玄之又玄　众妙之门
Fermat-Torricelli 问题与凸函数

521020910200 朱坦晟

2021 年 12 月 29 日

摘要

在线性规划求解极值问题时我们会用到许多与凸性有关的方法. 于是笔者想到了中学时期遇到过的一个经典的极值问题 – Fermat-Torricelli 问题, 我们在中学时对该问题的解法多借助几何图形的巧妙构造与三角不等式, 而现在则希望通过分析的方法, 借助凸函数这一工具, 对该问题进行探究.

1　问题的引入

1.1　最初的问题

已知平面上锐角三角形 ABC, 试引入平面上一点 P, 使得 $|PB| + |PB| + |PC|$ 最小.17 世纪时, 法国数学家 Fermat 向意大利的数学家和物理学家 Torricelli 提出了这个问题. Torricelli 证明了这个点的存在性, 并且有 $\angle APB = \angle BPC = \angle CPA = 120°$. 他还指出, 以 AB,BC,CA 为底边分别向外构造等边三角形 $\triangle ABC'$,$\triangle BCA'$,$\triangle CAB'$, 则 AA',BB',CC' 交于一点 P 即为所求. 该点也被称为 Fermat 点.

为了直观地引入该问题, 我们在这里先通过作图法对三点情况下的 Fermat-Torricelli 问题求解. 以下介绍两种比较简单的证明方法.

1.2　几何解法之一

证明. 如图1.1所示, 对于 $\triangle ABC$ 内任意一点 P, 连接 PA,PB,PC, 将 $\triangle ABP$ 绕 B 点逆时针旋转 $60°$ 得到 $\triangle C'BP'$, 则 $\triangle ABP \cong \triangle C'BP'$. 连接

PP' 和 CC', 则由三角形全等有

$$|PA|+|PB|+|PC|=|C'P|+|P'P|+|PC|$$

由三角形不等式有

$$|C'P|+|P'P|+|PC|\geq|CC'|$$

因此只要 P 和 P' 都落在 CC' 上即可取得最小值.

以下我们去说明取得最小值的条件. 考虑 P' 点, 若 P' 落在 CC' 上, 则有 $\angle C'P'B=180^\circ-\angle BP'P=120^\circ$, 所以 $\angle APB=120^\circ$. 同理, 我们可以根据对称性得到 $\angle BPC=\angle CPA=120^\circ$, 并且 P 为 AA',BB',CC' 的交点. □

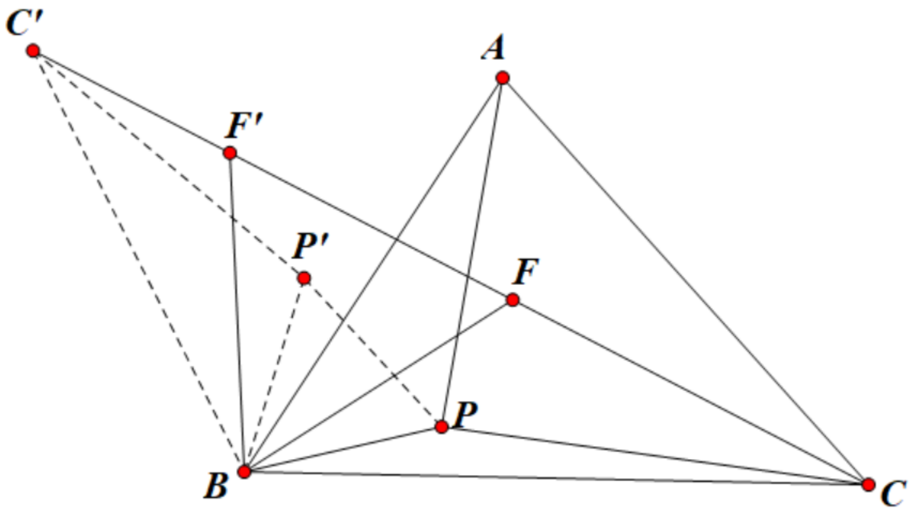

图 1.1: 利用旋转构造全等三角形

1.3 几何解法之二

证明. 如图1.2对于给定的锐角三角形 ABC, 我们不妨先固定线段 CP 的长度, 以 $|CP|$ 为半径画 $\odot C$. 接着我们以 A,B 为两个焦点作椭圆使其与 $\odot C$ 相切, 记切点为 F, 连接 FA,FB,FC. 作 AB 中点 F, 对于 $\odot C$ 上的任

意一点 P, 连接 PO 交椭圆于 H. 由椭圆的定义我们有

$$|HA| + |HB| = |FA| + |FB|$$

另一方面, 由三角不等式有

$$|PA| + |PB| \geq |HA| + |HB|$$

于是

$$|PA| + |PB| \geq |FA| + |FB|$$

当且仅当 P 与 F 重合时等号成立.

接下来, 对于 F 点, 我们由相切条件知 CF 为 F 处的法线, 并且由椭圆的光学性质有 CF 平分 $\angle AFB$. 由 $|CP|$ 的任意性知,CP 平分 $\angle APB$ 是长度和取得极小值的必要条件. 同理, 我们由对称性有 AP 平分 $\angle BPC$ 和 BP 平分 $\angle CPA$, 故 $\angle APB = \angle BPC = \angle CPA = 120°$ □

注 1. *事实上, 这里可以运用 Fermat 原理 (过空间中两定点的光, 其实际路径的光程最短) 直接得出 CF 平分 $\angle AFB$ 的结论.*

