

# 通过镜头看广义Heron优化-极小化问题

Eric C. Chi Kenneth Lange

**摘要** 最近一期《美国数学月刊 (Amer. Math. Monthly)》上刊登了 Mordukhovich, Nam 和 Salinas 的一篇文章,<sup>1)</sup> 他们在现代凸分析的框架下提出并解决了一个有趣的不可微广义经典 Heron (海伦) 问题, 并作出了推广 [22]. 广义 Heron 问题是对赋以 Euclid (欧几里得) 范数的  $\mathbb{R}^d$  中给定的  $k+1$  个闭凸集, 要求在最后一个集合中寻找这样的点以至于它到前  $k$  个集合距离的总和是最小的. 在后续工作中作者更进一步推广了 Heron 问题, 放宽其凸的假设, 研究其理论特性, 并对解凸集提出了次梯度算法. 在这里, 我们先从数值角度重温原始问题. 通过利用计算统计中优化-极小化 (majorization-minimization, 简称 MM) 原则和微积分的基本技术, 我们能够对求解的 Euclid 式的广义 Heron 问题构建一个非常快的算法.

## 1. 引言

本刊最近刊登了 Mordukhovich 等 3 人的一篇文章. 他们对经典 Heron 问题作出了如下推广 [22]. 在  $\mathbb{R}^d$  中给定闭凸集  $\{C_1, \dots, C_k\}$ , 在闭凸集  $S \subset \mathbb{R}^d$  中寻找点  $\mathbf{x}$ , 使得从  $\mathbf{x}$  到  $C_1, \dots, C_k$  的 Euclid 距离总和最小. 即,

$$\text{minimize } D(\mathbf{x}) := \sum_{i=1}^k d(\mathbf{x}, C_i) \quad \text{使服从 } \mathbf{x} \in S, \quad (1)$$

其中  $d(\mathbf{x}, \Omega) = \inf\{\|\mathbf{x} - \mathbf{y}\| : \mathbf{y} \in \Omega\}$ .

这个问题的提出由来已久. 最初是从解决一些特殊情况开始的. 当  $k=2$  时, 若  $C_1$  和  $C_2$  是单元素集,  $S$  是一条线, 这是最初由古代数学家 Alexandria 的 Heron 提出的问题. 对特殊情况  $k=3$ , 如  $C_1, C_2$  和  $C_3$  是单元素集,  $S = \mathbb{R}^2$ , 这个问题是 Fermat (Fermat) 在 400 年前提出并由 Torricelli (托里拆利) [13] 解决的. 在他的论文“流数的理论与应用”中, Simpson (辛普森) 将距离推广到加权距离. 在 19 世纪, Steiner (施泰纳) 在这方面作出了一些基础性的贡献, 他的名字有时与 [9, 11] 中的问题相联系. 到 20 世纪, 德国经济学家 Weber 把 Fermat 问题推广到任意数量的单元素集  $C_i$ . 在 1937 年 Weiszfeld 发表了解决 Fermat-Weber 问题 [28, 29] 的第一个迭代算法<sup>2)</sup>. 在当今时代, 在各种计算名目下, Fermat-Weber 问题得到更大的复兴. 这个问题以及相关的算法已成为定位理论中许

---

译自: The Amer. Math. Monthly, Vol.121 (2014), No.2, p.95-108, A Look at the Generalized Heron Problem through the Lens of Majorization-Minimization, Eric C. Chi and Kenneth Lange, figure number 3. Copyright ©2014 the Mathematical Association of America. Reprinted with permission. All rights reserved. 美国数学协会授予译文出版许可.

Eric C. Chi 于 2011 年在 David W. Scott 的指导下获得博士学位, 目前是洛杉矶加州大学人类遗传学系的博士后. 他的邮箱地址是 [ecchi@ucla.edu](mailto:ecchi@ucla.edu).

Kenneth Lange 是洛杉矶加州大学计算遗传学的 Rosenfeld 讲座教授, 也是人类遗传学系主任和生物数学与统计学的教授. 他的邮箱地址是 [klange@ucla.edu](mailto:klange@ucla.edu)

1) 见本刊 2012 年第 3 期文“凸分析方法解广义海伦问题”.——编注

2) Kuhn [15] 指出, Weiszfeld 的算法早已被重新发现了好几次.——原注

多高级模型 [18, 30] 的出发点.

著名的 Fermat-Weber 和广义 Heron 问题之间的关系早先是由 Mordukhovich 等指出的 [23]. 在随后的论文 [21, 23] 中, 他们将 Heron 问题进一步推广到在 Banach (巴拿赫) 空间的任意闭集,  $C_1, \dots, C_k$  和  $S$ . 关于如何利用最先进的变分分析工具来解决抽象形式的广义 Heron 问题, 读者可参考他们的论文.

这里, 我们只研究由 Mordukhovich 等人在 [23] 中提出的 Euclid 距离的特殊情况. 我们的目的是从算法设计的角度再次看一下这个简单, 但在我们看来, 又最中肯的问题. Mordukhovich 等人在文章 [21, 22, 23] 中提出了一种迭代次梯度算法去数值求解问题 (1) 及其推广形式. 关于求解的问题, 他们只保留目标函数的凸性的假设, 在我们看来这是一个稳健的选择. 事实上, 当 Euclid 范数替换成任意范数时, 次梯度算法照常适用. 然而, 一个很自然的问题是, 如果仅限于赋以 Euclid 距离的有限维问题, 是否有可能有更好的选择. 在这里, 我们用计算统计中的优化 - 极小化 (MM) 原则提出了一个推广的 Weiszfeld 算法. 虽然新算法具有同样的困扰 Weiszfeld 算法的奇异性 [15], 该难题可以通过微扰动问题 (1) 得到解决. 在本文的其余部分, 我们称之为广义 Heron 问题. 在极限情况下, 我们重新回到未扰动问题. 正如预期的那样, 利用一个问题的特殊结构是值得的. 对有 Euclid 距离的问题, 新的 MM 算法在计算速度上大大优于次梯度算法.

用 MM 原理来解释扰动问题还有其他好处. 光滑问题的 MM 算法的收敛性在理论上是公认的. 这一事实使我们能够不借助于全套的不可微函数的凸分析处理以及相应的次梯度算法工具, 就可以特征化原问题的解. 虽然这个数学知识本身是肯定值得学习的, 值得注意的是简单的工具可取得很大的进步. 可喜的是我们所展示的数值求解问题 (1) 的各具特色的迭代技巧是在经典的微积分范围之内. 我们的方案只要求本科或工科数学专业知识就能理解.

简短总结以下章节的安排, 我们首先回顾 MM 原理的背景材料和可微函数的凸分析. 其次是提出了针对问题 (1) 的 MM 算法的推导, 和一些相关数值例子. 最后我们证明了算法的收敛性并描述了解的特征.

