{\partial G}{\partial x_j}\mathrm{d}x_j = -\frac{\partial G}{\partial x_k}\mathrm{d}x_k, \tag{5}$$

将（4）式代入，得

$$\frac{\dfrac{\partial G}{\partial x_j}}{\dfrac{\partial R}{\partial x_j}} = \frac{\dfrac{\partial G}{\partial x_k}}{\dfrac{\partial R}{\partial x_k}} = \lambda. \tag{6}$$

x_j、x_k 为任意选取的，故有

$$\frac{\partial G}{\partial x_i} = \lambda\frac{\partial R}{\partial x_i}, \quad (i = 1, 2, \cdots, N), \tag{7}$$

也可写为

$$\frac{\partial}{\partial x_i}(G - \lambda R) = 0, \quad (i = 1, 2, \cdots, N). \tag{8}$$

（8）式即为 Lagrange 条件。它共有 N 个，加上约束条件（2），共可解出 N 个 x_i 及 λ。但解得的结果为最优分配的必要条件，仅当目标函数为凹时同时又为充分条件。或者目标函数要求为极小，同时它又是凸函数时（8）式也是充分条件。

（6）式又可写成

$$\frac{\partial G}{\partial R} = \lambda. \tag{9}$$

此式明确地表明了乘数 λ 的意义。它表明约束条件微量变化时，目标函数的增量与约束条件增量的比，或者目标函数的值依赖于约束条件的程度，故称之为敏度系数。在经济问题中 λ 有其经济含义。如 R 表明某种资源的限制条件，G 表示经济收益时，则 λ 表示该种资源的影子价格，即每增加一个单位的 R 能增加多少收益。已如前述，垄断性资源的价格不取决于成本加平均利润，而是与因使用此种资源获得之收益，即与影子价格有关。

在更一般的约束极值问题中，约束条件不止一个，而有 N 个，即

$$R_i(x_1, x_2, \cdots, x_N) = 0, \quad i = 1, 2, \cdots, M, \quad \text{且 } M < N. \tag{10}$$

对每一个约束条件，均作同上的推导，用矩阵表示可得：

$$\frac{\partial G}{\partial R} = \begin{bmatrix} \dfrac{\partial G}{\partial R_1} \\ \dfrac{\partial G}{\partial R_2} \\ \vdots \\ \vdots \\ \vdots \\ \dfrac{\partial G}{\partial R_M} \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \vdots \\ \vdots \\ \lambda_M \end{bmatrix} = \lambda, \tag{11}$$

令

$$\frac{\partial G}{\partial x} = \begin{bmatrix} \dfrac{\partial R_1}{\partial x_1} & \dfrac{\partial R_2}{\partial x_1} & \cdots & \dfrac{\partial R_M}{\partial x_1} \\ \dfrac{\partial R_1}{\partial x_2} & \cdots & \cdots & \dfrac{\partial R_M}{\partial x_2} \\ \vdots & \vdots & \vdots & \vdots \\ \dfrac{\partial R_1}{\partial x_N} & \cdots & \cdots & \dfrac{\partial R_M}{\partial x_N} \end{bmatrix}, \tag{12}$$

(11) 式右乘 (12) 式，得：

$$\frac{\partial R}{\partial x}\frac{\partial G}{\partial R} = \frac{\partial R}{\partial x}\lambda, \tag{13}$$

即

$$\left.\begin{array}{l} \dfrac{\partial G}{\partial R_1}\cdot\dfrac{\partial R_1}{\partial x_1} + \dfrac{\partial G}{\partial R_2}\cdot\dfrac{\partial R_2}{\partial x_1} + \cdots + \dfrac{\partial G}{\partial R_M}\cdot\dfrac{\partial R_M}{\partial x_1} \\ \qquad = \lambda_1\dfrac{\partial R_1}{\partial x_1} + \lambda_2\dfrac{\partial R_2}{\partial x_1} + \cdots + \lambda_M\dfrac{\partial R_M}{\partial x_1} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \dfrac{\partial G}{\partial R_1}\cdot\dfrac{\partial R_1}{\partial x_N} + \dfrac{\partial G}{\partial R_2}\cdot\dfrac{\partial R_2}{\partial x_N} + \cdots + \dfrac{\partial G}{\partial R_M}\cdot\dfrac{\partial R_M}{\partial x_N} \\ \qquad = \lambda_1\dfrac{\partial R_1}{\partial x_N} + \lambda_2\dfrac{\partial R_2}{\partial x_N} + \cdots + \lambda_M\dfrac{\partial R_M}{\partial x_N} \end{array}\right\}, \tag{14}$$