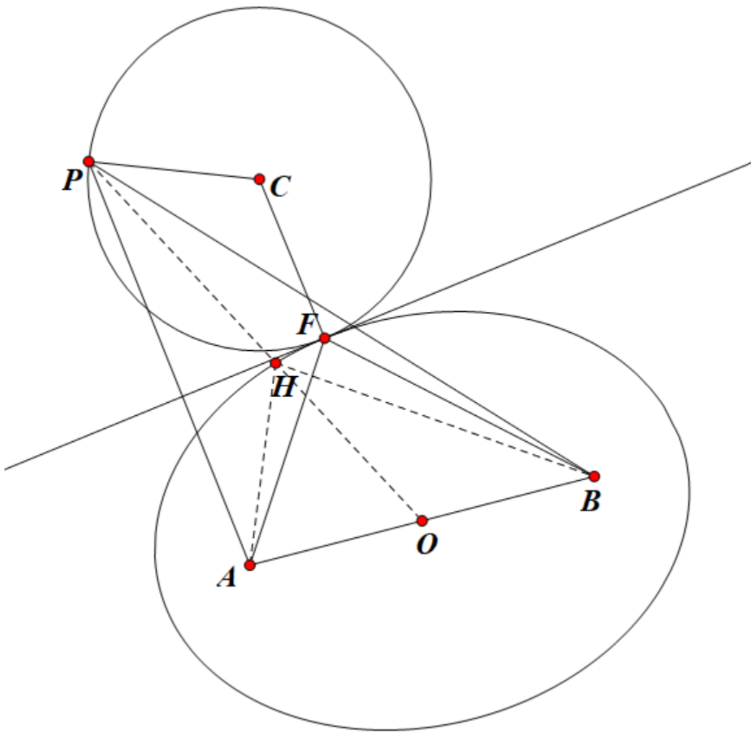

图 1.2: 利用椭圆的光学性质

1.4 相似问题的拓展

1.4.1 多点情形的 Fermat-Torricelli 问题

这是下文中将要讨论的问题, 即要寻找 $\mathbb{R}^n$ 中的一点使其到 $\mathbb{R}^n$ 中有限个给定点的距离之和最小.

给定点 $a_i \in \mathbb{R}^n, (i = 1, 2, \cdots, m), ||\cdot||$ 表示欧式范数. 定义函数

$$\varphi(x) \coloneqq \sum_{i=1}^{m} ||x - a_i|| \tag{1}$$

则 Fermat-Torricelli 问题可描述为

$$\min \varphi(x), \quad \forall x \in \mathbb{R}^n \tag{2}$$

1.4.2 Fermat-Weber 场址问题

给定点 $a_i \in \mathbb{R}^n, (i = 1, 2, \cdots, m), w_i (i = 1, 2, \cdots, m)$ 是 m 个正数, 则 Fermat-Weber 场址问题可描述为

$$\min \varphi(x) = \sum_{i=1}^{m} w_i ||x - a_i||, \quad \forall x \in \mathbb{R}^n$$

1.4.3 Steiner 最小树问题

一般研究二维欧式空间上的 Steiner 最小树问题, 即对于二维平面上的点集 P, 试添加新点集 S, 使得 $P \cup S$ 的最小生成树总长度最小.

在该问题下同样有三点 Fermat-Torricelli 问题类似的有趣性质: 点集 S 中任意点与之相连的边有 3 条, 且两两夹角均为 120°. 图1.3给出点集 P 为正方形四个顶点的情况.

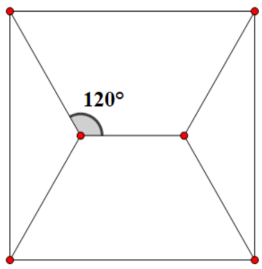

图 1.3: Steiner 最小树的示例 ($|P| = 4$)

2 凸函数是个新世界

该节引入凸函数, 主要用于说明 Fermat-Torricelli 问题是一个求解凸函数绝对极小值的问题, 并说明解的存在性与唯一性, 然后再去探讨极值点需要满足的性质.

2.1 $\varphi(x)$ 的凸性

首先给出凸函数的定义, 并说明我们将要用到的两个基本函数的凸性.

定义 2.1. *设 $f:\Omega \to \bar{\mathbb{R}} = (-\infty, +\infty]$ 是定义在凸集 $\Omega \subset \mathbb{R}^n$ 上的增广实值函数, 如果*

$$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda) f(y), \quad \forall x, y \in \Omega, \quad \lambda \in (0,1) \tag{3}$$

则称 f 在 Ω 上是凸的.

如果对于 $x \ne y$, 式(3)等号严格不成立, 则称 f 在 Ω 上是严格凸的.

命题 2.1. *以下函数是凸的 $(a \in \mathbb{R}^n)$:*

(i) $f(x) := \langle a, x \rangle, \quad \forall x \in \mathbb{R}^n$

(ii) $f(x) := ||x-a||, \quad \forall x \in \mathbb{R}^n$

证明. (i)

$$\begin{aligned} f(\lambda x + (1-\lambda)y) &= \langle \lambda x + (1-\lambda)y, a \rangle \\ &= \lambda \langle x, a \rangle + (1-\lambda)\langle y, a \rangle \\ &= \lambda f(x) + (1-\lambda) f(y) \end{aligned}$$

(ii)

$$\begin{aligned} f(\lambda x + (1-\lambda)y) &= ||\lambda x + (1-\lambda)y - a|| \\ &= ||\lambda(x-a) + (1-\lambda)(y-a)|| \\ &\le ||\lambda(x-a)|| + ||(1-\lambda)(y-a)|| \\ &= \lambda||(x-a)|| + (1-\lambda)||(y-a)|| \\ &= \lambda f(x) + (1-\lambda) f(y) \end{aligned}$$

□

为了说明我们研究的 $\varphi(x)$ 是凸函数, 还需要说明有限个凸函数之和的凸性.

命题 2.2. 设 $f_i:\mathbb{R}^n\to\bar{\mathbb{R}}$ 是凸函数,$i=1,2,\cdots,m$, 则和函数 $\sum\limits_{i=1}^{m} f_i$ 是凸函数

证明. 任取 $x,y\in\mathbb{R}^n,\lambda\in(0,1)$, 则有

$$\begin{aligned}(f_1+f_2)(\lambda x+(1-\lambda)y)&=f_1(\lambda x+(1-\lambda)y)+f_2(\lambda x+(1-\lambda)y)\\&\le\lambda f_1(x)+(1-\lambda)f_1(y)+\lambda f_2(x)+(1-\lambda)f_2(y)\\&=\lambda(f_1+f_2)(x)+(1-\lambda)(f_1+f_2)(y)\end{aligned}$$

于是 f_1+f_2 是凸函数, 由此可推出 $\sum\limits_{i=1}^{m} f_i$ 是凸函数 □

于是, 利用上述诸定义和命题, 我们可以得到 Fermat-Torricelli 问题中定义的函数1是一个连续的凸函数. 方便下文给出凸函数的有关性质和定理来探究 $\min\varphi(x)$.