## 2. MM 的原理

虽然 MM 原理是数值分析家 Ortega 和 Rheinboldt 首先提出的 [24], 但目前却在计算统计中是非常流行的 [1, 17]. 其基本思想是将一个困难的优化问题 (例如, 非可微) 变成一系列简单的问题 (例如, 平滑). MM 的原理要求通过一个替代函数  $g(\mathbf{y} | \mathbf{x})$  在给定的点  $\mathbf{x}$  处来优化目标函数  $f(\mathbf{y})$ . 优化是对所有的  $\mathbf{y} \in \mathbb{R}^d$ , 相切条件  $g(\mathbf{x} | \mathbf{x}) = f(\mathbf{x})$  和控制条件  $g(\mathbf{y} | \mathbf{x}) \geq f(\mathbf{y})$  的组合. 相关的 MM 算法是由下面的迭代来定义的

$$\mathbf{x}_{k+1} := \arg \min_{\mathbf{y} \in S} g(\mathbf{y} | \mathbf{x}_k). \quad (2)$$

因为

$$f(\mathbf{x}_{k+1}) \leq g(\mathbf{x}_{k+1} | \mathbf{x}_k) \leq g(\mathbf{x}_k | \mathbf{x}_k) = f(\mathbf{x}_k), \quad (3)$$

MM 的迭代产生一个下降算法, 驱动目标函数往下走. 寻找  $\mathbf{x}_{k+1}$ , 使得约束条件得到满足. 在适当的正则条件下, MM 算法保证收敛到原问题局部的最小值 [16].

### 3. 凸分析的背景

作为推导 MM 算法的前奏, 我们将从凸分析中的可微函数的范围内回顾一些基本事实. 进一步的处理可以在文献 [3, 4, 12, 25, 26] 中找到. 在这篇文章中, 我们用  $\langle \mathbf{a}, \mathbf{b} \rangle$  表示向量  $\mathbf{a}$  与  $\mathbf{b}$  之间的点积. 回想一下, 一个可微函数  $f(\mathbf{y})$  是凸的, 当且仅当其定义域  $S$  是凸的, 并对所有的  $\mathbf{x}, \mathbf{y} \in S$ , 有

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle. \quad (4)$$

在  $f(\mathbf{x})$  二次可微条件下, 当  $d^2 f(\mathbf{x})$  对所有的  $\mathbf{x}$  为正半定时, 它是凸的; 当  $d^2 f(\mathbf{x})$  对所有的  $\mathbf{x}$  为正定时, 它是严格凸的. 这些特征是  $f(\mathbf{y})$  二阶 Taylor (泰勒) 展开并应用支撑超平面不等式 (4) 的一个直接结果. 支撑超平面不等式 (4) 也导致了关于全局最小值的一个简洁的充要条件. 点  $\mathbf{x} \in S$  是  $f(\mathbf{y})$  在  $S$  上的全局极小点, 当且仅当对所有  $\mathbf{y} \in S$

$$\langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \geq 0. \quad (5)$$

直观地说, 每一个方向指向  $S$  必须是走上坡的.

我们通过回顾投影算子 [16] 来结束本节. 用  $P_\Omega(\mathbf{x})$  表示  $\mathbf{x}$  的投影到一个集  $\Omega \subset \mathbb{R}^d$ . 由定义,  $P_\Omega(\mathbf{x})$  满足

$$P_\Omega(\mathbf{x}) := \arg \min_{\mathbf{y} \in \Omega} \|\mathbf{x} - \mathbf{y}\|.$$

如果  $\Omega$  是  $\mathbb{R}^d$  的闭凸集, 则  $P_\Omega(\mathbf{x})$  存在且唯一. 进一步, 投影算子对所有的  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$  有

$$\|P_\Omega(\mathbf{x}) - P_\Omega(\mathbf{y})\| \leq \|\mathbf{x} - \mathbf{y}\|,$$

即它是非扩张的. 显然, 非扩张必需是连续的. 当  $\Omega$  是一个盒子, Euclid 球, 超平面, 或半空间时, 投影算子  $P_\Omega(\mathbf{x})$  有显式的公式. 对单位单纯形,  $\ell_1$  球, 半正定矩阵锥等, 存在快速计算  $P_\Omega(\mathbf{x})$  的算法 [10, 20].

投影算子和距离函数通过梯度恒等式  $\nabla d(\mathbf{x}, C)^2 = 2[\mathbf{x} - P_C(\mathbf{x})]$  而密切相关. 这个事实的一个标准的证明可以在文献 [12, p. 181] 中找到. 若  $d(\mathbf{x}, C)^2 > 0$ , 则链式法则给出

$$\nabla d(\mathbf{x}, C) = \nabla \sqrt{d(\mathbf{x}, C)^2} = \frac{\mathbf{x} - P_C(\mathbf{x})}{d(\mathbf{x}, C)}.$$

在  $C$  的内部, 很明显  $\nabla d(\mathbf{x}, C) = \mathbf{0}$ . 相反, 在  $C$  的边界点不能保证  $d(\mathbf{x}, C)$  的可微性.

### 4. Heron 问题的一个 MM 算法

因为它增加了一点额外的计算, 我们将问题 (1) 改写为 Simpson 形式.

$$\text{minimize } D(x) := \sum_{i=1}^k \gamma_i d(\mathbf{x}, C_i) \quad \text{使服从 } \mathbf{x} \in S, \quad (6)$$

正如 [23] 所建议的, 问题 (6) 涉及到正的权重  $\gamma_i$  与距离  $d(\mathbf{x}, C_i)$  的凸组合. 我们首先推导, 对于所有的  $i$ , 当  $S \cap C_i = \emptyset$  时解问题 (6) 的 MM 算法. 这次练习将为我们以后处理更一般的情况 (如  $S$  与一个或多个  $C_i$  相交) 打下基础. 在实践中二次优化不等式是可取的, 因为它可以使 MM 算法的极小化步骤达到精确解. 在我们的设置中需要两个相继的优化不等式来得到一个二次优化不等式. 直接从距离函数的定义可得到第 1 个简单的优化不等式

$$d(\mathbf{x}, C_i) \leq \|\mathbf{x} - P_{C_i}(\mathbf{x}_m)\|.$$

第 2 个是凹函数  $\sqrt{u}$  在区间  $(0, \infty)$  上的优化不等式

$$\sqrt{u} \leq \sqrt{u_m} + \frac{1}{2\sqrt{u_m}}(u - u_m). \quad (7)$$

这两个优化不等式的组合产生一个二次优化不等式

$$d(\mathbf{x}, C_i) \leq \|\mathbf{x}_m - P_{C_i}(\mathbf{x}_m)\| + \frac{\|\mathbf{x} - P_{C_i}(\mathbf{x}_m)\|^2 - \|\mathbf{x}_m - P_{C_i}(\mathbf{x}_m)\|^2}{2\|\mathbf{x}_m - P_{C_i}(\mathbf{x}_m)\|}. \quad (8)$$

对这些优化不等式关于  $i$  求和产生了  $D(\mathbf{x})$  的一个二次优化不等式和最终的 MM 算法映射

$$\psi(\mathbf{x}) = \arg \min_{\mathbf{z} \in S} \left\{ \frac{1}{2} \sum_{i=1}^k w_i \|\mathbf{z} - P_{C_i}(\mathbf{x})\|^2 \right\},$$

其中权重  $w_i = \gamma_i \|\mathbf{x} - P_{C_i}(\mathbf{x})\|^{-1}$ . 当  $C_i$  是单元素集且  $S = \mathbb{R}^d$  时, 映射  $\psi(\mathbf{x})$  就是解 Fermat-Weber 问题的 Weiszfeld 算法 [28, 29].

