

最优化问题的数学模型一般可以用约束集 X 和目标函数 f 进行表示. 集合 X 包含所有可用的决策 x , 函数 $f(x)$ 将 X 的元素映射到实数集上, 表示决策 x 带来的成本损失. 我们试图寻找一个最优的决策, 即 $x^* \in X$, 并且满足

$$f(x^*) \leq f(x), \quad \forall x \in X.$$

本书假定 x 是一个 n 维向量, 即 x 是一个由实数构成的 n 元数组 $(x_1, \dots, x_n)$, 因此约束集 X 是 n 维欧氏空间 $\mathbb{R}^n$ 的子集.

最优化问题是非常广泛的, 包含了很多具有不同结构的重要问题. 本书主要研究非线性规划问题. 我们将提供处理这些问题的方法, 并讨论这些问题与其他优化问题之间的联系.

最优化问题可以被分为连续优化问题和离散优化问题. 连续优化问题中, 约束集 X 包含无穷多个元素, 并且具有“连续性”特征. 典型的连续优化问题包含无约束优化问题, 即 $X = \mathbb{R}^n$, 以及 X 是被某些等式和不等式限定的约束优化问题. 一般而言, 可利用微积分和凸函数性质来分析连续优化问题.

离散优化问题就是非连续问题, 通常其约束集 X 是有限集. 典型的离散优化问题包含调度、路径规划以及匹配问题. 一类重要的离散问题是整数规划, 在这种问题中约束集的变量的取值是某一范围内的整数 (比如 0 或者 1). 离散优化问题一般利用组合数学和离散数学进行分析, 同时会利用一些特殊的方法以及一些和连续优化问题相关的方法.

非线性规划是指目标函数 f 是非线性函数, 或者约束集 X 是由非线性的等式和不等式给定的优化问题. 这种问题属于连续优化问题. 一些其他优化问题, 虽然也有各自的特点, 但是往往与非线性规划问题具有密切的关系.

线性规划问题是指 f 是线性函数, 并且 X 是由线性不等式约束给定的多面体的一类优化问题. 这种问题具有连续优化问题的很多特性, 同时它还具有组合优化的结构特性: 根据多面体表示的基本定理 [附录 B 命题 B.21(d)], 线性规划的最优解可以在多面体 X 的某个极点 (极点数目是有限的) 上取得. 因此, 寻找最优解的过程是在某个有限集中进行搜索. 实际上, 最流行的线性规划问题的求解方法, 即单纯形方法, 就是基于这种思想设计的. 然而, 我们也注意到其他重要的线性规划的方法, 比如内点法 (将在 2.6 节, 4.1 节

以及 4.2 节中讨论) 和第 5 章、第 6 章中的其他一些对偶方法, 都是基于线性规划的连续结构和非线性规划的思想提出的.

另一类同时具有连续优化问题和离散优化问题特点的优化问题是网络优化问题. 该类问题的约束集 X 是由包含节点和有向弧的图定义的 $\mathbb{R}^n$ 中的多面体. 该约束集的明显特征是其极点都具有整数分量, 而这对于通常的多面体是不成立的. 因此, 很多重要的组合优化或者整数规划问题, 比如匹配问题和最短路径问题, 都可以利用连续网络优化的方法求解.

本书重点研究非线性规划问题的连续性特性以及相应的数学分析方法, 然而, 我们也会涉及存在某些离散特性的问题. 我们将会对线性规划与非线性规划的联系进行讨论, 比如内点法和多面体的凸性 (见 2.5 节、2.6 节、4.1 节、4.4 节、B.3 节和 B.4 节). 2.1 节和 5.5 节将对网络优化问题的连续特性和离散特性进行讨论 (对此类问题的更加深入的讨论请参见本书作者网络优化专著 [Ber98]). 最后, 将对求解整数规划和组合优化的主要方法进行讨论, 比如分枝定界方法和拉格朗日松弛方法, 这些方法依赖于对偶性质以及连续优化子问题的解 (见 5.5 节和 6.3 节).