命题 2.3. $\varphi(x)$ 函数是连续的凸函数

证明. 设 $\varphi_i(x)=||x-a_i||,(i=1,2,\cdots,m).\varphi_i(x)$ 的连续性显然, 则有 $\varphi(x)$ 是连续函数. 而由命题2.1知 $\varphi_i(x)$ 是凸函数, 再由命题2.2可知 $\varphi(x)=\sum\limits_{i=1}^{m}\varphi_i(x)$ 是 $\mathbb{R}^n$ 上连续的凸函数. □

2.2 解的存在性与唯一性

为了说明 Fermat-Torricelli 问题的有解性, 我们还需要考虑连续函数的最值可以保证取得这一性质, 故我们有如下定理.

定理 2.4. *(Weierstrass) 设 $f:\Omega\to\mathbb{R}$ 是连续函数, 其中 Ω 是 $\mathbb{R}^n$ 的非空紧子集, 则存在 $x',x''\in\Omega$ 满足*

$$f(x')=\sup\{f(x)|x\in\Omega\},\quad f(x'')=\inf\{f(x)|x\in\Omega\}$$

利用该定理和 $\varphi(x)$ 函数的连续性, 可以得出 Fermat-Torricelli 问题的有解性.

命题 2.5. *Fermat-Torricelli 问题2至少有一个解*

证明. 由定理2.4和命题2.3即证. □

若我们给 Fermat-Torricelli 问题加上限制, 保证给定的有限个点 a_i $(i = 1, 2, \cdots, m)$ 不共线, 则可保证问题有唯一解. 后文中我们对于一般情形下 Fermat-Torricelli 问题的讨论也将以该假设为前提. 而为了说明唯一解, 我们需要用到严格凸函数的以下定理.

定理 2.6. *连续的严格凸函数 $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ 有唯一绝对极小值点*

证明. 反证法, 设 $\bar{x}, \bar{y}$ 是严格凸函数 $f : \mathbb{R}^n \to \mathbb{R}$ 的两个绝对极小值点, 且 $\bar{x} \neq \bar{y}$, $f(\bar{x}) = f(\bar{y})$, 则对于 $\forall \lambda \in (0, 1)$ 有

$$\begin{aligned} & f(\lambda\bar{x} + (1-\lambda)\bar{y}) \\ < & \lambda f(\bar{x}) + (1-\lambda) f(\bar{y}) \\ = & f(\bar{x}) \end{aligned}$$

这与 $\bar{x}, \bar{y}$ 是 f 的极小值点矛盾 □

利用该定理和给定点两两不共线的预设, 即可说明解的唯一性.

命题 2.7. *若 $a_i, (i = 1, 2, \cdots, m)$ 不共线, 则 Fermat-Torricelli 问题2有唯一解*

证明. 由定理2.6, 我们只需证明 $\varphi(x)$ 是严格凸函数即可.

用反证法, 假设 $\varphi(x)$ 不是严格凸函数, 故 $\exists \bar{x}, \bar{y} \in \mathbb{R}^n, \bar{x} \neq \bar{y}, \lambda \in (0, 1)$ 满足

$$\varphi(\lambda\bar{x} + (1-\lambda)\bar{y}) = \lambda\varphi(\bar{x}) + (1-\lambda)\varphi(\bar{y})$$

由 $\varphi(x) = \sum\limits_{i=1}^{m} \varphi_i(x)$ 及取等条件可得

$$\varphi_i(\lambda\bar{x} + (1-\lambda)\bar{y}) = \lambda\varphi_i(\bar{x}) + (1-\lambda)\varphi_i(\bar{y}), \quad (i = 1, 2, \cdots, m)$$

所以

$$||\lambda(\bar{x} - a_i) + (1-\lambda)(\bar{y} - a_i)|| = \lambda||\bar{x} - a_i|| + (1-\lambda)||\bar{y} - a_i||$$

由三角不等式的取等条件, 可得 $\lambda(\bar{x} - a_i)$ 和 $(1-\lambda)(\bar{y} - a_i)$ 共线

故可设

$$(1-\lambda)(\bar{y} - a_i) = t_i \lambda(\bar{x} - a_i)$$

记 $\gamma_i = \dfrac{1-\lambda}{\lambda t_i}$, 故

$$\begin{aligned}\bar{x} - a_i &= \gamma_i(\bar{y} - a_i) \\ \Rightarrow a_i &= \frac{1}{1-\gamma_i}\bar{x} + \frac{-\gamma_i}{1-\gamma_i}\bar{y} \in \mathcal{L}(\bar{x}, \bar{y})\end{aligned}$$

这表明 a_i $(i = 1, 2, \cdots, m)$ 都在直线 $\mathcal{L}(\bar{x}, \bar{y})$ 上, 与条件矛盾 □

2.3 凸函数与极值

在凸函数意义下, 次梯度给了我们解决极值问题的一般方法. 下面首先给出微分和次微分的有关定义.

定义 2.2. *(Frechet) 对函数 $f : \mathbb{R}^n \to \bar{\mathbb{R}}$,$\bar{x} \in \mathrm{dom} f$, 若 $\exists v \in \mathbb{R}^n$, 满足*

$$\lim_{x \to \bar{x}} \frac{f(x) - f(\bar{x}) - \langle v, x - \bar{x} \rangle}{||x - \bar{x}||} = 0$$

则称 f 在 $\bar{x}$ 处可微, 记 $\nabla f(\bar{x}) = v$.