刚才导出的  $D(\mathbf{x})$  的二次优化不等式可改写为

$$g(\mathbf{x} | \mathbf{x}_m) = \frac{1}{2} \left( \sum_{i=1}^k w_i \right) \left\| \mathbf{x} - \sum_i \alpha_i P_{C_i}(\mathbf{x}_m) \right\|^2 + c,$$

其中

$$\alpha_i = \frac{w_i}{\sum_{i=1}^k w_i}$$

依赖于权重  $w_i = \gamma_i \|\mathbf{x}_m - P_{C_i}(\mathbf{x}_m)\|^{-1}$ ,  $c$  是不依赖于  $\mathbf{x}$  的常数. 因此, MM 更新可简化为把前几次的迭代投影到集合  $C_i$  上的凸组合, 再投影到  $S$  上; 用记号表示, 即为

$$\mathbf{x}_{m+1} = P_S \left[ \sum_i \alpha_i P_{C_i}(\mathbf{x}_m) \right]. \quad (9)$$

当  $\mathbf{x}_m$  属于  $C_i$  时, 优化不等式 (8) 涉及到除以 0 的问题. 这奇点也同样困扰 Weiszfeld 算法. 幸运的是, 目标函数的扰动挽救了局面. 对小的正的  $\epsilon$ , 用相关函数

$$D_\epsilon(\mathbf{x}) = \sum_{j=1}^k \gamma_j \sqrt{d(\mathbf{x}, C_j)^2 + \epsilon}$$

简单地替换函数  $D(\mathbf{x})$ . Ben-Tal 和 Teboulle 在文献 [2] 中给出了这种扰动策略的进一步例子. 在任何情况下, 观察到光滑函数  $f_\epsilon(u) = \sqrt{u^2 + \epsilon}$  有导数

$$f'_\epsilon(u) = \frac{u}{\sqrt{u^2 + \epsilon}}, \quad f''_\epsilon(u) = \frac{\epsilon}{(u^2 + \epsilon)^{3/2}},$$

所以在区间  $[0, \infty)$  上  $f_\epsilon$  是严格递增和严格凸的. 因此, 函数  $D_\epsilon(\mathbf{x})$  也是凸的. 因为  $\sqrt{u^2 + \epsilon} - \sqrt{\epsilon}$  很好地逼近了  $u \geq 0$ , 这两个问题的解应该是很接近的. 事实上, 我们稍后会指出  $D_\epsilon(\mathbf{x})$  的最小值点趋于  $D(\mathbf{x})$  的最小值点, 当  $\epsilon$  趋向于 0 时. 在存在多个极小值的情况下, 这种说法必须被改写为聚点.

围绕当前迭代  $\mathbf{x}_m$  的优化不等式  $d(\mathbf{x}, C_j) \leq \|\mathbf{x} - P_{C_j}(\mathbf{x}_m)\|$  产生了优化不等式

$$\sqrt{d(\mathbf{x}, C_j)^2 + \epsilon} \leq \sqrt{\|\mathbf{x} - P_{C_j}(\mathbf{x}_m)\|^2 + \epsilon}.$$

应用优化公式 (7) 意味着进一步优化

$$D_\epsilon(\mathbf{x}) \leq \frac{1}{2} \sum_{j=1}^k \gamma_j \frac{\|\mathbf{x} - P_{C_j}(\mathbf{x}_m)\|^2}{\sqrt{\|\mathbf{x}_m - P_{C_j}(\mathbf{x}_m)\|^2 + \epsilon}} + c,$$

其中  $c$  是一个不相关的常数. 除了一个差别, 相应的 MM 更新  $\mathbf{x}_{m+1}$  同以前的 MM 的更新 (9) 是一样的. 这个差别在于权重  $w_j$  现在定义为好的不涉及奇点的公式

$$w_j = \frac{\gamma_j}{\sqrt{\|\mathbf{x}_m - P_{C_j}(\mathbf{x}_m)\|^2 + \epsilon}}.$$

## 5. 一些例子

下面考虑用 4 个例子说明 MM 算法的性能和构想我们预期的收敛. 在整个比较中, 次梯度算法 [22] 是作为一个基准. 该算法依赖于更新

$$\mathbf{x}_{m+1} = P_S \left[ \mathbf{x}_m - \eta_m \sum_{i=1}^k \gamma_i \mathbf{v}_{im} \right],$$

其中

$$\mathbf{v}_{im} = \begin{cases} \frac{\mathbf{x}_m - P_{C_i}(\mathbf{x}_m)}{d(\mathbf{x}_m, C_i)} & \text{若 } \mathbf{x}_m \notin C_i, \\ 0, & \text{若 } \mathbf{x}_m \in C_i. \end{cases}$$

非负常数  $\eta_m$  满足  $\sum_{m=1}^{\infty} \eta_m = \infty$  和  $\sum_{m=1}^{\infty} \eta_m^2 < \infty$ . 在所有的例子 (最后一个除外) 中权重  $\gamma_i$  都等于 1.

### 5.1 $\mathbb{R}^3$ 中 5 个立方体和球的例子

第 1 个例子是取自参考文献 [22]. 此三维例子涉及边长等于 2, 中心为  $(0, -4, 0)$ ,  $(-4, 2, -3)$ ,  $(-3, -4, 2)$ ,  $(-5, 4, 4)$  和  $(-1, 8, 1)$  的 5 个立方体  $C_i$ . 集合  $S$  是一个中心为  $(0, 2, 0)$ , 半径为 1 的球. 迭代开始在点  $\mathbf{x}_1 = (0, 2, 0) \in S$  和用步长为  $\eta_m = 1/m$  的次梯度算法. 表 1 显示了用  $\epsilon = 0$  的 MM 迭代. 在迭代 30 次范围内收敛到机器精度. 作为对比, 表 2 表明参数值  $(x_1, x_2, x_3)$  在次梯度迭代  $10^6$  次后仍在变化. 为了简便起见, 我们省略了取自同一文献 [22] 中的第 2 个例子, 即在  $\mathbb{R}^2$  中的 4 个正方形和一个圆盘的例子. 在这个例子中, MM 算法优于次梯度算法的优越性同样非常明显.