或

$$\left.\begin{array}{l}\dfrac{\partial G}{\partial x_1}+\dfrac{\partial G}{\partial x_1}+\cdots+\dfrac{\partial G}{\partial x_1}=\lambda_1\dfrac{\partial R_1}{\partial x_1}+\lambda_2\dfrac{\partial R_2}{\partial x_1}+\cdots+\lambda_M\dfrac{\partial R_M}{\partial x_1}\\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\ \dfrac{\partial G}{\partial x_N}+\dfrac{\partial G}{\partial x_N}+\cdots+\dfrac{\partial G}{\partial x_N}=\lambda_1\dfrac{\partial R_1}{\partial x_N}+\lambda_2\dfrac{\partial R_2}{\partial x_N}+\cdots+\lambda_M\dfrac{\partial R_M}{\partial x_N}\end{array}\right\}, \tag{15}$$

即

$$\left.\begin{array}{l}M\dfrac{\partial G}{\partial x_1}=\lambda_1\dfrac{\partial R_1}{\partial x_1}+\lambda_2\dfrac{\partial R_2}{\partial x_1}+\cdots+\lambda_M\dfrac{\partial R_M}{\partial x_1}\\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\ M\dfrac{\partial G}{\partial x_N}=\lambda_1\dfrac{\partial R_1}{\partial x_N}+\lambda_2\dfrac{\partial R_2}{\partial x_N}+\cdots+\lambda_M\dfrac{\partial R_M}{\partial x_N}\end{array}\right\}, \tag{16}$$

$$\frac{\partial G}{\partial x}=\begin{pmatrix}\dfrac{\partial R_1}{\partial x_1} & \dfrac{\partial R_2}{\partial x_1} & \cdots & \dfrac{\partial R_M}{\partial x_1}\\ \dfrac{\partial R_1}{\partial x_2} & \cdots & \cdots & \cdots\\ \vdots & \vdots & \vdots & \vdots\\ \dfrac{\partial R_1}{\partial x_N} & \dfrac{\partial R_2}{\partial x_N} & \cdots & \dfrac{\partial R_M}{\partial x_N}\end{pmatrix}\begin{pmatrix}\lambda_1'\\ \lambda_2'\\ \vdots\\ \lambda_M'\end{pmatrix}, \tag{17}$$

此处

$$\lambda_i'=\frac{1}{M}\lambda_i.$$

再写成代数式：

$$\frac{\partial G}{\partial x_i}=\sum_{j=1}^{M}\lambda_j'\frac{\partial R_j}{\partial x_j},\quad (i=1,2,\cdots,N). \tag{18}$$

（18）式即为多个约束条件的 Lagrange 最优化条件。至此我们完成了全部证明。上面的讨论说明，用择优分配原理可以证明 Lagrange 乘数法；反之，用 Lagrange 乘数法也可证明择优分配所得到的结果。但择优分配原理对最优解必须满足的条件作出了经济意义上的解释，那就是目标函数对于各个变量必须有相同的边际收益率，同时变量的变化不得违背约束条件的限制。

边际收益率相等是一个很重要的概念。在没有约束条件时，$\lambda=0$，此时我们可以假想存在着一个不起作用的约束条件，或者说目标函数对于约束条件的敏度系数为零。在 $\lambda=0$ 时最优解的必要条件转化为$\dfrac{\partial G}{\partial x}=0$，这就是普通微分方法求极值所要满足的必要条件。

回顾图 3，X 集合就是约束条件中一个系数的变化所组成的，它与目标函数值 G 集合之间通过敏度系数 Λ 集合存在着对应关系。图 3 所讨论的是约束

条件中的常数项变化而形成了 X 集合，事实上约束条件中任一系数变化都会有一个集合，而且都会通过与 Λ 的对应关系而最终与目标函数 G 集合有对应关系。Λ 即为边际收益率。可见边际收益率对应着确定的目标函数值，不论约束条件是如何规定的。