定义 2.3. *设 $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ 是凸函数, 且 $\bar{x} \in \mathrm{dom} f$, 称*

$$\partial f(\bar{x}) = \{v \in \mathbb{R}^n | \langle v, x - \bar{x} \rangle \leq f(x) - f(\bar{x}), \forall x \in \mathbb{R}^n\}$$

为函数 f 在 $\bar{x}$ 处的次微分, 且称 $v \in \partial f(\bar{x})$ 为 f 在 $\bar{x}$ 处的次梯度.

注 2. *为说明次微分是表示凸函数极值的普适方法. 我们此处简证凸函数的次微分是一个非空的凸紧集: 一方面, 可以由凸集分割定理 (Minkowski) 和凸函数上镜图 (epigraph) 的凸性得到次微分非空; 另一方面, 由命题2.1可以导出次微分的凸性.*

在讨论定义域为 $\mathbb{R}$ 的连续函数之极值时, 我们通常采用 Fermat 驻点定理. 但对于一般的凸函数, 不一定在每一点都满足可微条件, 在该种情形下, 我们仍可以用次微分 (次梯度) 来描述极小值.

命题 2.8. *设 $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ 是凸函数, 且 $\bar{x} \in \mathrm{dom} f$, 则 $\bar{x}$ 为局部 (全局) 极小值点当且仅当 $0 \in \partial f(\bar{x})$.*

证明. 对于 $\forall x \in \mathbb{R}^n$, 有

$$
\begin{aligned}
& f(\bar{x}) \le f(x) \\
\iff & 0 = \langle 0, x-\bar{x}\rangle \le f(x) - f(\bar{x}) \\
\iff & 0 \in \partial f(\bar{x})
\end{aligned}
$$

□

在解决极小值点的问题中, 凸函数的性质决定其能够把局部最优解推广到全局最优解.

定理 2.9. *对凸函数 $f:\mathbb{R}^n \to \bar{\mathbb{R}}$, 若 $\bar{x} \in \mathrm{dom}f$ 是局部极小值点, 则 $\bar{x}$ 为全局极小值点.*

证明. $\bar{x} \in \mathrm{dom}f$ 是局部极小值点, 故由定义知 $\exists\delta > 0, \forall x \in \mathbb{B}(\bar{x},\delta), f(x) \ge f(\bar{x})$. 取定 $x \in \mathbb{R}^n$, 定义

$$x_k \coloneqq (1-\frac{1}{k})\bar{x} + \frac{1}{k}x,\ k \in \mathbb{N}_+$$

则当 k 充分大时, 有 $x_k \in \mathbb{B}(\bar{x},\delta)$. 故由凸函数性质得

$$f(\bar{x}) \le f(x_k) \le (1-\frac{1}{k})f(\bar{x}) + \frac{1}{k}f(x)$$

所以

$$f(\bar{x}) \le f(x), \quad \forall x \in \mathbb{R}^n$$

即 $\bar{x}$ 是全局极小值点. □

最后为了统一微分和次微分, 我们下证在可微点的次微分实际上是单点集, 即只包含凸函数在该点的导数.

命题 2.10. *设 $f:\mathbb{R}^n \to \bar{\mathbb{R}}$ 是凸函数, 且在内点 $\bar{x} \in \mathrm{dom}f$ 处可微, 则有*

$$\partial f(\bar{x}) = \{\nabla f(\bar{x})\}$$

证明. 一方面, 由可微的定义知,$\forall\varepsilon > 0, \exists\delta > 0$, 对 $\forall x \in \mathbb{B}(\bar{x},\delta)$, 满足

$$-\varepsilon||x-\bar{x}|| \le f(x) - f(\bar{x}) - \langle\nabla f(\bar{x}), x-\bar{x}\rangle \le \varepsilon||x-\bar{x}|| \tag{4}$$

定义函数 $\phi(x) \coloneqq f(x) - f(\bar{x}) - \langle\nabla f(\bar{x}), x-\bar{x}\rangle + \varepsilon||x-\bar{x}||$

由命题2.1, 命题2.2可得 $\phi(x)$ 是凸函数. 并且我们有 $\forall x \in \mathbb{B}(\bar{x}, \delta), \phi(x) \geq 0 = \phi(\bar{x})$. 故由定理2.9知

$$\forall x \in \mathbb{R}^n, \phi(x) \geq 0$$

所以有

$$\langle \nabla f(\bar{x}), x - \bar{x} \rangle \leq f(x) - f(\bar{x}) + \varepsilon ||x - \bar{x}||$$

当 $\varepsilon \to 0$ 时,$\langle \nabla f(\bar{x}), x - \bar{x} \rangle \leq f(x) - f(\bar{x})$, 即 $\nabla f(\bar{x}) \in \partial f(\bar{x})$.

另一方面, 设 $v \in \partial f(\bar{x})$, 故有

$$\langle v, x - \bar{x} \rangle \leq f(x) - f(\bar{x}), \quad \forall x \in \mathbb{R}^n$$

又由式4知

$$f(x) - f(\bar{x}) \leq \langle \nabla f(\bar{x}), x - \bar{x} \rangle + \varepsilon ||x - \bar{x}||, \quad \forall x \in \mathbb{B}(\bar{x}, \delta)$$

故

$$\langle v - \nabla f(\bar{x}), x - \bar{x} \rangle \leq \varepsilon ||x - \bar{x}||, \quad \forall x \in \mathbb{B}(\bar{x}, \delta)$$

对于 Cauchy-Schwarz 不等式:

$$\langle v - \nabla f(\bar{x}), x - \bar{x} \rangle \leq ||v - \nabla f(\bar{x})|| \cdot ||x - \bar{x}||$$

考虑取等条件. 一定存在 $x_0 \in \mathbb{B}(\bar{x}, \delta), x_0 \neq \bar{x}$ 使得等号成立, 故

$$||v - \nabla f(\bar{x})|| \leq \varepsilon$$

由 ε 的任意性得

$$v = \nabla f(\bar{x})$$

综上,$\partial f(\bar{x}) = \{\nabla f(\bar{x})\}$. □

3 三点 Fermat-Torricelli 问题的最优解

对于三点的 Fermat-Torricelli 问题, 不论给定三点的位置关系如何, 我们总是能够找到唯一解.