表 1  $\mathbb{R}^3$  中立方体和球的例子: MM 算法

迭代	$x_1$	$x_2$	$x_3$
1	0.00000000000000	2.00000000000000	0.00000000000000
2	-0.93546738305698	1.66164748416805	0.10207032020482
3	-0.92881282698649	1.63915389878166	0.08424264751830
4	-0.92645373003448	1.63220797263449	0.08007815377225
5	-0.92567602259658	1.63004821970935	0.07911751670489
6	-0.92542515217106	1.62937435413374	0.07889815178685
7	-0.92534495711879	1.62916364685109	0.07884864943702
8	-0.92531944712805	1.62909766226627	0.07883765997470
9	-0.92531135783449	1.62907697582185	0.07883527888603
10	-0.92530879826106	1.62907048520349	0.07883478238381
20	-0.92530761702316	1.62906751412014	0.07883466748783
30	-0.92530761701184	1.62906751409212	0.07883466748878
50	-0.92530761701184	1.62906751409212	0.07883466748878

表 2  $\mathbb{R}^3$  中立方体和球的例子: 次梯度算法

迭代	$x_1$	$x_2$	$x_3$
1	0.000000000000000	2.000000000000000	0.000000000000000
10	-0.92583298353433	1.63051788239768	0.07947484741743
100	-0.92531325048300	1.62908232435160	0.07883822912883
1000	-0.92530767419684	1.62906766065418	0.07883468589312
10000	-0.92530761758555	1.62906751554109	0.07883466757273
100000	-0.92530761701755	1.62906751410641	0.07883466748904
1000000	-0.92530761701233	1.62906751409334	0.07883466748881
1500000	-0.92530761701231	1.62906751409328	0.07883466748881
2000000	-0.92530761701229	1.62906751409324	0.07883466748881

### 5.2 在 $\mathbb{R}^2$ 中到 3 个圆盘的最近点

这个例子取自文献 [21], 它说明了用一系列逼近函数  $D_{\epsilon_m}(\mathbf{x})$  的极小化的优点. 集合  $C_i$  是在  $\mathbb{R}^2$  3 个中心为  $(0, 2)$ ,  $(2, 0)$  和  $(-2, 0)$  的单位球; 集合  $S$  等于  $\mathbb{R}^2$ . 最小距离出现在  $(0, 1)$ , 这很容易通过检查文献 [21] 命题 4.3 的优化条件来验证. 图 1 显示了 50 个不同的出发 (点) 值和它们相应的迭代路径的不动点 (正方形). 沿着  $m$  条路径, 我们令  $\epsilon_m$  为  $\max\{10^{-m}, 10^{-16}\}$ . 当前的问题的解被取为下一个问题的初始点. 所有的解路径开始收敛到  $(0, 1)$  下面的一个点, 然后集体向  $(0, 1)$  前进. 利用系统地降低  $\epsilon$  的策略, 单位球的 MM 迭代通道变得很容易. 表 3 显示了从点  $(5, 7)$  开始的次梯度和 MM 迭代算法.

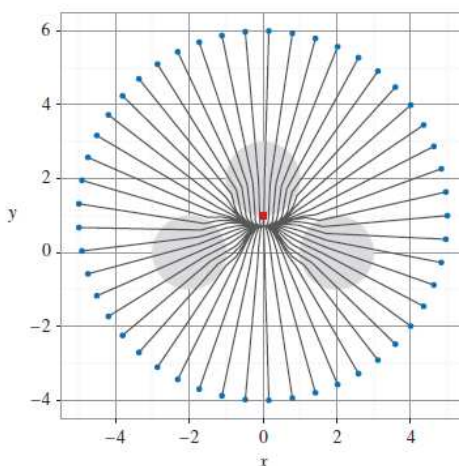


图 1 在  $\mathbb{R}^2$  中寻找找到 3 个圆盘的最近点

表 3  $\mathbb{R}^2$  中从点  $(5, 7)$  出发的 3 个圆盘例子

次梯度算法			MM 算法		
迭代	$x_1$	$x_2$	迭代	$x_1$	$x_2$
10	0.7092649	1.2369866	10	0.2674080	0.7570688
100	0.0558764	0.9973310	100	0.0000000	0.7249706
1,000	0.0046862	0.9993844	1,000	0.0000000	0.9998002
10,000	0.0003955	0.9999274	1,800	0.0000000	0.9999999
100,000	0.0000334	0.9999957	1,850	0.0000000	1.0000000
1,000,000	0.0000028	0.9999998	1,900	0.0000000	1.0000000

### 5.3 在 $\mathbb{R}^2$ 中的三点共线圆盘

在这里, 我们举例说明当有多于一个以上的解时 MM 算法的行为. 考虑  $\mathbb{R}^2$  上中心为  $(2, 0)$  和  $(-2, 0)$  两个单位球, 并取  $S$  是中心为原点的单位球. 沿着线段从  $(-1, 0)$  延伸到  $(1, 0)$ , 有一系列的解, 这可由文献 [22] 提供的定理 3.2 的最优性条件来检验. 图 2 显示出了 100 次不同的初值 (点) 迭代路径和其相应的不动点 (正方形). 在这个例子中,

我们取  $\epsilon = 0$ . 虽然迭代不能保证收敛, 原则上可以在多个聚点之间循环. 这种行为是不能在实践中观察到的. 依赖于他们的出发点, 简单迭代收敛到不同的不动点. 表 4 比较了次梯度方法和 MM 算法由点  $(1.5, 0.25)$  出发的迭代结果. 两种算法收敛到不同的点, 但完全不同的速度.

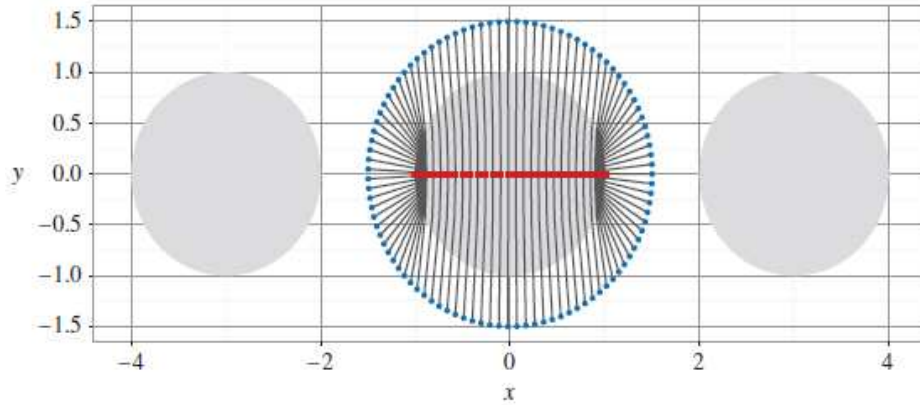


图 2 解的连续统的例子

表 4  $\mathbb{R}^2$  中三点共线圆盘从点  $(1.5, 0.25)$  出发的例子

次梯度算法			MM 算法		
迭代	$x_1$	$x_2$	迭代	$x_1$	$x_2$
10,000	0.9997648	0.0000223	10	0.9941149	0.0001308
100,000	0.9997648	0.0000040	20	0.9941149	0.0000000
1,000,000	0.9997648	0.0000007	30	0.9941149	0.0000000