六、择优分配原理在连续分配问题中的应用

前面讨论的分配问题，分配的部门是数得清的几个。当分配的部门无穷增多，每个部门的分配量趋近于零时，就成为连续分配问题。择优分配原理也可以应用于连续分配问题，而且可以得到更普遍的结果。连续分配问题的解在工程技术中有很广泛的应用。下面通过三个例子来说明：

1. 火车或汽车省功运行的速度分配规律。

已知火车运行的阻力为速度的凸单调递增函数。给定火车从甲站行驶到乙站的时间为 T，问火车司机应如何控制速度，使行驶中克服阻力所做的功为极小。

现在的目标函数是 $\min\int_0^S R(v)\mathrm{d}s$，约束条件是 $\int_0^S \frac{\mathrm{d}s}{v} = T$。

式中 S 为两站间的距离。欲求速度与距离的关系 $V=V(s)$。

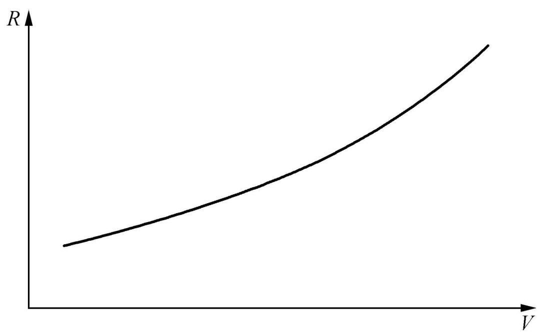

图 4　火车阻力和速度的关系

现将 S 等分为 N 小段，每段长均为 ΔS。于是现在的问题是如何将 T 分给这 N 个小段。每段分给的时间为 ΔT_i，则 $\sum_{i=1}^{N}\Delta T_i = T$。每段的速度 $V_i = \frac{\Delta S}{\Delta T_i}$，分定了 ΔT_i，就对应一定的 V_i。从图 4 知道，由于各段的距离均同为 ΔS，故每一段中 R 与 ΔT_i 的关系完全是同一条曲线。相对于图 2 现有 I、II、……许多条线，但这些线都重合为一条线。凸函数极小化的问题，完全可以用凹函数极大化的原理。在最优分配时边际收益率均相同，即平行切线的切点的横坐标对应最优分配量。现所有的 $g(x)$ 线均重合，平行切线的切点为同

一个点，说明各个小段应用同一个速度运行，即分配同样的 $\Delta T_i=\Delta T$。以上的讨论当 $N\to\infty$ 时同样成立，此时即转变为连续的分配问题。

2. 使成本为最低的生产进度安排

与上例相似的一个问题是：已知生产成本为生产速度的凸函数。现要求在一定的时间 T 内生产一定量的产品量 N，问应如何安排生产进度，可使总的产品成本为极小。

本问题与上例是完全对应的。也是将时间 T 分配给 N 件产品。每件产品的成本与生产进度的关系就是类似于图 4 的凸函数曲线，而总成本为每件产品的成本之和。根据上例的分析我们得知，成本最低的进度安排是均匀生产，任何突击和会战都会使成本增加。无论成本与生产速度之间函数关系的具体形状如何，只要是凸函数，上述的结论总是成立的。当然，这里假定了其他一切生产条件均不变。

严格地说，本问题与上例是有区别的。上例中阻力只是速度的函数，与加速度无关。而生产成本则与生产的加速度有关，即重新调整生产进度时将发生额外的费用。因此上例完全用静态的分配模型求解是可以的，而本例则理应建立一个包括生产调整费用的模型。但本例所得的结论显然是正确的，因为均匀生产不需用任何调整所发生的费用。

3. 锅炉中燃烧率的最优分配问题

众所周知，锅炉的吸收热量是燃烧率（以每单位时间单位炉篦面积上燃掉燃料的重量计量）的凹函数。在整个炉篦面积上由于各点通风分布的不同，离开传热面的距离不同，而且产生的燃气的路径也不同，所以同样燃烧一定量的煤，在不同的位置燃烧，锅炉吸收的热量是不同的，换言之，锅炉所能吸收的热量 G 是燃烧率 u 及炉篦坐标位置（x,y）的函数：

$$g=g(u,x,y),$$

现在的问题是：在一定数量的总燃烧量 U

$$U=\iint u(x,y)\mathrm{d}x\mathrm{d}y \tag{19}$$

的约束条件下，求能使锅炉吸收的总热量 G

$$G=\iint g(u,x,y)\mathrm{d}x\mathrm{d}y \tag{20}$$

为极大的燃烧率的分布函数 $U=u(x,y)$。

本例所讨论的问题与前面讨论的最优分配问题有一基本的不同点。前面讨论的问题中收益函数 g 仅为分配量的函数，而本问题中收益函数 g 不仅为分配量 u 的函数，而且是 u 所在位置的坐标（x,y）的函数，因此 g 函数已不能用如图 1 中的平面曲线来表示，而是一个三维空间中的一个曲面。