首先给出一个简单命题和两个引理, 方便我们将由凸函数得出的结论过渡到几何表示.

命题 3.1. 设 $f(x) \coloneqq ||x||, \forall x \in \mathbb{R}^n$, 则有

$$\partial f(x) = \begin{cases} \mathbb{B}, & x = 0 \\ \{\frac{x}{||x||}\}, & x \neq 0 \end{cases}$$

证明. $f(x)$ 在 $x \neq 0$ 处可微, 且 $\nabla f(x) = \frac{x}{||x||}$, 故 $\partial f(x) = \{\frac{x}{||x||}\}$.

当 $x = 0$ 时, 设 $v \in \partial f(0)$, 由定义2.3知

$$\langle v, x - 0 \rangle \leq f(x) - f(0), \quad \forall x \in \mathbb{R}^n$$

若取 $x = v$, 可得 $\langle v, v \rangle \leq ||v||$, 则 $||v|| \leq 1$, 即 $v \in \mathbb{B}$, 故有 $\partial f(0) \subset \mathbb{B}$

下设对 $\forall v \in \mathbb{B}$, 由 Cauchy-Schwarz 不等式

$$\langle v, x - 0 \rangle = \langle v, x \rangle \leq ||v|| \cdot ||x|| \leq ||x|| = f(x) - f(0)$$

所以 $v \in \partial f(0)$, 故有 $\mathbb{B} \subset \partial f(0)$

综上,$\partial f(0) = \mathbb{B}$. □

引理 3.2. 设 $a_1, a_2 \in \mathbb{R}^n$ 为单位向量, 则 $a_1 + a_2 \in \mathbb{B}$ 当且仅当 $\langle a_1, a_2 \rangle \leq -\frac{1}{2}$

证明.

$$\begin{aligned} a_1 + a_2 \in \mathbb{B} &\iff ||a_1 + a_2|| \leq 1 \\ &\iff ||a_1 + a_2||^2 \leq 1 \\ &\iff ||a_1||^2 + ||a_2||^2 + 2\langle a_1, a_2 \rangle \leq 1 \\ &\iff \langle a_1, a_2 \rangle \leq -\frac{1}{2} \end{aligned}$$

□

引理 3.3. 设 $a_1, a_2, a_3 \in \mathbb{R}^n$ 为单位向量, 则 $a_1 + a_2 + a_3 = 0$ 成立当且仅当

$$\langle a_1, a_2 \rangle = \langle a_2, a_3 \rangle = \langle a_1, a_3 \rangle = -\frac{1}{2} \tag{5}$$

证明. 一方面, 由 $a_1 + a_2 + a_3 = 0$ 可得

$$\begin{aligned} ||a_1||^2 + \langle a_1, a_2 \rangle + \langle a_1, a_3 \rangle = 0 \\ \langle a_2, a_1 \rangle + ||a_2||^2 + \langle a_2, a_3 \rangle = 0 \\ \langle a_3, a_1 \rangle + \langle a_3, a_2 \rangle + ||a_3||^2 = 0 \end{aligned}$$

故有等式5成立.

另一方面, 当等式5成立时,

$$(a_1+a_2+a_3)^2=\sum_{i=1}^{3}||a_i||^2+2\sum_{1\le i<j\le 3}\langle a_i,a_j\rangle=0$$

故 $a_1+a_2+a_3=0$. □

以下讨论三点 Fermat-Torricelli 问题最优解的充要条件.

记

$$\cos(u,v):=\frac{\langle u,v\rangle}{||u||\cdot||v||},\quad v_i=\frac{\bar{x}-a_i}{||\bar{x}-a_i||},||v_i||=1$$

定理 3.4. *对问题2, 分为两种情况讨论*

(i) 若 $\bar{x}\notin\{a_1,a_2,a_3\}$, 则 $\bar{x}$ 为最优解当且仅当

$$\cos(v_1,v_2)=\cos(v_2,v_3)=\cos(v_3,v_1)=-\frac{1}{2}$$

(ii) 若 $\bar{x}\in\{a_1,a_2,a_3\}$, 不妨设 $\bar{x}=a_1$, 则 $\bar{x}$ 为最优解当且仅当

$$\cos(v_2,v_3)\le-\frac{1}{2}$$

证明. (i) 此时 $\varphi(x)$ 在 $\bar{x}$ 处可微, 由命题3.1有

$$\nabla\varphi(\bar{x})=v_1+v_2+v_3$$

又由命题2.8和命题2.10知

$$v_1+v_2+v_3=0$$

根据引理3.3知上式成立当且仅当

$$\langle v_1,v_2\rangle=\langle v_2,v_3\rangle=\langle v_3,v_1\rangle=-\frac{1}{2}$$

即

$$\cos(v_1,v_2)=\cos(v_2,v_3)=\cos(v_3,v_1)=-\frac{1}{2}$$

(ii) 此时由命题2.8和命题3.1知

$$0 \in \partial\varphi(\bar{x}) = v_2 + v_3 + \mathbb{B}$$

故有

$$v_2 + v_3 \in \mathbb{B}$$

根据引理3.2可得

$$\langle v_2, v_3 \rangle \leq -\frac{1}{2}$$

即

$$\cos(v_2, v_3) \leq -\frac{1}{2}$$

□

至此, 我们已经用凸函数的性质解决了三点情况下的 Fermat-Torricelli 问题. 引入凸函数工具来对最值进行刻画. 令人兴奋的是, 我们成功告别几何直观, 用凸函数的次梯度解决了简单的最值问题.

4 一般情况下 Fermat-Torricelli 问题的最优解

对于一般情况下的 Fermat-Torricelli 问题, Endre Weiszfeld[**ref5**] 和 Harold W. Kuhn[**ref6**] 给出了数值算法.