#### 5.4 Kuhn (库恩) 问题

我们最后一个例子最初是由 Kuhn 设计的 [14], 来说明当迭代进入  $C_i$  之一时, Weiszfeld 的算法会如何停止. 虽然这种事件在实践中很少发生, 激发它发生的初始条件的特征一直是文献 [5, 6, 7, 8, 15] 强烈关注的对象. Weiszfeld 算法的偶然的失败驱使 Vardi 和 Zhang 等人 [27] 重新设计算法. 他们的算法保持了下降性, 但与我们的算法有显著的不同. 在任何情形下, 图 3 所示的示例中涉及到权重  $\gamma_i$  比例为 5 的两个点放置在  $(59, 0)$  和  $(20, 0)$ ; 两个权重比例为 13 的点, 放置在  $(-20, 48)$  和  $(-20, -48)$ . 最佳点是原点. Weiszfeld 算法开始于  $(44, 0)$ , 一次迭代后停止在  $(20, 0)$ . 相反, 我们的 MM 迭代 (点), 在  $\epsilon$  从 0.1

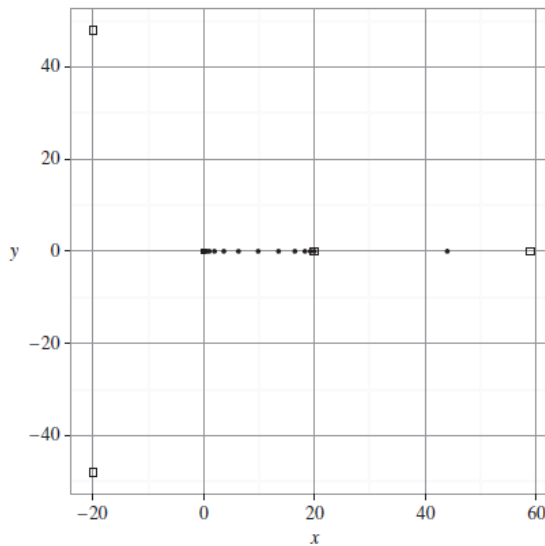


图 3 Weiszfeld 的算法不收敛的一个问题

降到 0 的情况下, 移动跨越了 (20.0), 在机器精度内用 99 步正确地收敛到 (0, 0). 表 5 比较了 MM 算法和次梯度方法进行的情况. 需要注意的是, 当  $\epsilon$  是 0.1 的时候, MM 算法冲过真正的答案, 然后在设置  $\epsilon$  为 0 之后, 回到 (0, 0). 对这个几乎平滑的问题, 次梯度算法早期扎实推进, 但其后会大幅减慢.

表 5 Kuhn 问题

次梯度算法			MM 算法		
迭代	$x_1$	$x_2$	迭代	$x_1$	$x_2$
10	8.6984831	0.0000000	10	1.9448925	0.0000000
1,000	1.2966354	0.0000000	30	-0.0011998	0.0000000
100,000	0.1845171	0.0000000	60	-0.0012011	0.0000000
10,000,000	0.0259854	0.0000000	90	0.0000000	0.0000000

## 6. 收敛理论

在着手证明收敛性之前, 需要谨慎的考虑最小值是否存在且唯一. 回想一下, 一个连续函数在紧集上存在极小值. 因此, 当  $S$  为有界, 问题 (6) 有极小值. 如果  $S$  是无界的, 那么可以用 1 个或更多  $C_i$  有界集替代. 在此情况下  $D(\mathbf{x})$  是强制在这个意义上:  $\lim_{\|\mathbf{x}\| \rightarrow \infty} D(\mathbf{x}) = \infty$ . 正如文献 [22] 中命题 3.1 指出, 强制性足以保证存在性. 由于  $D(\mathbf{x}) \leq D_\epsilon(\mathbf{x})$ , 当原来的准则  $D(\mathbf{x})$  是强制的, 则扰动准则  $D_\epsilon(\mathbf{x})$  也是强制的. 此后, 我们将假定  $S$  或至少一个  $C_i$  是有界的.

严格凸函数在凸集上具有至多一个最小值点. 函数  $|x|$  表明对于唯一性的这个充分条件几乎不可能是必要的. 在 Fermat-Weber 问题中, 那里的闭凸集  $C_i = \{\mathbf{x}_i\}$  是单元素集, 该函数  $D(\mathbf{x})$  是严格凸的, 当且仅当点  $\mathbf{x}_i$  是不共线的. 为了推广这个结果, 我们要求集  $C_i$  是不共线的. 几何上来说, 画一条直线穿过所有的  $C_i$  是不可能的. 当  $k > 2$  和  $\bigcap_{i=1}^k C_i = \emptyset$  时, 非共线可能实现. 我们还要求  $C_i$  是严格凸的. 一个集合  $C$  被说成是严格凸的, 如果连接  $C$  中两个不同点  $\mathbf{x}$  和  $\mathbf{y}$  的线段  $[\mathbf{x}, \mathbf{y}]$  位于  $C$  的内部. 换句话说,  $C$  的边界不包含线段. 一单元素集或闭球是严格凸的, 但一个封闭的盒子不是严格凸的.

**命题 6.1** 如果闭凸集  $C_1, \dots, C_k$  是严格凸的, 但不共线, 则  $D(\mathbf{x})$  是严格凸的.

**证明** 假设相反的是真实的, 并选择  $\mathbf{x} \neq \mathbf{y}$  和  $\alpha \in (0, 1)$ , 使

$$D[\alpha\mathbf{x} + (1-\alpha)\mathbf{y}] = \alpha D(\mathbf{x}) + (1-\alpha)D(\mathbf{y}), \quad (10)$$

设  $L$  为经过点  $\mathbf{x}$  和  $\mathbf{y}$  的线  $\{s\mathbf{x} + (1-s)\mathbf{y} : s \in \mathbb{R}\}$ . 那么至少存在一个  $C_j$  使得  $L \cap C_j = \emptyset$ . 特别地,  $\mathbf{x}, \mathbf{y}$  和  $\alpha\mathbf{x} + (1-\alpha)\mathbf{y}$  都落在  $C_j$  之外. 等式 (10) 意味着

$$\begin{aligned} & \alpha\|\mathbf{x} - P_{C_j}(\mathbf{x})\| + (1-\alpha)\|\mathbf{y} - P_{C_j}(\mathbf{y})\| \\ &= \|\alpha\mathbf{x} + (1-\alpha)\mathbf{y} - P_{C_j}[\alpha\mathbf{x} + (1-\alpha)\mathbf{y}]\| \\ &\leq \|\alpha\mathbf{x} + (1-\alpha)\mathbf{y} - \alpha P_{C_j}(\mathbf{x}) - (1-\alpha)P_{C_j}(\mathbf{y})\| \\ &\leq \alpha\|\mathbf{x} - P_{C_j}(\mathbf{x})\| + (1-\alpha)\|\mathbf{y} - P_{C_j}(\mathbf{y})\|. \end{aligned}$$