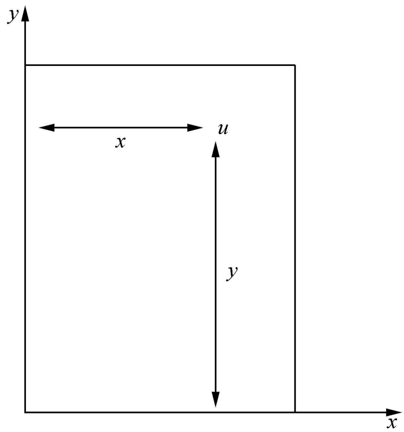

图 5　燃烧率的最优分布

这类问题可以用具有等周条件的变分问题的模型来解决，而且结果是解一个退化为一般代数方程的微分方程。但我们基于择优分配原理的讨论可知，最优解必须满足边际产出率相等的条件。现在的边际产出率是$\frac{\partial g(u,x,y)}{\partial u}$，故最优解必须满足的条件是

$$\frac{\partial g(u,x,y)}{\partial u}=\lambda. \tag{21}$$

λ 的值要根据约束条件（19）式来确定。u 值越大，对应的 λ 越小。U 即相当于前面讨论的总资源量的限制。

（21）式有它的经济含义。它说明在实现最优分配时，在炉篦的任一块小面积 $\mathrm{d}x\mathrm{d}y$ 上投入少许煤量 $\mathrm{d}u$ 都能使锅炉吸收的热量增加同样的数量λ。因此如果有某一小块$\frac{\partial g}{\partial u}$稍大一些的话，就应将该坐标处投的煤 $\mathrm{d}u$ 改投到该处。如此不断调整，由于 $g(u,x,y)$函数的凹性，最后必须使得各处的燃烧率的边际产出率相等。现 $g(u,x,y)$函数是已知的，故（21）式就是问题的解。这种形式的解要求资源 u 的分布是一个随坐标变化的函数，这与前面讨论的结果不同。进一步推广使择优分配原理的应用有更多方面的可能。

在连续分配问题中应用择优分配的概念，主要用于解决工程技术方面的问题。

七、动态问题中择优分配原理的应用

动态问题和静态问题的根本区别在于：动态问题的最优方案不但要考虑到决策所得的收益，而且与规定的始终点的条件有关，当始终点的条件改变时，最优方案跟着改变。一个动态问题往往可以在空间上或实践上分成若干

阶段。每一小段采取不同的策略，不但影响到该小段中的收益，而且影响该小段终点的状态，而且因此又影响到下一小段的方案的选取。以国民经济规划为例，始点条件是规划开始时的经济参数，如固定资产数、人口等；规划的目的可能是五年计划内人民消费总额为极大；同时要求期末的固定资产、人口等达到某一预定目标。要解出最优的积累比例、人口的增长率等。显然，如果期末要求达到的固定资产越多，则规划期内人民消费总额将减少，但这对下一个规划期生活的改善创造了条件。因此积累率的选取不但要满足目标函数极大，还应保证期末的经济参数。这一类经济规划问题往往还有一系列约束条件，如资源的限制、部门间经济联系的限制等。经济规划是一个非常复杂的问题。下面我们将讨论动态问题中最简单的情况，即最速降线，或称捷降线的问题。这是古典变分法中的一个典型问题。它要求解出从始点 $P_0(x_0, y_0)$ 到终点 $P_N(x_N, y_N)$ 的一条曲线的方程，使质点沿该曲线依靠重力从始点滑行至终点所花的时间为最少，忽略摩擦阻力。

连接 P_0P_N 的直线是两点间最短的线，可是质点沿直线运行时间未必最短，因为质点未能在开始时获得较大的加速度。如果开始时先垂直向下，加速度虽然最大，行经的距离却加长。因此所求的线可能是介乎上述两种情况之间的某一条曲线。捷降线的原理在铁路编组站驼峰纵断面的设计中有所应用，但要增加一个相邻车辆车钩夹角的限制条件。