在本章中, 我们首先考虑如下的非线性规划问题 (其中 $X = \mathbb{R}^n$)

$$\begin{aligned} & \text{minimize } f(x) \\ & \text{subject to } x \in \mathbb{R}^n. \end{aligned} \tag{UP}$$

在后面的章节中, 我们将研究 X 是 $\mathbb{R}^n$ 的子集的优化问题, 其中 X 可由等式和不等式约束给出, 在大多数情况下, 假定 f 是一个连续可微的函数, 并且是二阶连续可微的, f 的一阶和二阶导数在最优解的特征, 即充分必要条件中发挥了重要作用, 这是 1.1 节的主要内容. 一阶和二阶导数对于计算近似最优解也是非常重要的, 1.2 节~ 1.8 节将讨论这方面的若干算法和理论. 1.9 节将前几节的方法运用到了求解含有离散时间动态系统的最优控制问题中. 虽然本章主要研究的是无约束优化问题, 但是本章的内容的很多思想也是全书其余内容的重要基础.

1.1 最优性条件

虽然下面对非线性规划的最优性条件的证明比较复杂, 但是其主要思想是容易理解的. 因此, 在这一节中, 我们首先简要地对最优性条件的主要思想进行阐述, 然后在下一节再给出严格的结论及其证明.

局部最小值点与全局最小值点

向量 x^* 被称为是 f 的一个无约束局部最小值点, 是指该点处的函数值不大于其邻域内所有点的函数值, 即存在 $\varepsilon > 0$ 满足

$$f(x^*) \leq f(x), \quad \forall x \text{ 使得 } \|x - x^*\| < \varepsilon.$$

(除非有特别说明, 否则将采用欧氏范数 $\|x\| = \sqrt{x'x}$. 附录 A 对本书中用到的数学记号和术语进行了详细介绍.)

向量 x^* 被称为是 f 的一个无约束的全局最小值点, 是指该点的函数值不大于其他所有点的函数值, 即满足

$$f(x^*) \leq f(x), \quad \forall x \in \mathbb{R}^n.$$

无约束优化问题的局部或者全局最小值点 x^* 被称作是严格的, 如果相应的不等式对于 $x \neq x^*$ 都是严格成立的. 图 1.1.1 给出以上定义的图示.

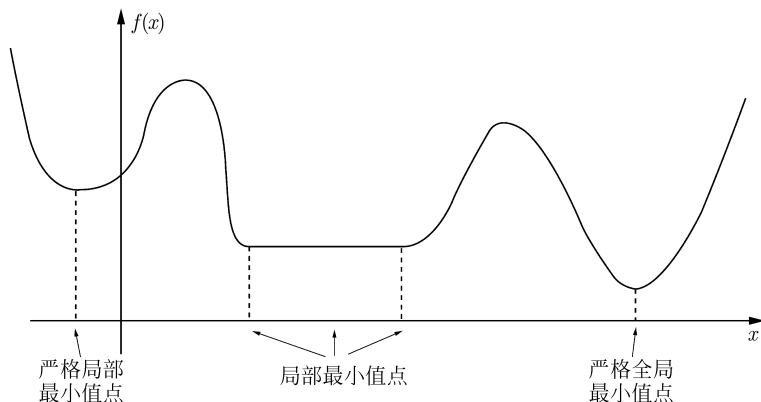


图 1.1.1 一维函数无约束优化问题的局部与全局最小值点.

局部和全局最小值点的定义还可以扩展到当 f 是定义在 $\mathbb{R}^n$ 的子集 X 上的函数的情况. 即 x^* 是 f 在 X 上的一个局部最小点值, 如果 $x^* \in X$ 并且存在 $\varepsilon > 0$, 对于所有满足 $\|x - x^*\| < \varepsilon$ 的 $x \in X$, 都有 $f(x^*) \leq f(x)$. 对于 f 在 X 上的全局最小值点和严格最小值点的定义可以类似地给出.

局部和全局的最大值点的定义也是类似的, 即如果 x^* 是 $-f$ 的无约束局部(全局)最小值点, 那么 x^* 是 f 的无约束局部(全局)最大值点.