由于点到  $C_j$  上的投影是唯一的, 这些夹在中间的不等式带来

$$P_{C_j}[\alpha\mathbf{x} + (1-\alpha)\mathbf{y}] = \alpha P_{C_j}(\mathbf{x}) + (1-\alpha)P_{C_j}(\mathbf{y}).$$



如果  $P_{C_j}(\mathbf{x}) \neq P_{C_j}(\mathbf{y})$ , 则  $C_j$  的严格凸性蕴涵着凸组合  $\alpha P_{C_j}(\mathbf{x}) + (1-\alpha)P_{C_j}(\mathbf{y})$  是  $C_j$  的内部. 因此, 这一点不能是到外点  $\alpha\mathbf{x} + (1-\alpha)\mathbf{y}$  最接近的点. 因此, 考虑  $P_{C_j}(\mathbf{x}) = P_{C_j}(\mathbf{y}) = \mathbf{z}$  的可能性. 等号可以发生在不等式中

$$\|\alpha\mathbf{x} + (1-\alpha)\mathbf{y} - \mathbf{z}\| \leq \alpha\|\mathbf{x} - \mathbf{z}\| + (1-\alpha)\|\mathbf{y} - \mathbf{z}\|$$

仅当对于某一  $t \neq 1$  有  $\mathbf{x} - \mathbf{z} = t(\mathbf{y} - \mathbf{z})$ . 这个关系表明,

$$\mathbf{z} = \frac{1}{1-t}\mathbf{x} - \frac{t}{1-t}\mathbf{y}$$

属于  $L \cap C_j$ , 与我们的假设矛盾. 因此,  $D(\mathbf{x})$  是严格凸的. ■

接下来的结果表明, 函数  $D_\epsilon(\mathbf{x})$  从  $D(\mathbf{x})$  继承了严格凸. 因此, 当  $D(\mathbf{x})$  是严格凸时,  $D_\epsilon(\mathbf{x})$  有唯一的最小值点.

**命题 6.2** 如果  $D(\mathbf{x})$  是严格凸的, 则  $D_\epsilon(\mathbf{x})$  也是严格凸的.

**证明** 固定任意  $\mathbf{x} \neq \mathbf{y}$  和  $\alpha \in (0, 1)$ . 严格凸  $D(\mathbf{x})$  意味着至少存在一个  $j$ , 使得

$$d(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}, C_j) < \alpha d(\mathbf{x}, C_j) + (1-\alpha)d(\mathbf{y}, C_j).$$

由于  $f_\epsilon(u) = \sqrt{u^2 + \epsilon}$  是一个严格递增和凸的函数, 即有严格不等式

$$\begin{aligned} \sqrt{d(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}, C_j)^2 + \epsilon} &< \sqrt{[\alpha d(\mathbf{x}, C_j) + (1-\alpha)d(\mathbf{y}, C_j)]^2 + \epsilon} \\ &\leq \alpha\sqrt{d(\mathbf{x}, C_j)^2 + \epsilon} + (1-\alpha)\sqrt{d(\mathbf{y}, C_j)^2 + \epsilon}. \end{aligned}$$

对  $j$  求和即得到所需的结果. ■

现在我们澄清函数  $D_\epsilon(\mathbf{x})$  和  $D(\mathbf{x})$  的极小值之间的关系.

**命题 6.3** 对于常数  $\epsilon_m$  趋于 0 的序列, 设  $\mathbf{y}_m$  是相应  $D_{\epsilon_m}(\mathbf{x})$  极小化的序列. 如果  $\mathbf{y}$  是  $D(\mathbf{x})$  唯一的最小值点,  $\mathbf{y}_m$  则趋向于  $\mathbf{y}$ . 如果  $D(\mathbf{x})$  具有多个极小, 那么每个序列  $\mathbf{y}_m$  的聚点是  $D(\mathbf{x})$  的极小化点.

**证明** 为了证明该结论, 对于任意的  $\mathbf{x} \in S$  和  $\epsilon_m \leq 1$  考虑不等式

$$D(\mathbf{y}_m) \leq D_{\epsilon_m}(\mathbf{y}_m) \leq D_{\epsilon_m}(\mathbf{x}) \leq D_1(\mathbf{x}).$$

沿着相应的子序列取极限证明了该序列  $\mathbf{y}_m$  的聚点也极小化  $D(\mathbf{x})$ . 因为所提供的序列  $\mathbf{y}_m$  是有界的, 它收敛到唯一的极小化值点  $\mathbf{y}$ . 若  $S$  是有界的, 则  $\mathbf{y}_m$  按定义是有界的.

另一方面, 如果任何  $C_j$  是有界的, 则  $D(\mathbf{x})$  是强制的, 而且不等式  $D(\mathbf{y}_m) \leq D_1(\mathbf{x})$  促使  $\mathbf{y}_m$  是有界的. ■

MM 的算法的收敛性理论取决于算法映射  $\psi(\mathbf{x}) \equiv \arg \min_{\mathbf{y}} g(\mathbf{y} | \mathbf{x})$  的性质. 为了便于参考, 我们陈述一个简单的 Meyer (梅耶) 单调收敛定理 [19], 它在我们提供收敛证明中起到了关键作用.

**命题 6.4** 设  $f(\mathbf{x})$  是域  $S$  上一个连续函数,  $\psi(\mathbf{x})$  是对所有的  $\mathbf{x} \in S$  且  $\psi(\mathbf{x}) \neq \mathbf{x}$  满足  $f(\psi(\mathbf{x})) < f(\mathbf{x})$  的从  $S$  到  $S$  中的连续算法映射. 假设对于某些初始点  $\mathbf{x}_0$ , 集合

$$\mathcal{L}_f(\mathbf{x}_0) \equiv \{\mathbf{x} \in S : f(\mathbf{x}) \leq f(\mathbf{x}_0)\}$$

是紧的. 则 (a) 所有聚点是  $\psi(\mathbf{x})$  的不动点, (b)  $\lim_{m \rightarrow \infty} \|\mathbf{x}_{m+1} - \mathbf{x}_m\| = 0$ .

注意, 命题 6.4 还确保迭代序列  $\mathbf{x}_{m+1} = \psi(\mathbf{x}_m)$  至少存在一个聚点. 此外, 如果  $\psi(\mathbf{x})$  的不动点是  $f(\mathbf{x})$  的一个驻点的话, 可以立刻推出 MM 的迭代序列 (9) 收敛到  $f(\mathbf{x})$  的一个驻点.  $\psi(\mathbf{x})$  拥有有限个不动点.