为便于下文证明的叙述, 我们先给出聚点定理和凸包有关的定义.

定理 4.1. *(Bolzano-Weierstrass)* $\mathbb{R}^n$ *中的有界集合* S *包含无穷多个点, 那么至少有一点* $x \in \mathbb{R}^n$ *为* S *的一个聚点.*

证明. 因为 S 有界, 所以 $\exists J^{(0)} = I_1^{(0)} \times I_2^{(0)} \times \cdots \times I_n^{(0)}$, 其中 $I_i^{(0)} = [a_i^{(0)}, b_i^{(0)}]$ $(i = 1, 2, \cdots, m)$ 是一维的有界闭区间, 满足 $S \subset J^{(0)}$.

接下来对 $J^{(0)}$ 进行二分. 对于 $\forall i = 1, 2, \cdots, m$, 把 $I_i^{(0)} = [a_i^{(0)}, b_i^{(0)}]$ 对半分为两个区间 $[a_i^{(0)}, \frac{a_i^{(0)}+b_i^{(0)}}{2}]$ 和 $[\frac{a_i^{(0)}+b_i^{(0)}}{2}, b_i^{(0)}]$, 至少有一个区间含有 S 中的无穷多个点, 把该区间记为 $I_i^{(1)} = [a_i^{(1)}, b_i^{(1)}]$, 故可得

$$J^{(1)} = I_1^{(1)} \times I_2^{(1)} \times \cdots \times I_n^{(1)}$$

以此类推能够得到

$$J^{(k)} = I_1^{(k)} \times I_2^{(k)} \times \cdots \times I_n^{(k)} \subset J^{(0)}$$

每个一维区间 $I_i^{(k)}=[a_i^{(k)},b_i^{(k)}]$ 的长度为 $|\frac{b_i^{(0)}-a_i^{(0)}}{2^k}|\to 0$, 且有 $I_i^{(0)}\supset I_i^{(1)}\supset\cdots\supset I_i^{(k)}\supset\cdots$, 故由闭区间套定理可知 $a_i^{(k)},b_i^{(k)}$ 收敛于一点 t_i. 所以 $t=(t_1,t_2,\cdots,t_m)^T$ 是 S 的一个聚点 □

定义 4.1. 设 Ω 是 $\mathbb{R}^n$ 的子集,Ω 的凸包定义为

$$\mathrm{conv}\Omega=\{\sum_{i=1}^{m}\lambda_i a_i|\sum_{i=1}^{m}\lambda_i=1,\lambda_i\geq 0,a_i\in\Omega,m\in\mathbb{N}\}$$

命题 4.2. 设 Ω 是 $\mathbb{R}^n$ 中有限点集 $\{a_1,a_2,\cdots,a_m\}$ 生成的凸包, 则 Ω 是有界的.

证明. 任取一点 a_1, 由凸包的定义4.1知 $\forall x\in\Omega, x=\sum_{i=1}^m\lambda_i a_i\quad(\sum_{i=1}^m\lambda_i=1,\ \lambda_i\geq 0)$. 由命题2.1可得

$$\begin{aligned}||x-a_1||&=||\sum_{i=1}^{m}\lambda_i a_i-a_1||\\&\leq\sum_{i=1}^{m}\lambda_i||a_i-a_1||\\&\leq\sum_{i=1}^{m}||a_i-a_1||\end{aligned}$$

于是 $\Omega\subset\mathbb{B}(a_1,\sum_{i=1}^m||a_i-a_1||)$, 故 Ω 有界. □

下面开始正式介绍 Weiszfeld 算法. 此处声明, 对于一般情形下的 Fermat-Torricelli 问题, 我们只考虑给定点两两不共线时的情况.

Weiszfeld 算法首先假定 $a_1,a_2,\cdots,a_m$ 不共线, 则由命题2.7保证问题2有唯一解.

当 $x\notin\{a_1,a_2,\cdots,a_m\}$ 时,$\varphi(x)$ 是可微的, 我们有

$$\nabla\varphi(x)=\sum_{i=1}^{m}\frac{x-a_i}{||x-a_i||},\quad x\notin\{a_1,a_2,\cdots,a_m\}$$

令 $\nabla\varphi(x)=0$, 解得

$$x=\frac{\sum\limits_{i=1}^{m}\dfrac{a_i}{||x-a_i||}}{\sum\limits_{i=1}^{m}\dfrac{1}{||x-a_i||}}$$

定义函数

$$f(x)=\begin{cases}\dfrac{\sum\limits_{i=1}^{m}\dfrac{a_i}{||x-a_i||}}{\sum\limits_{i=1}^{m}\dfrac{1}{||x-a_i||}}, & x\notin\{a_1,a_2,\cdots,a_m\}\\ \\ x, & x\in\{a_1,a_2,\cdots,a_m\}\end{cases}\tag{6}$$

所以若 $\bar{x}\notin\{a_1,a_2,\cdots,a_m\}$ 且 $\bar{x}=f(\bar{x})$, 则由上述构造可得 $\bar{x}$ 是最优解.

选取任意点 $x_1\in\mathbb{R}^n$, 构造序列 $\{x_n\}$ 满足 $x_{n+1}\coloneqq f(x_n)$. 以下我们先证该序列最后一定收敛于某一个 $\bar{x}$.