最优性的必要条件

如果目标函数可微, 那么就可利用梯度和泰勒展开去比较某个向量的函数值及其邻域内的向量函数值的大小关系. 特别地, 我们考虑在给定向量 x^* 上给予一个微小扰动 Δx , 从而利用一阶近似得到目标函数的变化量为

$$f(x^* + \Delta x) - f(x^*) \approx \nabla f(x^*)' \Delta x.$$

二阶展开后, 目标函数的变化量为

$$f(x^* + \Delta x) - f(x^*) \approx \nabla f(x^*)' \Delta x + \frac{1}{2} \Delta x' \nabla^2 f(x^*) \Delta x.$$

如果 x^* 是一个无约束的局部最小值点, 那么由微小扰动 Δx 产生的目标函数的一阶变化满足

$$\nabla f(x^*)' \Delta x = \sum_{i=1}^n \frac{\partial f(x^*)}{\partial x_i} \Delta x_i \geq 0.$$

对于所有坐标 $i = 1, \dots, n$, 取 Δx 为单位坐标向量 (某个分量为 1, 其余均为 0) 的正数或负数倍, 我们分别得到 $\partial f(x^*)/\partial x_i \geq 0$ 和 $\partial f(x^*)/\partial x_i \leq 0$. 因此, 可得到最优性的必要条件

$$\nabla f(x^*) = 0.$$

[Fermat 在 1637 年的短论文 “Methodus as Disquirendam Maximam et Minimam” 中最早提出了该条件, 但是并没有给出证明.] 这个条件将在下节命题 1.1.1 中严格证明.

我们期望由微小扰动 Δx 产生的目标函数的二阶变化也是非负的, 即

$$\nabla f(x^*)' \Delta x + \frac{1}{2} \Delta x' \nabla^2 f(x^*) \Delta x \geq 0.$$

由 $\nabla f(x^*)' \Delta x = 0$, 得到

$$\Delta x' \nabla^2 f(x^*) \Delta x \geq 0,$$

由此可知

$$\nabla^2 f(x^*): \text{半正定}.$$

我们将在命题 1.1.1 中证明这个必要条件. 附录 A 中给出了正定和半正定矩阵的定义和性质.

在后面的讨论中, 我们将满足条件 $\nabla f(x^*) = 0$ 的向量 x^* 称为驻点.

凸目标函数的情形

附录 B 中给出了凸性的相关定义和性质, 在作者的专著 [BNO03] 中对此进行了更加深入的讨论. 凸性在非线性规划中发挥着很重要的作用, 这是因为当目标函数 f 为凸函数时, 局部最小值点和全局最小值点是一致的. 图 1.1.2 给出了图示, 附录 B 的命题 B.10 给出了严格证明.

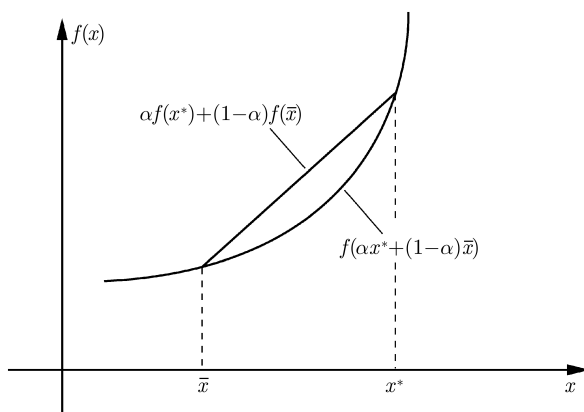


图 1.1.2 凸函数的局部最小值点也是全局最小值点的图示.

假设 f 是凸函数, 且令 x^* 和 $\bar{x}$ 满足 $f(\bar{x}) < f(x^*)$. 利用凸性, 对于所有 $\alpha \in (0, 1)$, 有

$$f(\alpha x^* + (1 - \alpha)\bar{x}) \leq \alpha f(x^*) + (1 - \alpha)f(\bar{x}) < f(x^*).$$

所以, 除了 x^* , 对连接 x^* 和 $\bar{x}$ 的线段上的所有点, 都有 f 的函数值严格小于 $f(x^*)$. 因此, 如果 x^* 不是一个全局最小值点 [从而存在 $\bar{x}$, 满足 $f(\bar{x}) < f(x^*)$], 那么 x^* 也不可能是局部最小值点.