让我们验证命题 6.4 的条件来极小化  $D_\epsilon(\mathbf{x})$ . 由于  $S$  是紧的或  $D_\epsilon(\mathbf{x})$  是强制的, 函数  $D_\epsilon(\mathbf{x})$  在域  $S$  上是连续的, 并且集合  $\mathcal{L}_{D_\epsilon}(\mathbf{x}_0)$  对于任何初始点  $\mathbf{x}_0$  是紧的. 该算法的映射连续性可由投影映射的连续性得出. 最后, 当  $\mathbf{x} \neq \psi(\mathbf{x})$  时我们需要证明,  $D_\epsilon(\psi(\mathbf{x})) < D_\epsilon(\mathbf{x})$ . 首先观察到  $\psi(\mathbf{x}) = \mathbf{x}$  当且仅当 MM 替代函数满足  $g_\epsilon(\mathbf{x}|\mathbf{x}) = \min_{\mathbf{y}} g_\epsilon(\mathbf{y}|\mathbf{x})$ . 由于  $g_\epsilon(\mathbf{y}|\mathbf{x})$  有唯一的极小, 当  $\mathbf{x}$  不是  $\psi$  的不动点时, 不等式  $g_\epsilon(\psi(\mathbf{x})|\mathbf{x}) < g_\epsilon(\mathbf{x}|\mathbf{x})$  严格成立. 这迫使目标函数  $D_\epsilon(\mathbf{x})$  是下降的, 并使 MM 算法在驻点外严格单调.

我们现在讨论, 算法映射  $\psi(\mathbf{x})$  的不动点就是  $D_\epsilon(\mathbf{x})$  的驻点. 事实上, 我们将证明这两个点集是相同的. 要做到这一点, 我们需要确定  $D_\epsilon(\mathbf{x})$  和  $g_\epsilon(\mathbf{x}|\mathbf{y})$  的梯度. 记得  $f_\epsilon(u)$  是严格递增和严格凸的. 因此函数  $f_\epsilon(\|\mathbf{x}\|)$  和  $f_\epsilon[d(\mathbf{x}, C_j)]$  是凸的. 更显著的事实是, 这两个函数是连续可微的. 当  $\mathbf{x} \neq \mathbf{0}$  时, 函数  $\|\mathbf{x}\|$  是可微的. 同样地, 当  $\mathbf{x} \notin C_j$  时, 函数  $d(\mathbf{x}, C_j)$  是可微的. 因此, 由链式法则我们可分别得到

$$\nabla f_\epsilon(\|\mathbf{x}\|) = \frac{\|\mathbf{x}\|}{\sqrt{\|\mathbf{x}\|^2 + \epsilon}} \frac{\mathbf{x}}{\|\mathbf{x}\|} = \frac{\mathbf{x}}{\sqrt{\|\mathbf{x}\|^2 + \epsilon}}, \quad (11)$$

$$\nabla f_\epsilon[d(\mathbf{x}, C_j)] = \frac{d(\mathbf{x}, C_j)}{\sqrt{d(\mathbf{x}, C_j)^2 + \epsilon}} \frac{\mathbf{x} - P_{C_j}(\mathbf{x})}{d(\mathbf{x}, C_j)} = \frac{\mathbf{x} - P_{C_j}(\mathbf{x})}{\sqrt{d(\mathbf{x}, C_j)^2 + \epsilon}}. \quad (12)$$

由于连续性, 利用相应的趋于  $\mathbf{0}$  的极限, 我们期待梯度对于  $\mathbf{x} = \mathbf{0}$  和  $\mathbf{x} \in C_j$  是有定义的. 对前者来说, 展开式

$$\sqrt{\|\mathbf{x}\|^2 + \epsilon} - \sqrt{\epsilon} = \sqrt{\epsilon} \sqrt{1 + \frac{\|\mathbf{x}\|^2}{\epsilon}} - \sqrt{\epsilon} = \frac{1}{2} \frac{\|\mathbf{x}\|^2}{\sqrt{\epsilon}} + \sqrt{\epsilon} o\left(\frac{\|\mathbf{x}\|^2}{\epsilon}\right)$$

表明  $\nabla f_\epsilon(\|\mathbf{0}\|) = \mathbf{0}$ . 对后者的情况, 展开式为

$$\sqrt{d(\mathbf{y}, C_j)^2 + \epsilon} - \sqrt{\epsilon} = \frac{1}{2} \frac{d(\mathbf{y}, C_j)^2}{\sqrt{\epsilon}} + \sqrt{\epsilon} o\left[\frac{d(\mathbf{y}, C_j)^2}{\epsilon}\right],$$

对于  $\mathbf{x} \in C_j$ ,  $d(\mathbf{y}, C_j)$  的界  $d(\mathbf{y}, C_j) = |d(\mathbf{y}, C_j) - d(\mathbf{x}, C_j)| \leq \|\mathbf{y} - \mathbf{x}\|$  同样表明  $\nabla f_\epsilon[d(\mathbf{x}, C_j)] = \mathbf{0}$ . 因此, 方程 (11) 和 (12) 对所有的  $\mathbf{x} \in \mathbb{R}^d$  仍然成立. 由此可见, 这两个  $D_\epsilon(\mathbf{x})$  和  $g_\epsilon(\mathbf{x}|\mathbf{y})$  在  $\mathbb{R}^d$  上是可微的, 且梯度分别为

$$\nabla D_\epsilon(\mathbf{x}) = \sum_{j=1}^k \gamma_j \frac{\mathbf{x} - P_{C_j}(\mathbf{x})}{\sqrt{d(\mathbf{x}, C_j)^2 + \epsilon}}$$

和

$$\nabla g_\epsilon(\mathbf{x}|\mathbf{y}) = \sum_{j=1}^k \gamma_j \frac{\mathbf{x} - P_{C_j}(\mathbf{y})}{\sqrt{d(\mathbf{y}, C_j)^2 + \epsilon}}. \quad (13)$$

注意  $\mathbf{y} \in S$  在  $S$  上极小化  $D_\epsilon(\mathbf{x})$ , 当且仅当对所有的  $\mathbf{x} \in S$  有

$$\sum_{j=1}^k \gamma_j \frac{\langle \mathbf{y} - P_{C_j}(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle}{\sqrt{d(\mathbf{y}, C_j)^2 + \epsilon}} \geq 0.$$

这个不等式等同于对所有的  $\mathbf{x} \in S$  成立的不等式

$$\langle \nabla g_\epsilon(\mathbf{y} | \mathbf{y}), \mathbf{x} - \mathbf{y} \rangle \geq 0,$$

而此不等式成立, 当且仅当  $\mathbf{y}$  是  $\psi(\mathbf{x})$  的一个不动点. 若  $D(\mathbf{x})$  是严格凸的, 则  $D_\epsilon(\mathbf{x})$  有唯一的最小值, 且  $\psi(\mathbf{x})$  仅有一个不动点.