命题 4.3. *(序列的收敛性) 若 $f(x)\neq x$, 则有 $\varphi(f(x))<\varphi(x)$.*

证明. 因为 $f(x)\neq x$, 故 $x\notin\{a_1,a_2,\cdots,a_m\}$.
定义函数

$$g(z)\coloneqq\sum_{i=1}^{m}\frac{||z-a_i||^2}{||x-a_i||}$$

有性质:

(i) 对于 $g_i(z)\coloneqq\dfrac{||z-a_i||^2}{||x-a_i||}$,$\forall z_1,z_2\in\mathbb{R}^n,z_1\neq z_2,\lambda\in(0,1)$, 我们有

$$\begin{aligned}g_i(\lambda z_1+(1-\lambda)z_2)&=\frac{||\lambda(z_1-a_i)+(1-\lambda)(z_2-a_i)||^2}{||x-a_i||}\\&\leq\lambda^2\frac{||z_1-a_i||^2}{||x-a_i||}+(1-\lambda)^2\frac{||z_2-a_i||^2}{||x-a_i||}\\&<\lambda\frac{||z_1-a_i||^2}{||x-a_i||}+(1-\lambda)\frac{||z_2-a_i||^2}{||x-a_i||}\\&=\lambda g_i(z_1)+(1-\lambda)g_i(z_2)\end{aligned}$$

故 $g_i(z)$ 是严格凸函数, 由命题2.2知 $g(z)=\sum\limits_{i=1}^{m}g_i(z)$ 是严格凸函数

(ii) $g(z)$ 可微, 令

$$\nabla g(z)=\sum_{i=1}^{m}\frac{2(z-a_i)}{||x-a_i||}=0$$

解得

$$z = \frac{\sum_{i=1}^{m} \frac{a_i}{||x-a_i||}}{\sum_{i=1}^{m} \frac{1}{||x-a_i||}} = f(x)$$

因此由命题2.8和命题2.10知 $z = f(x)$ 是 $g(z)$ 的唯一极小值点. 故

$$g(f(x)) < g(x) = \varphi(x)$$

且

$$\begin{aligned} g(f(x)) &= \sum_{i=1}^{m} \frac{||f(x)-a_i||^2}{||x-a_i||} \\ &= \sum_{i=1}^{m} \frac{||(x-a_i)+(f(x)-a_i)-(x-a_i)||^2}{||x-a_i||} \\ &= \varphi(x) + 2(\varphi(f(x)) - \varphi(x)) + \sum_{i=1}^{m} \frac{(||f(x)-a_i|| - ||x-a_i||)^2}{||x-a_i||} \\ &< \varphi(x) \end{aligned}$$

故

$$\varphi(f(x)) < \varphi(x) - \frac{1}{2}\sum_{i=1}^{m} \frac{(||f(x)-a_i|| - ||x-a_i||)^2}{||x-a_i||} \le \varphi(x)$$

□

为了最后的讨论, 接下来我们先分别在是否为最优解的情形下讨论顶点处的性质.

命题 4.4. *(顶点的性质)* 记 $R_j = \sum_{i \neq j} \frac{a_i - a_j}{||a_i - a_j||}, \quad j = 1, 2, \cdots, m.$

(i) 顶点 a_j 是最优解, 当且仅当

$$||R_j|| \le 1 \tag{7}$$

(ii) 记 $f^s(x) = f(f^{s-1}(x)), f^0(x) = x$. 若 a_j 不是最优解, 则 $\exists \delta > 0, s \in \mathbb{N}_+, \forall x \in \mathbb{B}(a_j, \delta)$, 满足

$$||f^{s-1}(x) - a_j|| \le \delta \quad 且 \quad ||f^s(x) - a_j|| > \delta \tag{8}$$

证明. (i) 由命题2.8和命题3.1知顶点 a_j 是最优解, 当且仅当

$$0 \in \partial\varphi(a_j) = -R_j + \mathbb{B}$$

故

$$||R_j|| \le 1$$

(ii) 对 $\forall x \notin \{a_1, a_2, \cdots, a_m\}$ 有

$$f(x) - a_j = \frac{\sum\limits_{i \ne j} \dfrac{a_i - a_j}{||x - a_i||}}{\sum\limits_{i=1}^{m} \dfrac{1}{||x - a_i||}}$$

于是

$$\lim_{x \to a_j} \frac{f(x) - a_j}{||x - a_j||} = \lim_{x \to a_j} \frac{\sum\limits_{i \ne j} \dfrac{a_i - a_j}{||x - a_i||}}{1 + \sum\limits_{i \ne j} \dfrac{||x - a_j||}{||x - a_i||}} = R_j$$

$$\lim_{x \to a_j} \frac{||f(x) - a_j||}{||x - a_j||} = ||R_j|| < \infty \tag{9}$$

由式7知 $||R_j|| > 1$, 故 $\forall \varepsilon \in (0, ||R_j|| - 1), \exists \delta > 0, \forall x \in \mathbb{B}(a_j, \delta)$, 有

$$||f(x) - a_j|| > (1 + \varepsilon)||x - a_j||$$

所以

$$||f^s(x) - a_j|| > (1 + \varepsilon)||f^{s-1}(x) - a_j|| > \cdots > (1 + \varepsilon)^s ||x - a_j||$$

当 $s \to \infty$ 时,$||f^s(x) - a_j|| \to \infty$, 于是 $\exists s \in \mathbb{N}_+$ 使得式8成立.

□

最后再证明序列收敛得到的极限 $\bar{x}$ 为 Fermat-Torricelli 问题的唯一解, 我们就成功验证了 Weiszfeld 算法.

命题 4.5. *(序列极限为最优解)* *设* $x_n \notin \{a_1, a_2, \cdots, a_m\}, n \in \mathbb{N}_+$, *则* $\{x_n\}$ *收敛于* $\bar{x}$, *即为问题2的唯一解.*

证明. 由命题2.7已知最优解 $\bar{x}$ 的存在性和唯一性.

1) 若 $\exists n = n_0$, 满足 $x_{n_0} = x_{n_0+1}$. 则有 $f(x_{n_0}) = x_{n_0}$, 且当 $n \geq n_0$ 时, $x_n = x_{n_0}$, 所以 $\lim\limits_{n\to\infty} x_n = x_{n_0} = \bar{x}$ 是最优解.