另一个重要的事实是, 如果 f 是凸函数, 那么一阶条件 $\nabla f(x^*) = 0$ 也是最优性的充分条件. 这将在下节中的命题 1.1.2 中证明. 该证明过程是根据凸函数的基本性质: 利用梯度的性质, 对函数在 x^* 处进行线性近似, 即 $f(x^*) + \nabla f(x^*)'(x - x^*)$, 因为该式是 $f(x)$ 的下界估计, 所以如果 $\nabla f(x^*) = 0$, 那么对于所有的 x , 都有 $f(x^*) \leq f(x)$ (见附录 B 中命题 B.3).

图 1.1.3 说明如果 f 不是凸函数, 那么一阶和二阶必要条件可能无法保证 x^* 是局部最优解 (即局部最小值点).

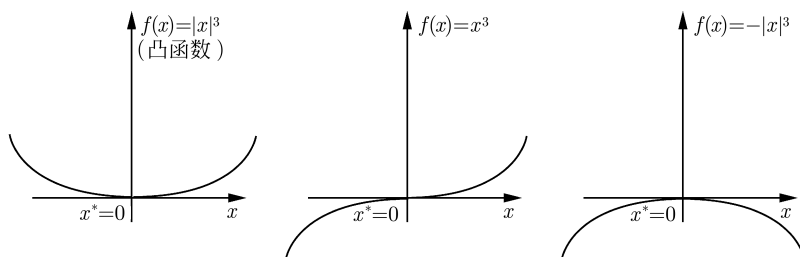


图 1.1.3 单变量函数优化问题的一阶最优性必要条件, 即零斜率 $[\nabla f(x^*) = 0]$ 以及二阶最优性必要条件, 即非负曲率 $[\nabla^2 f(x^*) \geq 0]$ 的图示. 一阶条件不仅在局部最小值点处满足, 同样在局部最大值以及“反射”点处满足 (如中间图所示). 在一些情况中, 比如 $f(x) = x^3$ 和 $f(x) = -|x|^3$, 二阶条件同样也在局部最大值点和反射点处满足. 如果 f 是凸函数, 条件 $\nabla f(x^*) = 0$ 是 x^* 为全局最优解的充分必要条件.

最优性的充分条件

假设向量 x^* 满足一阶最优性必要条件

$$\nabla f(x^*) = 0, \quad (1.1)$$

并且满足二阶最优性必要条件的增强形式

$$\nabla^2 f(x^*) : \text{正定}, \quad (1.2)$$

(即 Hessian 矩阵是正定的, 而非半正定), 那么对所有的 $\Delta x \neq 0$, 我们有

$$\Delta x' \nabla^2 f(x^*) \Delta x > 0,$$

说明在 x^* 处, 由微小非零扰动 Δx 产生的 f 的二阶变化为正. 所以, 如果 x^* 出现微小变动, f 会严格增大, 从而说明以上的条件式 (1.1) 和式 (1.2) 是 x^* 为局部最优解的充分条件. 这将在命题 1.1.3 中得到证明.

不满足充分条件式 (1.1) 和式 (1.2) 的局部最小值点被称为 奇异点; 否则被称为 非奇异点. 奇异的局部最小值点是比较难处理的, 这是因为首先 f 不满足凸性性质时, 其最优性不能通过容易验证的充分条件来保证. 其次, 在它们的邻域内, 常用的优化算法的收敛速度会变慢且 (或) 不稳定.

二次目标函数

考虑二次目标函数

$$f(x) = \frac{1}{2}x'Qx - b'x,$$

这里 Q 是一个对称的 $n \times n$ 矩阵, b 是 $\mathbb{R}^n$ 中的向量. 如果 x^* 是 f 的一个局部最小值点, 那么由最优性必要条件, 我们有

$$\nabla f(x^*) = Qx^* - b = 0, \quad \nabla^2 f(x^*) = Q : \text{半正定}.$$

所以如果 Q 不是半正定的, f 可能没有局部最小值点. 如果 Q 半正定, 那么 f 为凸函数 [附录 B 命题 B.4(d)], 所以任何满足一阶条件 $\nabla f(x^*) = Qx^* - b = 0$ 的向量 x^* 都是 f 的局部最小值点. 另外如果 Q 是奇异的, 方程 $\nabla f(x^*) = Qx^* - b = 0$ 可能无解. 然而如果 Q 是正定的 (因此由附录 A 的命题 A.20 知其可逆), 方程 $Qx^* - b = 0$ 存在唯一解, 且向量 $x^* = Q^{-1}b$ 是唯一的全局最小值点. 这和下面即将给出的命题 1.1.2(a) 是一致的. 该命题断言严格凸函数最多只能有一个全局最小值点 [f 是严格凸, 当且仅当 Q 是正定矩阵; 附录 B 中命题 B.4(d)]. 图 1.1.4 给出了各种特殊情况的图示.