因此, 命题 6.3 和命题 6.4 一起表明,  $\mathbf{y}$  是 (6) 的一个解, 如果存在一个序列  $\epsilon_m$  趋于零和点序列  $\mathbf{y}_m$  趋向于  $\mathbf{y}$ , 对所有的  $\mathbf{x} \in S$  满足

$$\left\langle -\sum_{j=1}^k \gamma_j \frac{\mathbf{y}_m - P_{C_j}(\mathbf{y}_m)}{\sqrt{d(\mathbf{y}_m, C_j)^2 + \epsilon_m}}, \mathbf{x} - \mathbf{y}_m \right\rangle \leq 0, \quad (14)$$

上面的充分条件可以成为必要, 如果  $D(\mathbf{x})$  为严格凸的. 作为一个必要的完整性检查, 当集合  $S \cap C_j$  都是空和权重  $\gamma_j$  是相同的情况下, 我们重新获得在文献 [22] 定理 3.2 中给出的最佳点特征, 虽然是在严格凸性这个更严格的假设之下.

## 7. 结束语

无可否认, 运用 MM 的原则有一定的技巧. 在这里介绍的优化不等式是特别应用于 Euclid 距离. 变更相关范数将需要彻底的修订. 然而, 应用 MM 原则时, 相应的 MM 算法是有效的, 容易编程, 且有直观的吸引力. 这里, 这个原则被应用到只用光滑凸分析的初等原理来有效的数值求解 Euclid 版本的广义 Heron 问题. 我们也提出了一个简单而精确的近似来消除 MM 算法和 Weiszfeld 的早期算法的奇异性. 类似优点可应用到更广泛的优化问题. MM 算法在解决高维成像, 基因组学, 统计学和其他的领域问题的能力见证了一贯地运用简单想法的威力. 数学家们热衷于在寻找新的应用.

致谢 (略)

## 参考文献

- [1] M. P. Becker, I. Yang, K. Lange, EM algorithms without missing data, Stat. Methods Med. Res. 6 (1997) 38–54, available at [http://deepblue.lib.umich.edu/bitstream/handle/2027.42/68889/10.1177\\_096228029700600104.pdf?sequence=2](http://deepblue.lib.umich.edu/bitstream/handle/2027.42/68889/10.1177_096228029700600104.pdf?sequence=2).
- [2] A. Ben-Tal, M. Teboulle, A smoothing technique for nondifferentiable optimization problems, in Optimization, Edited by S. Dolecki, Lecture Notes in Mathematics, Vol. 1405, Springer, Berlin/Heidelberg, 1989, 1–11.
- [3] D. P. Bertsekas, Convex Optimization Theory, Athena Scientific, Belmont, MA, 2009.
- [4] J. M. Borwein, A. S. Lewis, Convex Analysis and Nonlinear Optimization: Theory and Examples, Springer, New York, 2000.
- [5] J. Brimberg, The Fermat-Weber location problem revisited, Math. Program. 71 (1995) 71–76.
- [6] ———, Further notes on convergence of the Weiszfeld algorithm, Yugosl. J. Oper. Res. 13 (2003) 199–206.
- [7] L. Cánovas, A. Marín, R. Cañflavate, On the convergence of the Weiszfeld algorithm, Math. Program. 93 (2002) 327–330.
- [8] R. Chandrasekaran, A. Tamir, Open questions concerning Weiszfeld's algorithm for the Fermat-Weber location problem, Math. Program. 44 (1989) 293–295.
- [9] R. Courant, H. Robbins, What is Mathematics?, Oxford University Press, New York, 1961.

- [10] J. Duchi, S. Shalev-Shwartz, Y. Singer, T. Chandra, Efficient projections onto the  $\ell_1$ -ball for learning in high dimensions, in 25th International Conference on Machine Learning — Helsinki 2008, Proc. Int. Conf. Mach. Learn., The International Machine Learning Society, 2008, 272–279.
- [11] S. Gueron, R. Tessler, The Fermat-Steiner problem, Amer. Math. Monthly 109 (2002) 443–451, available at <http://dx.doi.org/10.2307/2695644>.
- [12] J.-B. Hiriart-Urruty, C. Lemaréchal, Fundamentals of Convex Analysis, Springer-Verlag, Berlin, 2001.
- [13] J. Krarup, S. Vajda, On Torricelli’s geometrical solution to a problem of Fermat, IMA J. Math. Appl. Bus. Indust. 8 (1997) 215–224.
- [14] H. W. Kuhn, On a pair of dual nonlinear programs, in Nonlinear Programming, Edited by J. Abadie, North-Holland Publishing, Amsterdam, 1967, 37–54.
- [15] ———, A note on Fermat’s problem, Math. Program. 4 (1973) 98–107.
- [16] K. Lange, Numerical Analysis for Statisticians, second edition, Springer, New York, 2010.
- [17] K. Lange, D. R. Hunter, I. Yang, Optimization transfer using surrogate objective functions (with discussion), J. Comput. Graph. Statist. 9 (2000) 1–20.
- [18] R. F. Love, J. G. Morris, G. O. Wesolowsky, Facilities Location: Models and Methods, Publications in Operations Research Series, Vol. 7, North-Holland Publishing, Amsterdam, 1988.
- [19] R. Meyer, Sufficient conditions for the convergence of monotonic mathematical programming algorithms, J. Comput. System Sci. 12 (1976) 108–121, available at [http://dx.doi.org/10.1016/S0022-0000\(76\)80021-9](http://dx.doi.org/10.1016/S0022-0000(76)80021-9).
- [20] C. Michelot, A finite algorithm for finding the projection of a point onto the canonical simplex of  $\mathbb{R}^n$ , J. Optim. Theory Appl. 50 (1986) 195–200, available at <http://dx.doi.org/10.1007/BF00938486>.
- [21] B. Mordukhovich, N. M. Nam, Applications of variational analysis to a generalized Fermat-Torricelli problem, J. Optim. Theory Appl. 148 (2011) 431–454, available at <http://dx.doi.org/10.1007/s10957-010-9761-7>.
- [22] B. Mordukhovich, N. M. Nam, J. Salinas, Solving a generalized Heron problem by means of convex analysis, Amer. Math. Monthly 119 (2012) 87–99.
- [23] ———, Applications of variational analysis to a generalized Heron problem, Appl. Anal. (2011) 1–28.
- [24] J. M. Ortega, W. C. Rheinboldt, Iterative Solutions of Nonlinear Equations in Several Variables, Academic, New York, 1970.
- [25] R. T. Rockafellar, Convex Analysis, Princeton University Press, Princeton, NJ, 1996.
- [26] A. Ruszczyński, Nonlinear Optimization, Princeton University Press, Princeton, NJ, 2006.
- [27] Y. Vardi, C.-H. Zhang, A modified Weiszfeld algorithm for the Fermat-Weber location problem, Math. Program. 90 (2001) 559–566.
- [28] E. Weiszfeld, Sur le point pour lequel la somme des distances de  $n$  points donnés est minimum, Tôhoku Mathematics Journal 43 (1937) 355–386.
- [29] E. Weiszfeld, F. Plastria, On the point for which the sum of the distances to  $n$  given points is minimum, Ann. Oper. Res. 167 (2009) 7–41, available at <http://dx.doi.org/10.1007/s10479-008-0352-z>.
- [30] G. O. Wesolowsky, The Weber problem: Its history and perspectives, Location Science 1 (1993) 5–23.

(秦孟兆 译 李胜台 校)