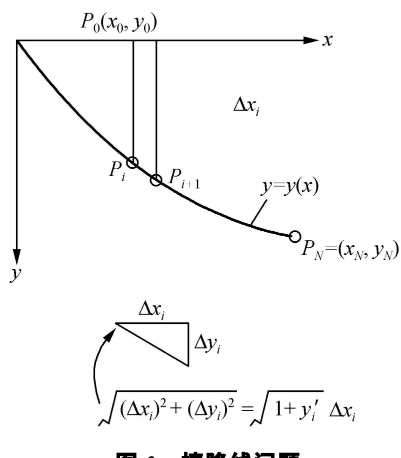

图 6 捷降线问题

从图 6 中设 $y=y(x)$ 为所求的捷降线，曲线经过 P_i 及 P_{i+1} 两点。根据物理学定理，质点到达 P_1 点时的速度为 $\sqrt{2gy_i}$，y_i 为 P_1 点的纵坐标。按照积分学的原理，质点从 P_0 到 P_N 所花的时间 G 为：

$$G=\int_{x_0}^{x_N}\frac{\sqrt{1+y'^2}}{\sqrt{2gy}}\mathrm{d}x, \tag{22}$$

或者写成和式

$$G=\lim_{\Delta x_i\to 0}\sum_{i=0}^{N-1}\frac{\sqrt{1+y_i'^2}}{\sqrt{2g\sum\limits_{j=0}^{i}\Delta y_j}}\Delta x_i,\tag{23}$$

更一般地考虑被积函数为一个一般的函数 $g(y,y')$，则

$$G=\lim_{\Delta x_i\to 0}\sum_{i=0}^{N-1}g\left(y_0+\sum_{j=0}^{i}\Delta y_j,y_i'\right)\Delta x_i\tag{24}$$

从择优分配的概念来考虑捷降线问题，我们很自然地将 y_N-y_0 视为持有的资源量，因为 $y_N-y_0=\Delta Y$ 为全部落差，质点的速度是依靠落差而获得的。现在的问题是如何将落差 ΔY 分配给各个小段，使收益函数 G 取极值。一旦对于各个小段 Δx_i 分得的落差 Δy_i 被确定，整个曲线 $y=y(x)$就得到了。由于总的落差是有限的，于是在 G 取得极值的同时，必须满足约束条件：

$$\sum_{i=0}^{N-1}\Delta y_i=y_N-y_0=\Delta Y,\tag{25}$$

同时 Δy_i 又与 y_i' 存在着关系：

$$\Delta y_i=y_i'\cdot\Delta x_i,\quad (i=0、1、\cdots、N-1).\tag{26}$$

（26）式也是必须满足的约束条件。

（24）、（25）、（26）式已构成了从择优分配概念描述的捷降线的变分问题。但捷降线问题还不是变分问题的普遍的提法，因为（23）或（24）式的被积函数或 g 函数中未出现 x，即收益与分配部门所在的坐标位置 x 无关。在更一般的变分问题中被积函数或收益函数同时与 x，y，y'有关。即应取极值的目标函数 G 为

$$G=\int_{x_0}^{x_N}g(x,y,y')\mathrm{d}x,\tag{27}$$

边界条件为

$$\text{当}\quad \left.\begin{aligned}x=x_0,y=y_0\\x=x_N,y=y_N\end{aligned}\right\},\tag{28}$$

x_0,y_0,x_N,y_N 均为给定的数。

现在可以将变分法所要解决的问题从分配与收益的意义来理解，并用择优分配的原理来解决。$g(x,y,y')\mathrm{d}x$ 为位于 x 的无穷多个待分配部门之一的收益函数。y 为累计投入的资源量，在问题中新投入的总资源量限制为 y_N-y_0，待分配的部门的坐标位置在 x_0 至 x_N 之间。

将目标函数写成和式

$$G = \lim_{\Delta xi \to 0} \sum_{i=0}^{N-1} g\Big(x_0 + \sum_{j=0}^{i} \Delta x_j, y_0 + \sum_{j=0}^{i} \Delta y_j, y_i'\Big)\Delta x_i. \tag{29}$$