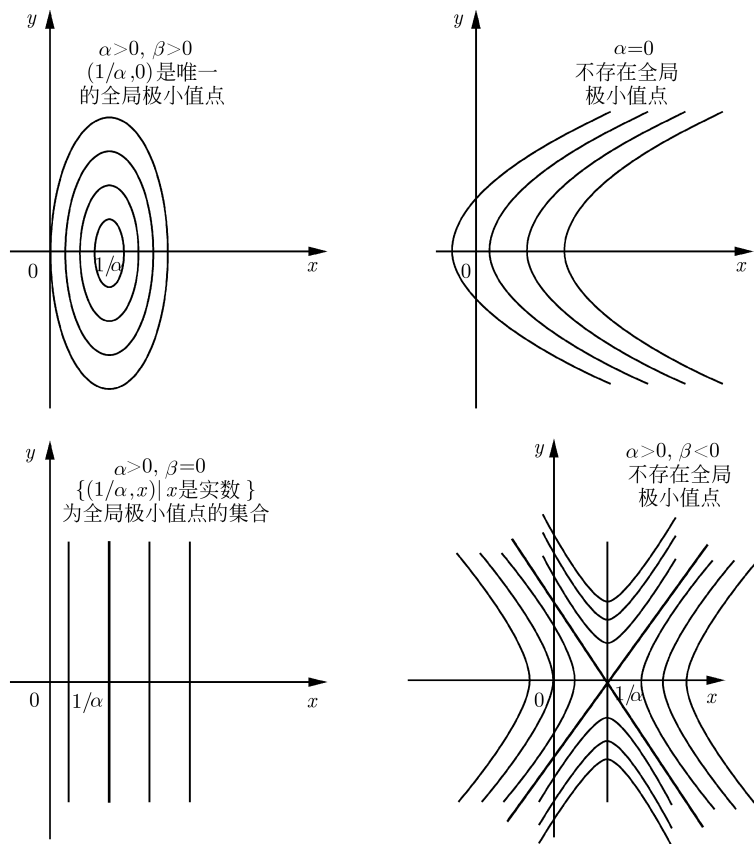


图 1.1.4 二次目标函数 $f: \mathbb{R}^2 \mapsto \mathbb{R}$ 的等值线图示. 该目标函数由

$$f(x, y) = \frac{1}{2}(\alpha x^2 + \beta y^2) - x$$

给出, 其中 α 和 β 的取值情况如各图所示.

二次目标函数在非线性规划中很重要, 这种函数在实际应用中很广泛. 从泰勒展开式

$$f(x) = f(x^*) + \frac{1}{2}(x - x^*)' \nabla^2 f(x^*)(x - x^*) + o(\|x - x^*\|^2),$$

可以看到一个非二次的目标函数在非奇异点的局部最小值点 x^* [$\nabla^2 f(x^*)$: 正定] 附近, 可以用二次函数来近似. 这说明我们可以利用正定二次函数进行很多演算分析和试验, 并且可将这些结论扩展到更一般的目标函数中去. 然而对于局部最小值点附近 Hessian 矩阵不存在或者为奇异的情形, 高于二阶的泰勒展开项是不能忽略的, 并且基于二次目标函数的演算分析很可能会导致错误的结论.