2) 设 $x_{n+1} \neq x_n, \forall n \in \mathbb{N}_+$. 由命题4.3知 $\varphi(x)$ 单调递减且 $\varphi(x) > 0$, 故 $\varphi(x)$ 收敛, 故有
$$\lim_{n\to\infty} (\varphi(x_{n+1}) - \varphi(x_n)) = 0 \tag{10}$$
而由序列 $\{x_n\}$ 的构造知
$$x_n = \sum_{i=1}^{m} \frac{\frac{1}{||x-a_i||}}{\sum\limits_{i=1}^{m} \frac{1}{||x-a_i||}} a_i \in \mathrm{conv}\{a_1, a_2, \cdots, a_m\}$$
由命题4.2知 $\mathrm{conv}\{a_1, a_2, \cdots, a_m\}$ 有界, 故由定理4.1知存在收敛子列 $\{x_{k_n}\} \subset \{x_n\}$, $\lim\limits_{n\to\infty} x_{k_n} = \bar{z}$. 又由式10有
$$\lim_{n\to\infty} (\varphi(x_{k_{n+1}}) - \varphi(x_{k_n})) = 0$$
根据 $\varphi(x)$ 的连续性, 得到
$$\varphi(\bar{z}) = \varphi(f(\bar{z})) \tag{11}$$
   (i) 若 $\bar{z} \notin \{a_1, a_2, \cdots, a_m\}$, 假设 $\bar{z} \neq f(\bar{z})$, 则由命题4.3得 $\varphi(f(\bar{z})) < \varphi(\bar{z})$, 与式11矛盾, 因此 $\bar{z} = f(\bar{z})$. 由6知 $\bar{z} = \bar{x}$ 是最优解.

   (ii) 若 $\bar{z} \in \{a_1, a_2, \cdots, a_m\}$, 不妨设 $\bar{z} = a_1$.
   我们假设 $\bar{z} \neq \bar{x}$, 那么就 $\exists \delta > 0$, 保证 $\bar{x}, a_2, a_3, \cdots, a_m \notin \mathbb{B}(a_1, \delta)$. 由命题4.4知 $\forall x \in \mathbb{B}(a_1, \delta), \exists s \in \mathbb{N}_+$, 使得
$$||f^{s-1}(x) - a_1|| \leq \delta, \quad ||f^s(x) - a_1|| > \delta \tag{12}$$
   不失一般性, 设 $\{x_{k_n}\} \subset \mathbb{B}(a_1, \delta)$. 如式12所示, 令 $x = x_{k_1}$, 可构造出 x_{l_1} 满足
$$l_1 \geq k_1, \quad x_{l_1} \in \mathbb{B}(a_1, \delta), \quad f(x_{l_1}) \notin \mathbb{B}(a_1, \delta)$$
   再取 $x = x_{k_{l_1+1}}$, 可构造 x_{l_2} 满足
$$l_2 \geq k_{l_1+1} > l_1, \quad x_{l_2} \in \mathbb{B}(a_1, \delta), \quad f(x_{l_2}) \notin \mathbb{B}(a_1, \delta)$$

以此类推, 可以构造序列 $\{x_{l_n}\} \subset \{x_n\}$ 满足

$$x_{l_n} \in \mathbb{B}(a_1, \delta), \quad f(x_{l_n}) \notin \mathbb{B}(a_1, \delta), \quad \forall n \in \mathbb{N}_+$$

于是再由定理4.1知存在收敛子列 $\{x_{l_{u_n}}\} \subset \{x_{l_n}\}, \lim_{n\to\infty} x_{l_{u_n}} = \bar{w}$. 满足

$$||\bar{w} - a_1|| \leq \delta, ||f(\bar{w}) - a_1|| \geq \delta$$

类比式11对 $\bar{z}$ 的推导亦有

$$\varphi(\bar{w}) = \varphi(f(\bar{w})) \tag{13}$$

同理对 $\bar{w}$ 进行讨论

(a) 若 $\bar{w} \notin \{a_1, a_2, \cdots, a_m\}$, 则 $\bar{w} = f(\bar{w})$ 是最优解.
但是 $\bar{w} \in \mathbb{B}(a_1, \delta), \bar{x} \notin \mathbb{B}(a_1, \delta)$, 于是 $\bar{w} \neq \bar{x}$, 这和最优解的唯一性矛盾.

(b) 若 $\bar{w} \in \{a_1, a_2, \cdots, a_m\}$, 那么由 $\bar{x}, a_2, a_3, \cdots, a_m \notin \mathbb{B}(a_1, \delta)$ 可知 $\bar{w} = a_1$.
所以有

$$\lim_{x\to a_1} \frac{||f(x) - a_1||}{||x - a_1||} = \lim_{n\to\infty} \frac{||f(x_{l_{u_n}}) - a_1||}{||x_{l_{u_n}} - a_1||} = \infty$$

这与式9矛盾.

故假设不成立,$\bar{z} = \bar{x}$ 是最优解.

□

5 缺陷与不足

本文从中学时期遇到的问题出发, 对于数学分析课本中的凸函数之定义和应用稍加拓展, 用来解决一个常见的问题. 文章中仍有许多缺陷. 第一, 对于点 a_i $(i = 1, 2, \cdots, m,\ m > 3)$ 共线时的情形并没有讨论; 第二, 仅研究了凸函数的一小部分应用, 对于次梯度算法等求解问题的一般化方法并没有深入探究; 第三, 解决问题的方法有局限性, 并没有详细说明所使用方法的缘由, 仅是对其加以验证; 第四, 对 Fermat-Torricelli 问题的实际应用并没有加以说明. 最后, 欢迎读者批评指正.

References

[1] Boyd S, Boyd S P, Vandenberghe L. Convex optimization[M]. Cambridge university press, 2004.

[2] Mordukhovich B S, Nam N M. An easy path to convex analysis and applications[J]. Synthesis Lectures on Mathematics and Statistics, 2013, 6(2): 1-218.

[3] Krantz S G. Convex analysis[M]. CRC Press, 2014.

[4] 常庚哲, 史济怀. 数学分析教程 [M]. 北京: 高等教育出版社, 2003.

[5] E. Weiszfeld, "Sur le point pour lequel la somme des distances de n points donnés est minimum," Tohoku Mathematics Journal, vol. 43, pp. 355-386, 1937.

[6] H. W. Kuhn, "A note on Fermat-Torricelli problem," Math. Program., vol. 54, pp. 98-107, 1973.