约束条件(25)、(26)式依旧成立。

于是我们将无约束的变分问题转化为具有约束条件的广义的资源分配问题。

设按问题的物理意义或经济意义,在所讨论的 x 和 y 的变化范围内这个和式是收敛的。现应用 Lagrange 乘数法,构成新函数 H:

$$H = \lim_{\Delta x_i \to 0}\Big[\sum_{i=0}^{N-1} g\Big(x_0 + \sum_{j=0}^{i} \Delta x_j, y_0 + \sum_{j=0}^{i} \Delta y_j, y_i'\Big)\Delta x_i + \sum_{i=0}^{N-1}\lambda_i(\Delta y_i - y_i'\Delta x_i) + \Delta\Big(\sum_{i=0}^{N-1}\Delta y_i - \Delta y\Big)\Big]. \tag{30}$$

此处的每一个 λ_i 均不同,它们随着 i 或者说随着 x 而变,是 x 的函数。

为使 H 取得极值,可以变化各个 Δy_i 来实现,也可以变化 y_i' 来实现,尽管由于(26)式的联系,二者实质是相同的,但在数学上确有不同的形式。我们将先后从这两方面来加以讨论。假定 $g(x,y,y')$对于 y 及 y'的偏导数存在,且(30)式为一致收敛。现在首先令变化 y_i 使 H 达到极值,则有:

$$\frac{\partial H}{\partial y_i'} = \lim_{\Delta x_i \to 0}\Big[g_{y'}\Big(x_0 + \sum_{0}^{i}\Delta x_j, y_0 + \sum_{0}^{i}\Delta y_j, y_i'\Big)\Delta x_i - \lambda_i \Delta x_i\Big] = 0,$$

或

$$\lim_{\Delta x_i \to 0} g_{y'}\Big(x_0 + \sum_{0}^{i}\Delta x_j, y_0 + \sum_{0}^{i}\Delta y_j, y_i'\Big) - \lambda_i = 0, \quad (i = 0,1,2,\cdots,N-1), \tag{31}$$

(31)式对于任何一个中间点 x_i 处均成立,因此可写为:

$$g_{y'}(x,y,y') = \lambda, \tag{32}$$

此处 λ 为 x 的函数。

其次再来变化 Δy_i,使 H 取极值,则有:

$$\frac{\partial H}{\partial \Delta y_k} = \lim_{\Delta x_i \to 0}\sum_{i=k}^{N-1} g_k\Big(x_0 + \sum_{j=0}^{i}\Delta x_j, y_0 + \sum_{j=0}^{i}\Delta y_j, y_i'\Big)\Delta x_i + \lambda_k + \Lambda = 0,$$
$$(k = 0,1,2,\cdots,N-1). \tag{33}$$

因为 H 中对 g 求和时足标等于及大于 k 的项内均包含有 Δy_k。(33)共有 N 个等式。取极限得:

$$\int_{x_k}^{x_N} g_y(x,y,y')\mathrm{d}x + \lambda_k + \Lambda = 0, \tag{34}$$

假设被积函数 g_y 的原函数为 $R(x,y,y')$，则（34）式可写成：

$$R(x_N, y_N, y_N') - R(x_k, y_k, y_k') + \lambda_k + \Lambda = 0. \tag{35}$$

（35）式中第一项为常数，第二项、第三项对于 $(x_0 < x_k < x_N)$ 中的任一点 x_k 均成立。将（35）式对 x 求导，注意到 Λ 为常数，得：

$$-g_y(x,y,y') + \frac{\mathrm{d}\lambda}{\mathrm{d}x} = 0, \tag{36}$$

将（32）式对 x 求导：

$$\frac{\mathrm{d}}{\mathrm{d}x} g_{y'}(x,y,y') = \frac{\mathrm{d}\lambda}{\mathrm{d}x}, \tag{37}$$

将（36）式代入（37）式即得欧拉方程

$$g_y - \frac{\mathrm{d}}{\mathrm{d}x} g_{y'} = 0. \tag{38}$$

至此，我们按择优分配的概念同样得出了欧拉方程。择优分配原理不但可用于解决静态问题，亦可用于解决动态问题。资源的限制可以从广义来理解，等周条件、边界条件等都可以转化作资源限制来处理，于是这个原理就可灵活地应用到许多提法不同的问题。择优分配对不同的最优化问题可以作出共同的解释，我们认为它是最优化方法的一个普遍基础，所以称它为择优分配量最优化原理。