最优解的存在性

在很多情况下, 判断在集合 X 上 f 函数的全局最小值点是否存在是很有用的. 一般来说, 极小值点可能并不存在. 比如, 对于标量函数 $f(x) = x$ 和 $f(x) = e^x$, 在整个实数集上就没有全局极小值点. 前者当 x 趋于 $-\infty$ 时函数值递减并趋于 $-\infty$, 而后者随着 x 趋于 $-\infty$ 时函数值趋于 0, 但是函数值总为正. 当 x 在 X 上变动时, $f(x)$ 的取值集合

$$\{f(x) \mid x \in X\},$$

有两个可能

(1) 集合 $\{f(x) \mid x \in X\}$ 有下界, 即存在标量 M , 对所有的 $x \in X$, 有 $M \leq f(x)$. 在这种情况下, $\{f(x) \mid x \in X\}$ 最大的下界是一个实数, 记为 $\inf_{x \in X} f(x)$. 比如 $\inf_{x \in \mathbb{R}} e^x = 0$ 且 $\inf_{x < 0} e^x = 0$.

(2) 集合 $\{f(x) \mid x \in X\}$ 无下界, 即该集合包含任意小的实数. 在这种情况下, 记

$$\inf_{x \in X} f(x) = -\infty.$$

如果 f 是连续函数, 且 X 是 $\mathbb{R}^n$ 的紧子集, 那么可以保证至少存在一个全局极小值点, 这是由于 Weierstrass 定理(见附录 A 命题 A.8). 命题 A.8 也说明, 如果 f 是连续的, X 是闭集, 并且 f 在 X 上是强制性的, 即对任意序列 $\{x^k\} \subset X$, $\|x^k\| \rightarrow \infty$, $f(x^k) \rightarrow \infty$, 那么 f 也至少存在一个全局最小值点. 2.1.2 节对最优解的存在性问题进行了更深入的探讨, 并且 X 可以是无界的.

为什么需要最优性条件?

在分析优化问题时, 最优性条件无疑是最根本的. 在实践中, 最优性条件在很多情况下也都是非常重要的, 然而在某些情况下, 这一点可能不是很明显.

解决优化问题时, 利用最优性条件的最直接的方法如下: 首先, 找到所有满足一阶必要条件 $\nabla f(x) = 0$ 的解, 然后 (如果不知道 f 是否为凸函数), 在每一个解处检查二阶必要条件 ($\nabla^2 f$ 为半正定), 删除不满足条件的解, 最后对于剩下的解, 检查 $\nabla^2 f$ 是否为正定的, 如果某个解正定, 那么这个解就是一个严格的局部最小值点.

另一个方法是找到所有满足必要条件的解, 并且把其中目标函数值最小的点作为全局最小值点. 此时全局最小值点的存在性是必要的. 比如对于一维方程

$$f(x) = x^2 - x^4,$$

满足必要条件

$$\nabla f(x) = 2x - 4x^3 = 0$$

的点为 $0, 1/\sqrt{2}, -1/\sqrt{2}$, 其中在 0 点取到最小的目标函数值, 但是 0 点并不是全局最小值点, 因为我们需要首先判断全局最小值点是否存在. 事实上, 在这个例子中, $0, 1/\sqrt{2}, -1/\sqrt{2}$ 都不是全局最小值点, 因为 f 随着 $|x| \rightarrow \infty$ 趋于 $-\infty$, 没有全局最小值点. 下面给出另一个例子.

例 1.1.1(代数-几何均值不等式) 证明如下经典不等式 [根据 Cauchy (1821)]

$$(x_1 x_2 \cdots x_n)^{1/n} \leq \frac{\sum_{i=1}^n x_i}{n}$$

对任意的正数集合 $x_i, i = 1, \dots, n$ 均成立. 通过引入变量代换

$$y_i = \ln(x_i), \quad i = 1, \dots, n,$$

我们有 $x_i = e^{y_i}$, 所以不等式可以被写作

$$e^{\frac{y_1 + \cdots + y_n}{n}} \leq \frac{e^{y_1} + \cdots + e^{y_n}}{n},$$

对于所有的标量 $y_1, \dots, y_n$ 都成立. 通过变量代换, 变量的非负性约束可以忽略.

证明上述不等式的一种方法是最小化函数

$$\frac{e^{y_1} + \cdots + e^{y_n}}{n} - e^{\frac{y_1 + \cdots + y_n}{n}},$$

并且其最小值为 0 . 另一种更好的方法是使用最优性条件, 对任意 s , 对满足 $y_1 + \cdots + y_n = s$ 的 $y = (y_1, \dots, y_n)$, 最小化

$$e^{y_1} + \cdots + e^{y_n},$$

并且证明其最优值不小于 $ne^{s/n}$. 为此, 我们使用约束 $y_1 + \cdots + y_n = s$ 消除变量 y_n , 从而得到一个关于 $y_1, \dots, y_{n-1}$ 的等价无约束最小化问题

$$g(y_1, \dots, y_{n-1}) = e^{y_1} + \cdots + e^{y_{n-1}} + e^{s - y_1 - \cdots - y_{n-1}},$$

由必要条件 $\partial g / \partial y_i = 0$, 得到等式

$$e^{y_i} = e^{s - y_1 - \cdots - y_{n-1}}, \quad i = 1, \dots, n-1,$$

或者

$$y_i = s - y_1 - \cdots - y_{n-1}, \quad i = 1, \dots, n-1.$$

该系统只有一个解: 对所有 i , 有 $y_i^* = s/n$. 如果我们能证明存在一个全局最小值点, 那么这个点就一定是全局最小值点. 事实上, 可以看出 $g(y_1, \dots, y_{n-1})$ 是强制的, 所以它存在无约束的全局最小值点, 因此 $(s/n, \dots, s/n)$ 就是全局最小值点, 从而 $e^{y_1} + \cdots + e^{y_n}$ 的最优值为 $ne^{s/n}$. 至此我们证明了代数-几何均值不等式.

然而需要指出, 除了在非常特殊的情况之外, 我们往往很难直接利用最优性条件得到优化问题的最优解. 这是因为求解方程 $\nabla f(x) = 0$ 往往是非常复杂的; 从算法上讲, 有时求解该方程和求解原优化问题的难度是一样的.

除非考虑迭代求解算法, 否则最优性条件的有效性并不明显. 然而最优性条件是很多算法发展和分析的理论基础. 具体而言, 算法通过检验当前解是否满足各种最优性条件来判断这些解是否是最优解, 并且在达到最优解时算法停止. 更进一步, 不同算法在局部最小值点邻域中的行为依赖于在该最小值点处不同的最优性条件是否得到了满足. 比如充分条件在不同算法收敛速度的比较中发挥了很重要的作用.

在微观经济学理论中, 最优性条件也是非常有用的, 其中最优性条件提供了分析的理论基础. 在该领域中, 人们的目的不在于找到最优解, 而是要研究通过改变问题的数据, 如何影响到最优解. 比如一个经济学家可能会对原材料的价格如何影响产品的可用性感兴趣. 在该问题中, 假设产量是优化问题中的变量, 是由相应的生产商确定的. 这种类型的分析称为灵敏度分析, 下面对该问题进行讨论.

灵敏度*

下面分析当一个参数向量变化时最优解如何变化. 特别地, 考虑一个优化问题

$$\begin{aligned} & \text{minimize } f(x, a) \\ & \text{subject to } x \in \mathbb{R}^n, \end{aligned}$$

这里 $f: \mathbb{R}^{m+n} \mapsto \mathbb{R}$ 是一个二阶可微的函数, 其中参数向量 a 为 m 维. 令 $x(a)$ 表示对应 a 的全局最优解, 并假设它存在且唯一. 根据一阶必要条件, 有

$$\nabla_x f(x(a), a) = 0, \quad \forall a \in \mathbb{R}^m,$$

对 a 求导数, 得到

$$\nabla x(a) \nabla_{xx}^2 f(x(a), a) + \nabla_{xa}^2 f(x(a), a) = 0,$$

其中, $m \times n$ 的梯度矩阵 $\nabla x(a)$ 的分量是 $x(a)$ 对 a 的不同元素的一阶偏导数. 假设矩阵可逆, 有

$$\nabla x(a) = -\nabla_{xa}^2 f(x(a), a) (\nabla_{xx}^2 f(x(a), a))^{-1}, \quad (1.3)$$

式 (1.3) 给出了最优解 x 关于 a 的一阶导数.