（32）式表明了收益函数 y' 的边际收益率 $g_{y'}$ 不再是一个常数，而是一个 x 的函数 λ。

上面的推导过程也提示了求欧拉方程数值解的一个方法。（34）式可写为：

$$\int_{x_0}^{x_N} g_y(x,y,y')\mathrm{d}x - \int_{x_0}^{x_k} g_y(x,y,y')\mathrm{d}x + \lambda_k + \Lambda = 0, \tag{39}$$

于是

$$\lambda_k = \int_{x_0}^{x_k} g_y(x,y,y')\mathrm{d}x - \int_{x_0}^{x_N} g_y(x,y,y')\mathrm{d}x - \Lambda, \tag{40}$$

但 $\int_{x_0}^{x_k} g_y(x,y,y')\mathrm{d}x$ 为常数，可与 Λ 合并，令

$$\int_{x_0}^{x_k} g_y(x,y,y')\mathrm{d}x + \Lambda = \Lambda', \tag{41}$$

再将（40）式写成极限求和式，得：

$$\lambda_k = \lim_{\Delta x_i \to 0} \sum_{i=0}^{k} g_y\Big(x_0 + \sum_{j=0}^{i} \Delta x_j, y_0 + \sum_{j=0}^{i} \Delta y_j, y_i'\Big) \Delta x_i - \Lambda',$$
$$k = 0,1,2,\cdots,N-1 \tag{42}$$

将（42）式与（26）式、（31）式联立，当设定一个 Λ'之后，对于每一个 i 均可求出 Δy_i，$\Delta y_i'$ 及λ_k。逐步解到 $i=N-1$ 时，每一个 Δy_i 均可求得。如果各个 Δy_i 能满足（25）式 $\sum_{i=0}^{N-1} \Delta y_i = \Delta Y$，则说明设定的 Λ'是合适的，否则可另设 Λ'的数值重新逐步求解，直到（25）式满足为止。实际计算时可绘出 Λ'与 $\sum_{i=0}^{N-1} \Delta y_i$ 的关系曲线，从曲线上取 $\sum_{i=0}^{N-1} \Delta y_i = \Delta Y$ 处Λ'的值，可减少试凑的计算工作量。对应每一个 Λ'值计算得到的每一条 $y=y(x)$曲线都是问题的最优解，所不同的只是终点条件不同。

八、欧拉方程的经济含义

在捷降线的研究中将落差 ΔY 视作资源，将落差在横坐标中进行分配，结果可以得出捷降线的解。一般地，将纵坐标的改变视作资源用择优分配的原理推导欧拉方程时，正如前面表述的，还要借助于 Lagrange 乘数法。事实上，根据择优分配的结果——边际收益率相等为最优化的必要条件，我们可以直接得出欧拉方程。这个边际收益率相等的具体表述，已在第五节中给出。现在我们将它应用到动态问题的最优解，并以此直接得出变分法中的欧拉方程。

现讨论最基本的变分问题，它仍由（27）及（28）式所规定。设在图7中 P_0 及 P_N 为给定的两个端点，而 $y=y^*(x)$为能使（27）式达到极限的最优解，$P_i(x_i, y_i)$为极限曲线上任一点。目标函数（27）式可写成和式：

$$G = \lim_{\Delta x \to 0} \sum g(x, y, y') \Delta x. \tag{43}$$

任一通过 P_0、P_N 两点的曲线 $y=y(x)$上的任一点 P，有坐标 x 及 y，并有该处的 $y' = \frac{\mathrm{d}y}{\mathrm{d}x}$。将这三个数值代入 $g(x, y, y')\Delta x$ 可得出构成G 的一个微单元。但仅在极值曲线 $y=y^*(x)$上才能使（43）式的和为极大（小）。

参考图7，在极值曲线 $y=y^*(x)$上任一点 P_i 处，设保留该点的坐标(x_i, y_i)不动，而令 y_i' 作微量改变。此时在 P_i 点处的 G 的微单元 $g(x_i, y_i, y_i')\Delta x$ 将产生相应的改变。当忽略高阶项时其改变量为：

$$g_{y'}(x_i, y_i, y_i')\Delta x \Delta y_i'. \tag{44}$$

这就是由于在 P_i 点处曲线的方向作了 $\Delta y_i'$ 的改变而对目标函数G 的影

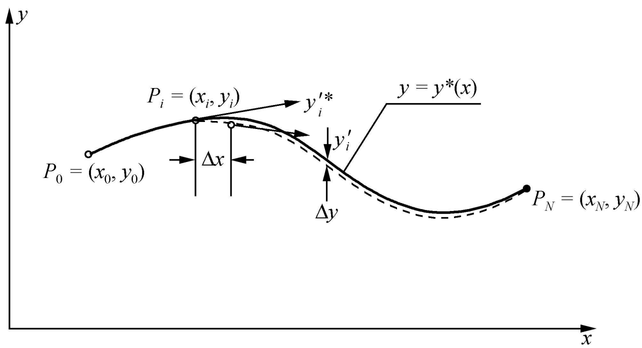

图 7　从边际收益概念推导欧拉方程

响，或者说，这就是由于曲线切线方向改变而得到的边际收入。但由于在 P_i 处的方向改变，使得以后曲线的纵坐标跟着发生变化。原来横坐标递增 Δx，纵坐标改变 Δy，现在 Δy 本身将改变 $\Delta^2 y$，即 y 的二阶差分。故微单元的改变为：

$$g_y(x,y,y')\Delta x\Delta^2 y.$$

从 P 点以后每个微单元都有这样的改变量，故一共使目标函数 G 改变了

$$\int_{x_i}^{x_N} g_y(x,y,y')\mathrm{d}x\Delta^2 y. \tag{45}$$

这是由于曲线在 P_i 处方向改变而不得不引起的一笔边际损失，因为约束条件是曲线最后必须通过 $P_N(x_N,y_N)$。

如果 $y=y^*(x)$是极值曲线，则我们可以断言，边际收入必定等于边际损失。如果二者不等，则通过这样的改变将有净的收益，亦即 $y=y^*(x)$还可以改善，它不是极值曲线。因此有

$$g_{y'}(x_i,y_i,y_i')\Delta x\Delta y_i' + \int_{x_i}^{x_N} g_y(x,y,y')\mathrm{d}x\Delta^2 y = 0. \tag{46}$$

上述的极值条件在任一点处 x_i 均应满足，因此特定的点 $P(x_i,y_i)$可换为曲线上的任意点 $P(x,y)$，又因 $\Delta y'=\dfrac{\Delta^2 y}{\Delta x}$，故有：

$$g_{y'}(x,y,y')\Delta^2 y + \int_{x}^{x_N} g_y(x,y,y')\mathrm{d}x\Delta^2 y = 0. \tag{47}$$

对 x 微分，并注意到 x_N 为常数，得

$$\frac{\mathrm{d}}{\mathrm{d}x}g_{y'}(x,y,y') - g_y(x,y,y') = 0, \tag{48}$$

即为欧拉方程。

从上述的推导过程我们可以察知，欧拉方程是有它的经济含义的。它说明，极值曲线具有这样的特性，在曲线的任一点稍微改变其方向所得的边际收益（损失），将恰为以后曲线位置改变所发生的边际损失（收益）所抵消；曲线无论在哪一点，朝哪个方向移动都不能改善目标函数的值。

参 考 文 献

Bellman, R., and S. Dreyfus, *Applied Dynamic Programming*. Princeton, N. J.: Princeton University Press, 1962.

Principle of Optimal Allocation and Its Applications

YUSHI MAO
(*Unirule Institute of Economics*)

Abstract The problem dealt with in this paper is how to allocate resources to achieve maximum effectiveness. Starting from a simple problem, to distribute chemical fertilizer on two plots of land to obtain maximum yield by continuous improvements on the margin, the method of Lagrange Multiplier can be deducted. This idea can be extended to dynamic problems. Borrowing the problem of Brachistochrone, to find a path descending along which a movable particle is to reach a given point in a vertical plane in the shortest time by the force of gravity, the same idea of resources allocation is applied to dynamic cases, where the height difference is treated as resources. Through further generalization, the Euler differential equation in calculus of variation can be derived. The Principle of Optimal Allocation is a fundamental principle of optimization and explains the economic meanings of both the Lagrange Multiplier method and the Euler equation in calculus of variation. This Principle has a wide area of applications.

JEL Classification B41, C02, C61