为了让上述的分析更严谨, 必须保证 $x(a)$ 存在并且是关于 a 的可微函数. 这里主要的分析工具是隐函数定理 (附录 A 命题 A.25). 利用这个定理, 我们可以将 $x(a)$ 定义在以某个最小值点 $\bar{x} = x(\bar{a})$ 为中心的开球中, 其中 $\bar{x}$ 是与 $\bar{a}$ 对应的最小值点, 这里假定 Hessian 矩阵 $\nabla_{xx}^2 f(\bar{x}, \bar{a})$ 是正定的. 当假定局部最小值点 $\bar{x}$ 非奇异时, 上述分析是可行的.

3.2 节给出了关于灵敏度的更加详细的分析, 并且给出了对于 $\nabla x(a)$ 的约束表达式 (1.3).

1.1.1 主要的最优性条件

下面对之前提到的最优性条件进行严格的叙述与证明.

命题 1.1.1 (最优性必要条件) 令 x^* 为 $f: \mathbb{R}^n \mapsto \mathbb{R}$ 的无约束局部最小值点, 并且假设 f 在开集 S 上连续可微, $x^* \in S$, 那么有

$$\nabla f(x^*) = 0. \quad (\text{一阶必要条件})$$

如果 f 在 S 中是二阶连续可微的, 那么有

$$\nabla^2 f(x^*): \text{半正定}. \quad (\text{二阶必要条件})$$

证明 对给定的 $d \in \mathbb{R}^n$, 利用链式法则对函数 $g(\alpha) = f(x^* + \alpha d)$ 关于标量 α 求导, 我们得到

$$0 \leq \lim_{\alpha \downarrow 0} \frac{f(x^* + \alpha d) - f(x^*)}{\alpha} = \frac{dg(0)}{d\alpha} = d' \nabla f(x^*),$$

上述不等式利用了 x^* 为局部最小值点的假设. 因为 d 是任意选取的, 那么如果把 d 用 $-d$ 替换, 不等式仍然成立, 所以对于所有 $d \in \mathbb{R}^n$, 都有 $d' \nabla f(x^*) = 0$, 从而有 $\nabla f(x^*) = 0$.

假设 f 二阶连续可微, 令 d 为 $\mathbb{R}^n$ 中的任意一个向量, 对于所有 $\alpha \in \mathbb{R}$, 利用二阶泰勒展开可以得到

$$f(x^* + \alpha d) - f(x^*) = \alpha \nabla f(x^*)' d + \frac{\alpha^2}{2} d' \nabla^2 f(x^*) d + o(\alpha^2).$$

利用条件 $\nabla f(x^*) = 0$ 以及 x^* 的局部最优性, 可知存在一个充分小的 $\varepsilon > 0$, 满足对于所有 $\alpha \in (0, \varepsilon)$

$$0 \leq \frac{f(x^* + \alpha d) - f(x^*)}{\alpha^2} = \frac{1}{2} d' \nabla^2 f(x^*) d + \frac{o(\alpha^2)}{\alpha^2}.$$

取 $\alpha \rightarrow 0$ 的极限, 利用

$$\lim_{\alpha \rightarrow 0} \frac{o(\alpha^2)}{\alpha^2} = 0,$$

得到 $d' \nabla^2 f(x^*) d \geq 0$, 从而 $\nabla^2 f(x^*)$ 是正定的. $\square$

下面的命题考虑了凸目标函数的情形.

命题 1.1.2 (凸目标函数) 令 $f: X \mapsto \mathbb{R}$ 为凸集 X 上的一个凸函数.

(a) 在 X 上的 f 的最小值点也是 X 的全局最小值点. 如果 f 是严格凸的, 那么至多存在一个全局最小值点.

(b) 如果 f 为凸函数, X 为开集, 那么向量 $x^* \in X$ 是 f 在 X 上全局最优解的充要条件是 $\nabla f(x^*) = 0$.

证明 (a) 部分在附录 B 的命题 B.10 中已证. 为了证明 (b), 利用附录 B 中的命题 B.3, 我们有

$$f(x) \geq f(x^*) + \nabla f(x^*)'(x - x^*), \quad \forall x \in X.$$

如果 $\nabla f(x^*) = 0$, 那么对所有 $x \in X$, 有 $f(x) \geq f(x^*)$, 所以 x^* 是一个全局最小值点. $\square$